Estimating the Effectiveness of City Connects on Middle School Outcomes

Author: Chen An

Boston College Electronic Thesis or Dissertation, 2015

BOSTON COLLEGE

Lynch School of Education

Department of
Educational Research, Measurement, and Evaluation

ESTIMATING THE EFFECTIVENESS OF CITY CONNECTS ON MIDDLE SCHOOL OUTCOMES

Dissertation
by

CHEN AN

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

August 2015

ESTIMATING THE EFFECTIVENESS OF CITY CONNECTES

ON MIDDLE SCHOOL OUTCOMES

by

Chen An

Dr. Henry I. Braun, Dissertation Chair

ABSTRACT

City Connects is a school-based model that identifies the strengths and needs of every student and links each child to a tailored set of intervention, prevention, and enrichment services in the school or community. The purpose of this study was to conduct a comprehensive evaluation of the City Connects treatment effects on academic performance (both MCAS scores and grade point average (GPA) grades) in middle school using student longitudinal records. Parallel analyses were conducted: one evaluated the City Connects elementary intervention (serving kindergarten to fifth grades) and the other one evaluated the City Connects middle school intervention (serving sixth to eighth grades). A series of two-level hierarchical linear models with middle school achievement scores adjusted and/or propensity score weights applied were used to answer the research questions of interest. In addition, to make a causal inference, a sensitivity analysis was conducted to examine whether or not the estimated treatment effects resulted from the first two analyses were robust to the presence of unobserved selection bias.

The results showed that students who were exposed to the City Connects elementary intervention significantly outperformed their counterparts, who graduated from the comparison elementary schools, on academic achievement in all middle school

grades. However, in the case of the City Connects intervention schools that served middle school grades, since all students only received a maximum of one year of City Connects middle school intervention, it was still too soon to expect any significant changes. Moreover, the estimated treatment effects of the City Connects elementary intervention were only mildly sensitive to the presence of some forms of hidden bias, which made the causal inference of City Connects on middle school academic achievement quite plausible.

ACKNOWLEDGMENTS

Writing a dissertation is a long and tedious journey. I am so lucky that I am not alone. I could have never reached to this point without the guidance, support, thoughtfulness, and understanding of lots of people around me. First, I would like to express my deepest gratitude to my dissertation chair, Dr. Henry Braun, for guiding me through all the important issues in my dissertation, encouraging me to reach my limit, and always supporting me to develop professionally. I will never forget how many drafts that Dr. Braun revised for me with great patience and how these revisions are full of intelligence and thoughtfulness. My dissertation has gained great depth during these two years due to his diligent and brilliant inputs.

My deepest appreciation also goes to Dr. Mary Walsh, one of my committee members and my boss at City Connects. I am greatly thankful for being able to use the City Connects data and for her unconditional support in exploring the topic of my interest. I have learned so much during my years with City Connects and this experience is invaluable.

In addition, I would also like to express my sincere gratefulness to Dr. Mandy (Zhushan) Li, the third member of my dissertation committee. Dr. Li, thank you so much for your support throughout the process and your final re-arrangement to make my defense possible on that day. I will never forget about it and please give my great blessings to the new born.

Next, I would like to thank Mrs. Anastasia Raczek, my advisor at City Connects. Thank you for your support and encouragement for these many years. And of course, thank you so much for reviewing my dissertation. You profound knowledge of the City

Connects intervention polished my dissertation and sharpened my understanding of the theories and practices regarding student support. Dr. Terrence J LeeStJohn, thank you so much for helping me dividing the first two research questions and answering all my annoying emails and questions patiently and timely. Without your suggestions, the first two research questions could have never been this clearly-defined.

Moreover, I would like to thank Dr. Matt Gregas, Miss Rossella Calvi, Mrs. Wei Qiu, and the whole team at the Boston College Research Services. Thank you so much for your support, guidance, and trainings in using the Linux cluster and in helping me programming scripts in different statistical software packages.

I would also like to thank Zihan Li, my former roommate, for his tremendous efforts in helping me figuring out the maximum number of trials that I could simulate given the limited computer capacity, various creative ways to increase efficiency, and how to set up 3-dimensioanl graphs using different mathematical applications. And a special thank you to all my beloved friends, Baoxin Zhu, Mark Shen, Lefei Tong, Gong Lv, and Xiaodan Huang, for generously lending me your high-performing desktops for such a long time. Without your help, I could have never got the analysis completed within a realistic time frame.

Also, Wenfeng Huang, my dear fiancé, thank you so much for being patient and understanding. Your support is what really pushes me to move forward. I am blessed to have you as part of my life.

Last, I would like to thank myself, for not giving up and always struggling to move on. And please allow me to send a big salute to those sleepless nights and stressful moments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

CHAPTER 1. INTRODUCTION

Middle school presents many challenges to students. It serves as a transitional period during which young adolescents start to cope with dramatically-changing social life and develop independent thinking skills. It is also an essential period to prepare students academically for high school.

There is an urgent need to improve academic achievement in U.S. middle school. According to the Third International Math and Science Study (TIMSS) 2011, U.S. fourth-graders scored higher than most of their peers and ranked among the top 15 countries and regions in Mathematics and among the top 10 in Science; however, by the eighth grade, U.S. students dropped to the 24th place in Mathematics and the 23th place in Science on the list of 57 participating countries and regions (Provasnik et al., 2012). The results of Programme for International Student Assessment (PISA) 2012 showed similarly mediocre academic performances among 15-year-old U.S. students: mean scores in Mathematics and Science for U.S. students are 481 and 491, both of which are below the OECD average of 494 and 501, separately. Although the mean score in Reading for U.S. students is higher than the international average, the difference is negligible (the mean score is 498 for U.S. students as compared to the international average of 496) (OECD, 2013).

Furthermore, the No Child Left Behind Act of 2001 (NCLB) emphasized the importance of academic success and held schools accountable for student academic achievement (2002). Under NCLB, schools are sanctioned if they fail to make adequate yearly progress (AYP) on their students' test performance. In the state of Massachusetts, students in middle school grades (Grades 6, 7, and 8) are mandated to take the state-wide standardized assessment (Massachusetts Comprehensive Assessment System or MCAS) annually in the subjects of English Language Arts (ELA) and Mathematics. Consequently, it is of great importance to accurately identify and thoroughly evaluate schools or educational programs that are effective in boosting student academic performance, so that their success can inform the practices in other schools.

Urdan and Klein (1998) argued that in order to meet early adolescents' developmental needs so that they can achieve academic success, a focus should be placed on the interaction between individuals and the context that includes the school, home, and community. Heller, Calderon, and Medrich echoed this point and encouraged middle school reform frameworks that possess the following features in order to foster student achievement: "an strengthened curriculum, a modified school organization and practices, increased demands on students with essential support provided, improved teacher quality, and the inclusion of parents and community in the learning process" (2003, p. 11).

One such program is the City Connects intervention. City Connects implements theoretically-guided practices for student support in high-poverty, urban schools. It began in 2001 in response to the recognition that non-academic barriers to learning seriously impede students' ability to benefit from instruction in underperforming schools (Walsh & Brabeck, 2006). City Connects is implemented by a School Site Coordinator who collaborates with classroom teachers to identify strengths and needs in academic, social-emotional/behavioral, family, and health/medical domains for every student in each class. City Connects connects students with a tailored set of prevention, intervention, and enrichment services provided by local community agencies; documents the service plan; and provides follow-ups to assure delivery and assess effectiveness (Walsh, Kenny, Wieneke & Harrington, 2008). Initially serving kindergarten to fifth grades, City Connects was later extended to pre-kindergarten and middle school in 2008.

One of the major indicators of the success of the City Connects intervention is its impact on academic achievement. Past City Connects investigations have demonstrated that the City Connects intervention had a significant positive impact on students' academic and thriving report card scores in elementary school. In addition, City Connects also had a significantly positive impact on GPA in middle school. Although a similar positive impact of City Connects on MCAS scores was not observed in elementary grades, the intervention was associated with long-term MCAS gains that manifested in later middle school grades (Walsh et al., 2014).

1.1 Purpose and Research Questions

The purpose of this study was to conduct a comprehensive evaluation of City Connects treatment effects on academic performance (both MCAS scores and grade point average (GPA) grades) in middle school using student longitudinal records. Parallel analyses were conducted: one evaluates the City Connects elementary intervention (serving kindergarten to fifth grades) and the other one evaluates the City Connects middle school intervention (serving sixth to eighth grades). Three research questions were addressed in this study:

RQ1/RQ2. What is the impact of the City Connects elementary/middle school intervention on middle school achievement as measured by standardized MCAS scores and criterion-referenced GPA grades?

This includes three sub-questions:

a. After controlling for student characteristics and pre-existing academic achievement differences, does receiving the City Connects intervention in elementary/middle school help students succeed in middle school and does this success persist through the entire middle school years (RQ1a and RQ2a)?

b. Do students who graduated from different City Connects elementary schools follow significantly different academic achievement patterns in middle school (RQ1b); and do students who received the City Connects intervention in different middle schools generate significantly different academic achievement patterns in middle school (RQ2b)?

c. If City Connects elementary/middle school effects are observed, to what extent can this be accounted for by both student and other school characteristics (RQ1c and RQ2c)?

RQ3. Are the estimated treatment effects resulted from the first two analyses robust to the presence of unobserved selection bias that may jeopardize causal inferences?

1.2 Significance of the Study

This study differs from other City Connects middle school outcome evaluation analyses in five critical ways: 1) it was the first time a longitudinal approach was employed to track cohorts of students progressing through middle school; 2) it examined the relative school effectiveness among City Connects elementary schools so that further improvement could be made to address the special strengths or needs of each school; 3) it was also the first time that the City Connects middle school intervention was scrutinized; 4) it included institution-level characteristics so that the unique contribution of the City Connects intervention could be disentangled from confounding factors; 5) it tested the robustness of the conclusions to hidden bias so that causal inferences could be reasonably made.

This study is potentially of great significance to the current evaluation of the City Connects intervention, the understanding of differential City Connects effects if they exist, and the literature on value-added models (VAMs).

First, it is one of the major responsibilities of the City Connects evaluation team to obtain credible estimates of the treatment effects. The proposed analysis is an exciting supplement to the existing evidence in order to get approximately unbiased estimates of the effectiveness of the City Connects intervention. Furthermore, with the addition of including institution-level characteristics in the outcome models and conducting a sensitivity analysis to hidden bias, reasonable causal inferences can be made.

Second, the use of VAMs approach will further help improving the quality of the City Connects evaluation by identifying schools that are most or least successful in obtaining benefits from the implementation of City Connects so that more aimed intervention will be able to take place. In addition, VAM scores can be used to promote the City Connects intervention to a broader audience. Interested principals, school district administrators, and other educators can relate their own schools to specific City Connects schools that are similar to theirs in every possible way. Observing how these City Connects schools progressed over years and gradually outperformed other schools, if it is the case, will be critical in helping them make the decision to join the intervention in the future.

Third, VAMs have been widely used for school accountability. However, few of them have been employed to examine school effectiveness via a complex structure with multiple longitudinal data points and with an intervention. The study provided additional empirical evidence of using such models in more complicated reality to answer research questions of interest.

1.3 Dissertation Organization

Chapter 1 emphasizes the importance of academic success in middle school, followed by a brief overview of City Connects, an educational intervention that has been empirically demonstrated to help students succeed in middle school. The chapter then introduces the purpose, research questions of interest, and the significance of this study. The chapter ends with an outline of the organization of the dissertation.

Chapter 2 begins with a discussion of the difference between experiments and quasi-experiments, followed by an explanation of the selection bias threat to causal inferences, the key issue associated with the latter. The chapter then reviews both statistical methods and research design elements to address the issue. Next, the rationale of the City Connects intervention, its major challenges, and the current evaluation models and results are elaborated. The next section proposes statistical techniques tailored to address selection bias, both overt and hidden bias, which threatens the validity of causal inferences of City Connects due to the quasi-experimental nature of any City Connects study. At last, in order to examine differential City Connects school effectiveness, VAMs are introduced; using VAMs in addressing research questions of interest and the VAM choice are justified.

Chapter 3 presents detailed exposition of the three research questions and proposes statistical models to address them. The outcome variables of interest and the student- and school-level covariates used in the models are scrutinized, followed by a discussion of centering decisions. The last section of Chapter 3 is devoted to a preliminary analysis examining descriptive baseline student characteristics and assessing covariate balance between the two groups after applying the corresponding statistics that aims at removing overt selection bias.

Chapter 4 presents empirical data analysis results. Each research question is answered via a series of statistical analyses. Tables of results and visual displays of findings are shown. Finally, Chapter 5 discusses the results, explores implications of the findings to the VAMs and sensitivity analysis literature, discusses limitations, and suggests potential directions for future research.

CHAPTER 2. LITERATURE REVIEW

2.1 Experimental Designs versus Quasi-experiments

First proposed by Fisher, experimental designs have been viewed as the “gold standard” of all research designs (1935). As Kirk stated in 1995, “an experimental design is a plan for assigning subjects to experimental conditions and the statistical analysis associated with the plan” (p.1). According to Kirk, the primary goal of an experimental design is to identify the causal relationship between the independent (the assumed causes) and dependent variables (the outcomes). The key element of such a design is the utilization of randomization in the assignment of units to the treatments under study.

Random assignment applies chance procedures to ensure that participants have a known probability to be assigned to the treatment and the control groups. The logic behind it is that with large sample size and through randomization, the differences in all the relevant observed and unobserved characteristics between treatment and control groups before the intervention are small and due to chance rather than some systematic discrepancies among them. This allows one to claim that observed statistically significant differences in outcomes between the two groups after the intervention can reasonably be attributed to the intervention alone.

Over the past decades, such randomized controlled trials (RCTs) have been advocated by researchers and educators to estimate causal relationships. According to the What Works Clearinghouse (WWC), an initiative of the U.S. Department of Education's Institute of Education Sciences (IES) that was created in 2002 to be a central and trusted source of scientific evidence for what works in education, well-designed and well-implemented RCTs are considered strong evidence, while quasi-experimental designs with equating may only meet “standards with reservations” (2011, p.11).

Chapter 2. Literature Review

However, in reality, due to practical and ethical reasons, experimental designs may not always be feasible so researchers turn to quasi-experiments instead. Shadish and his colleagues (2002) defined quasi-experiments as follows:

> Quasi-experiments share with all other experiments a similar purpose – to test descriptive causal hypotheses about manipulable causes – as well as many structural details, such as the frequent presence of control groups and pretest measures, to support a counterfactual inference about what would have happened in the absence of treatment. But, by definition, quasi-experiments lack random assignment. Assignment to conditions is by means of self-selection, by which units choose treatment for themselves, or by means of administrator selection, by which teachers, bureaucrats, legislators, therapists, physicians, or others decide which persons should get which treatment. (pp. 13-14)

Quasi-experiments have been widely applied in education for years. For example, it is sometimes infeasible to randomly assign students to different treatments such as after-school tutoring programs because students need to be identified with learning disadvantages to participate. As a result, students in the control group (do not receive the program) and those in the experiment group (do receive the program) are different to begin with. Another example occurs in the study of the effectiveness of a specific charter school program: students who are enrolled in a specific charter school may differ in many ways (such as prior academic achievement and motivation) from those who are not. These differences will be confounded with the treatment effect that researchers would like to estimate.

Thus, a major disadvantage of quasi-experiments is that they potentially suffer from selection bias, a type of threat to internal validity. Internal validity in both experimental designs and quasi-experiments is about the credibility of causal inferences. It refers to “inferences about whether observed co-variation between A and B reflects a causal relationship from A to B in the form in which the variables were manipulated or measured” (Shadish et al., 2002, p.53). In other words, are the differences we observed between the experimental and the control group on the outcome due primarily to the intervention? Are there any other extraneous variables that influence the outcome?

Campbell and Stanley (1963) identified eight threats that can jeopardize internal validity of causal inferences. They are selection, history, maturation, testing, instrumentation, statistical regression, experimental mortality/attrition, and selection interactions. Selection refers to the effect of having non-equivalent treatment and control groups. Treatment and control groups should be statistically equivalent on all the observed and unobserved variables at the beginning (in other words, they are different only by chance). However, without randomization, it is difficult to defend the claim of equivalence of the treatment and control groups.

2.2 Methods to Reduce Selection Bias

2.2.1 Regression Adjustments

In order to reduce selection bias, researchers have done extensive work to search for appropriate statistical adjustments to make the two groups as similar as possible. The most straightforward statistical adjustment is through regression analysis, which estimates the effect of the intervention on the outcome conditional on one or more covariates. Regression adjustment removes differences in outcomes that can be accounted for by differences in the observed covariates: what remains is attributed to the treatment effect.

There are several major drawbacks to use regression adjustments to eliminate selection bias. First, regression adjustments use the outcome measures in the estimation equations and thus may result in manipulating regression models to achieve favorable results. As Rubin (2001) argued,

> The most important feature of experiments is that we must decide on the way data will be collected before observing the outcome data. If we could try hundreds of designs and for each see the resultant answer, we could capitalize on random variation in answers and choose the design that generated the answer we wanted (p.169).

In other words, it is essential to design an experiment before obtaining the outcomes to avoid manipulating data to achieve desired results.

Second, Stuart and Rubin (2007) stated that “when there are large differences in the covariate distributions between the groups, standard model-based adjustments rely heavily on extrapolation and model-based assumptions” (p.157). Rubin (2001) mentioned some basic conditions for regression analysis to be trustworthy. If these conditions are not met, as Rubin argued, “the differences between the distributions of covariates in the two groups must be regarded as substantial, and regression adjustment will be unreliable and cannot be trusted (2001, p.174).

The third drawback of regression analysis is related to its’ inability to make strong causal inferences. Correlation does not equal causation. Berk (2004) stressed the point that one cannot make causal inferences based on a regression analysis alone. To be specific, he stated the following:

- Standardized coefficients do not represent the causal importance of a variable.
- Contributions to explained variance for different predictors do not represent the causal importance of a variable.
- A good overall fit does not demonstrate that a causal model is correct. (p.224)

Berk (2004) also argued that only when one has a correctly-specified regression model, together with information on how reasonable the variables in the model are able to explain the outcome based on literature and common sense, can one gain confidence in applying various regression diagnostics, specification tests, and mathematical formalisms to make the causal argument.

2.2.2 Propensity Scoring Matching Methods

One family of methods that has drawn growing attention to reduce selection bias when using observational data employs matching. The basic idea of the matching methods is to select a sub-sample of the comparison group that is statistically equivalent to the treatment group on all the observed covariates. The covariate distributions of the two groups are approximately the same, thus creating equivalent treatment and control groups that approximate what is accomplished in an experiment, although one should keep in mind that unobserved differences between the two groups may remain.

An important breakthrough for such methods is the introduction of the propensity score (PS) matching by Rosenbaum and Rubin. In 1983, they published a seminal paper on the theory and the application of propensity scores to analyze observational studies. Since then, a variety of PS models were developed and refined. Guo and Fraser (2010) summarized all the existing PS models as a three-step analytic process:

The first step is to generate propensity scores. Analysts select a number of covariates that are considered to create the “imbalance” between the treatment and the control groups. This imbalance, represented by the statistically significantly differences in the outcome between the two groups, is believed to be a result of the non-random selection process that needs to be carefully modeled. A typical practice is to estimate through logistic regression the conditional probability of receiving treatment. Those conditional probabilities are defined as propensity scores.

The second step to apply propensity scores to the sample in different ways. Analysts can either use propensity scores to match participants; or use them as sampling weights to avoid losing participants; or conduct analysis of weighted mean differences using kernel or local linear regression. If propensity scores are used for matching, the third step will involve some post-matching analyses, such as multivariate analysis based on the matched sample and stratification.

PS matching makes participants in the treatment and the comparison groups as similar as possible in terms of propensity scores. The major advantage is that only a single score has to be used, thus solving the problem of matching multiple covariates simultaneously. Sometimes one may find neither an exact match nor a proximate match with tolerance from the comparison pool for every member of the treatment group, and thus have to drop unmatched participants. In this sense, PS matching can be viewed as a resampling procedure. Although the original sample is unbalanced on observed covariates between treatment and comparison conditions, the new sample will be balanced on such covariates by using propensity scores.

In contrast to the regression analysis approach, one of the advantages of PS matching models is that they do not use outcome data to estimate propensity scores. As a result, these models avoid capitalizing on random variation. In addition, PS matching models limit reliance on the untestable regression assumptions and thus are robust to violations of such assumptions. However, they are still dependent on logistic regression assumptions and all the variables in the models need to be measured without error or nearly so.

However, PS models are not above criticism. To begin with, they reduce selection bias by using observed covariates to balance the two groups; however, one can argue that significant hidden biases (unobserved factors that are highly correlated with the outcome) may remain. Rubin made it clear that the major drawback of PS methods is that they only adjust for observed covariates and he stated, “this is always a limitation of nonrandomized studies compared with randomized studies, where the randomization tends to balance the distribution of all covariates, observed and unobserved” (1997, p.762). The presence of hidden bias and the robustness of PS models to such bias can be assessed through simulation studies and sensitivity analysis (Cornfield, et al., 1959; Montgomery, Richards, & Braun, 1986; Rosenbaum, 1991).

Second, critics of PS models are skeptical about the tenability of assumptions that these models generally hold in real settings. Michalopoulos, Bloom, and Hill (2004) assessed estimates that were obtained from non-experimental approaches, including various PS methods such as one-to-one matching and PS weights, using results from a six-state random assignment study of mandatory welfare-to-work programs. The basic method to estimate bias was to examine the difference in outcomes between the randomly selected control group and the non-randomly chosen comparison group (through PS methods). They concluded that all non-experimental estimators displayed significant bias. To be specific, they discovered that in-state comparison groups produced smaller bias than out-of-state groups and long-term outcomes suffered from greater bias than short-term ones, suggesting that PS methods correct less well if the two study groups are not exposed to the same ecological or social context.

Third, critics of PS methods compared the results of regression analysis and those of PS methods and did not find significant differences. For instance, in 2004, Shah, Laupacis, Hux, and Austin systematically reviewed published observational studies that applied both regression analysis and PS methods to control for confounding covariates. They concluded that the two approaches yielded similar results. However, in some rare cases, regression analysis indicated statistically significant association which was not found with propensity scores. Shah and his colleagues then attributed this to PS methods being slightly more conservative measures of association than regression analyses. They also argued that dropping cases due to not being able to find good matches may result in reduced statistical power for certain PS methods.

2.2.3 Research Designs

In addition to statistical adjustments discussed before, statisticians and econometricians also developed a variety of quasi-experimental designs to duplicate key features of randomized experiments. Examples of such designs include regression discontinuity (RD) and interrupted time series (ITS) designs. These designs are believed to be more powerful tools in making causal arguments than statistical adjustments. As Rubin (2008) stated, “For objective causal inference, design trumps analysis” (p. 1). Cook and Shadish (2012) ranked causal research designs and statistical adjustments as follows: an experiment first, RD next, followed by ITS, and then different matching methods.

2.3 The City Connects Intervention

Beginning in the 1960’s (Coleman, et al., 1966; Harrington, 1962), it has been recognized that life outside of school has consequences for achievement in school, especially for students growing up poor. Achievement gaps persist between poor and not-poor students due to both within school and out of school factors (Barton, 2004; Becker & Luthar, 2002; Berliner, 2009; Hanushek, Kain, Markman, & Rivkin, 2003; Wright, Horn, & Sanders, 1997). Children in poverty experience non-academic barriers to learning (Adelman & Taylor, 2005; Anderson-Butcher et al., 2008; Walsh & Murphy, 2003), which may impede them from engaging in daily school activities and making best use of their academic time. These barriers include: physical and mental health issues such as poor nutrition and depression; behavioral issues such as disruptive and unruly behavior, alcohol, tobacco, and drug use; social-emotional issues such as defective impulse-control or anger management; family issues such as family violence, abuse and neglect, and homelessness; negative peer influences; and experiences of traumatic events (Ohio Mental Health Network for School Success, 2004). Not surprisingly, schools cannot address all these factors by themselves and strong school-community partnerships are needed.

School counselors play an essential role in building these partnerships. Amatea and Clark (2005) identified four key roles of school counselors as perceived by school administrators: the innovative school leader, the collaborative case consultant, the responsive direct service provider, and the administrative team player. The current national model proposed by American School Counselor Association (2003) strongly emphasized the importance of school counselors' collaboration with parents and other educators in order to create an environment that promotes student achievement.

However, building strong school-community partnerships requires overcoming many challenges including "(a) fragmentation and a limited range of services, (b) turf conflicts between school-based and community-based providers, and (c) insufficient funding" (Walsh & Depaul, 2008, p.769). The City Connects model was developed to provide a comprehensive practice in building strong school-community partnerships in respond to such challenges.

2.3.1 An Overview of City Connects

Students whose academic and social/emotional, health, and family strengths and needs are being met may exhibit an increased capacity to come to school prepared to engage and learn (Ayoub & Fischer, 2006; Noguera, 2011). The mission of City Connects is to help children engage and learn in school by connecting each child with a tailored set of prevention, intervention, and enrichment services that she or he needs to thrive (City Connects, 2013). This school-based intervention is systemic, making student support a core function of a school. City Connects does not directly impact pedagogy and classroom instructional practices. Instead, it aims at facilitating a positive school climate in which: academic learning is promoted; students' emotional and behavioral problems are addressed; job satisfaction for school personnel is enhanced; communication and understanding between parents and teachers are strengthened; and referrals to services are focused. As a result, students' academic achievement improves.

A full-time City Connects staff member in each school called the School Site Coordinator (SSC) is at the core of the intervention. A SSC is a licensed school counselor or school social worker who holds a master's degree. In the fall of the school year, the SSC collaborates with each classroom teacher to develop a customized support plan for every student by identifying strengths and needs across four domains (academic, social/emotional/behavioral, health, and family) and identifying appropriate school- and/or community-based services and enrichments. During this Whole Class Review (WCR) process, the SSC, the classroom teacher, and a third staff member from the school or a community partner collaborate to systematically assess each student's strengths and needs across the four domains and place students into the following four tiers (City Connects, 2013):

- Tier 1. Strengths & minimal Risks
- Tier 2a. Strengths & Mild Risks
- Tier 2b. Strengths & Moderate Risks
- Tier 3. Strengths & Severe Risk

Moreover, students with intensive needs (e.g. students being categorized into Tier 3) may also receive an Individual Student Review (ISR), which is a process that brings together a wider group of professionals (e.g. educational team facilitators, school psychologists, teachers, principals, nurses, and occasional community partner staff members) to discuss goals and strategies to help those students.

Once the WCR and the subsequent ISR process are finished, SSCs and the group will develop a unique plan for each student; SSCs will then refer students to tailored sets of services offered by school- or community-based providers to address students' needs and enhance their strengths. These services fall into three broad categories:

- Preventive and Enrichment: examples include before and after school programs, sports activities, academic and youth development enrichments;
- Early Intervention: examples include academic support, English as a Second Language (ESL), classroom-based social skills and health intervention, adult mentoring, tutoring, family support and assistance;
- Intensive/Crisis Intervention: check-in with SSCs, mental health and family counseling, informal screening/diagnostic, SPED evaluation and screening.

SSCs are responsible for identifying appropriate services and corresponding service providers for students, organizing existing resources, formalizing partnerships with community agencies, and building new partnerships. Finally, SSCs are also responsible for documenting the service plan and following-up to assure service delivery. The SSC also serves as a primary point of contact for families in the school and takes part in additional programmatic responsibilities, such as leading small social skills groups with students.

2.3.2 Challenges in Categorizing the City Connects Treatment

It is essential to understand the complexity of the Boston Public Schools (BPS) data and the nature of the City Connects intervention. Perhaps the most problematic factor related to school data is student mobility. Ideally students would enroll in one of the BPS schools and stay there for their entire education career. In reality, however, students often transfer in and out of schools. Some students may even repeat grades at different schools. Further, if a student transfers out of the district, his or her entire subsequent record will be completely missing from the dataset.

Additionally, implementation of City Connects in BPS schools has varied over time. It was first implemented in six elementary and K–8 schools in two large neighborhoods; two schools from a different geographical area were added after one year, and two years after that, the district requested an expansion to all seven elementary schools in that area. Due to later expansions of the intervention as well as school closings over time, the number of years participating schools implemented City Connects is not consistent. The City Connects intervention itself is also an evolving program. Launched more than ten years ago, the strategies and practices of City Connects have evolved over time, although the core components have been consistent.

Given these complexities, it is necessary to clearly define the City Connects dose and dosage variables. Due to student mobility, the City Connects dose, a dummy variable that represents treatment group membership, was defined as ever attended a City Connects school during a year when the City Connects intervention was implemented in that school. However, it is inadequate to use a single variable to represent the City Connects treatment. The City Connects dosage variable is used to represent the number of years experiencing the City Connects treatment. To tackle the varying nature of the City Connects intervention, Lee-St. John (2013) developed a *two-dimensional stratification strategy* to examine causal treatment-effects, which is a dosage framework to capture all possible cohort and treatment patterns[1] of the City Connects intervention. Cohort and treatment-pattern served as the two dimensions and together they comprised a set of exhaustive, discrete, and mutually exclusive classification cells. A treatment effect was estimated for each cell and then summary patterns were identified.

2.3.3 Past Cross-Sectional Outcome Models and Results

The evaluation design of the City Connects is a quasi-experiment because participating schools were identified by the district to receive the full program treatment. Further, because serving all students within a school is a critical feature of City Connects, individual students could not be assigned to treatment or control conditions. Since neither students nor schools are randomly assigned to the City Connects intervention, the estimation of the treatment effects must contend with the problem of selection bias. To address this problem, the analysis of the City Connects intervention applied PS weighting methods at the student-level to adjust the treatment and comparison groups to be approximately equivalent in all aspects except treatment assignment (An and Wong, 2012). The reason for choosing PS weighting from among various PS methods is because it maintained the largest possible sample size at a time when only a limited comparison group of seven randomly-selected non-City Connects schools was available to the project.

The major outcome models of the City Connects evaluation are weighted ordinary least square regression models: MCAS Mathematics and ELA (Grades 3-8) scores and Report Card (Grades 3-5) grades are used as the outcomes; students' demographics

[1] Depending on numbers of years participating in CCNX (4-6 years or less than 3 years), grade of initial enrollment in CCNX (during Grades K-2 or 3-5), and retention (being required to repeat a grade or not in elementary school), students were classified into a series of treatment-patterns.

(gender, race, bilingual status, free or reduced lunch status, special education status, and school mobility) are used as covariates; the City Connects dose and dosage variables are the treatment effects to be estimated; and PS weights are applied to all the regressions to make the two groups approximately equivalent on observables at the baseline[2]. The weights are produced by means of multinomial regression models that predict the probability of receiving each level of the City Connects dosage based on an extensive student profiles (gender, race, bilingual status, free or reduced lunch status, special education status, school mobility, distance to school, age, and Report Card Reading and Math grades) at the baseline. The weights are the inverse of the predicted probabilities. Some extreme weights that could have significantly influenced the results were trimmed (City Connects, 2011).

2.4 Addressing Selection Bias in this Study

The first section of Chapter 2 discussed quasi-experiments, their limitations in making credible causal inferences, and the solutions proposed in literature. The general implementation and evaluation models of City Connects were elaborated in the second section. In the third section, statistical techniques to deal with selection bias for this specific study will be presented.

2.4.1 Propensity Score Weighting to Remove Overt Selection Bias

In this study, for the purpose of maintaining the largest possible sample size, PS weights were used to remove the pre-existing differences between the treatment and the control groups. Therefore, it is necessary to clarify the process and introduce proper statistics to examine covariate balancing.

[2] For elementary Report Card grades, the baseline is Grade 1 fall records; for MCAS scores and middle school GPA grades, the baseline is Grade 2 fall records.

With a binary treatment indicator Z, where $Z = 1$ for treated units and $Z = 0$ for control units, Rosenbaum and Rubin (1983a) defined $e(x)$, the PS, as the conditional probability of receiving the treatment given the pre-treatment variables x:

$$e(x) = \Pr(Z = 1 \mid X = x) \qquad (2.1)$$

Propensity scores can be used as sample weights. As Guo and Fraser (2010) defined, for estimating the average treatment effect, PS weights are expressed as

$$\mathrm{w}(\mathrm{Z}, \mathrm{x}) = \frac{\mathrm{Z}}{\hat{\mathrm{e}}(\mathrm{x})} + \frac{1-\mathrm{Z}}{1-\hat{\mathrm{e}}(\mathrm{x})} \qquad (2.2)$$

where $\hat{e}(x)$ is the estimated probabilities of receiving treatment.

Under this definition, for a treatment unit ($Z = 1$), the PS weight is $1/\hat{e}(x)$; while for a control unit ($Z = 0$), the PS weight is $1/1\text{-}\hat{e}(x)$.

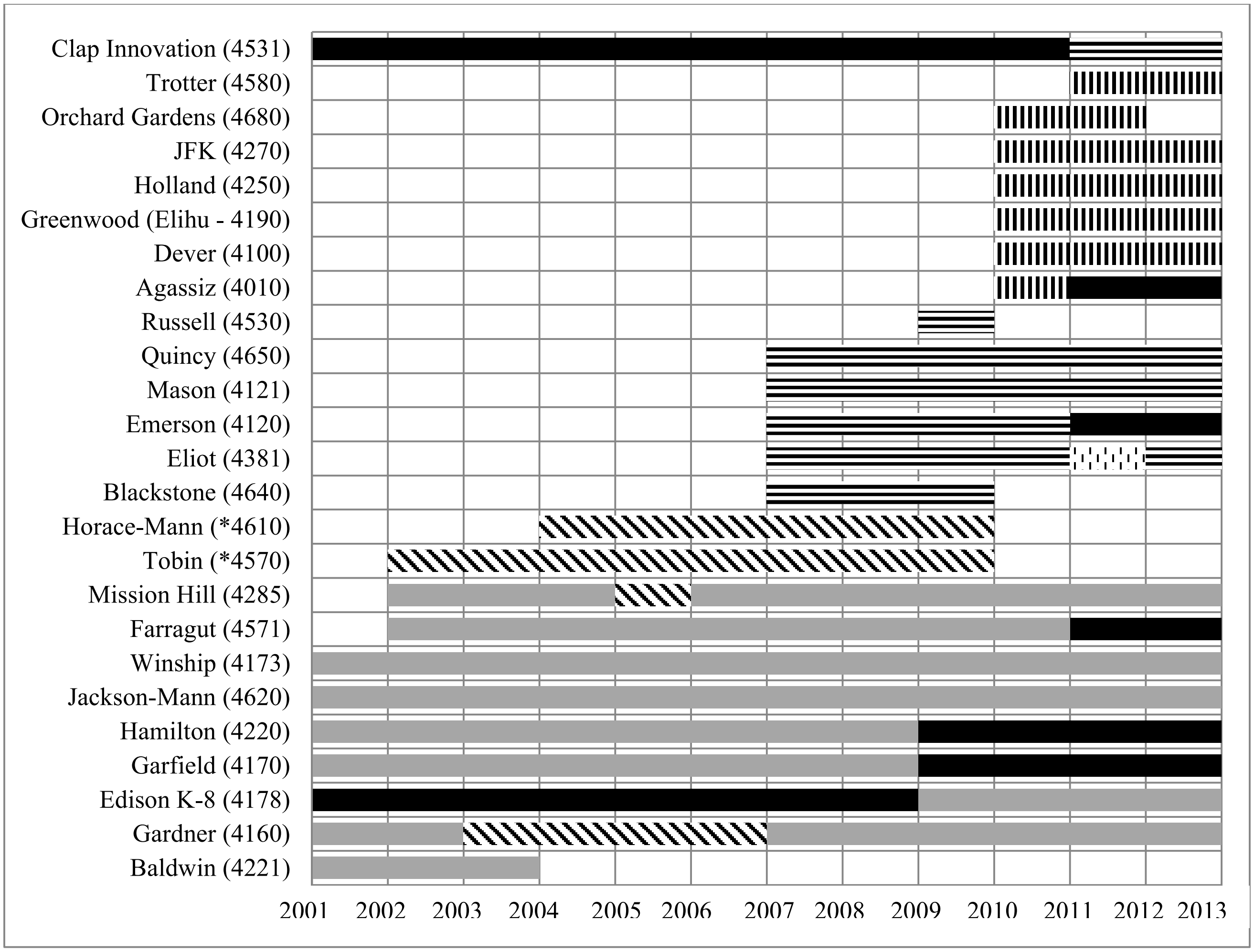
To deal with treatment with multiple conditions (dosage), Imbens (2000) extended the basic definition and developed the generalized PS as the conditional probability of receiving a particular level of the treatment given those pre-treatment variables. Using multinomial logistic regression, the inverse of these estimated probabilities are the generalized PS weights.

Theoretically speaking, either City Connects dose or dosage can be estimated to generate PS weights. However, dosage is not a linear function that can be easily modelled. First, since dosage is defined as years spent with City Connects, it is possible for students who repeat grades to have higher dosage. If one assumes higher dosage will benefit students more, then the results could be confusing because repeaters are more likely to be low-performing students with high dosage. If there were a high percentage of repeaters in the relatively high dosage group, the results may be distorted. Second, schools joined the City Connects intervention at different time points (see Figure 2.1):

eight “legacy schools” have stayed with City Connects since 2001 or 2002; five “Cluster 2 schools” were added in 2007; a number of turnaround schools have started to implement City Connects since 2010. Additionally, some schools had the City Connects intervention for a short period of time and decided to withdraw from the study (e.g., Russell Elementary School). During the past thirteen years, City Connects has developed and improved its implementation model exponentially. The strategies and practices of the early City Connects model when legacy schools started were different from what were implemented later on. Therefore, the one-year dosage received by a student enrolled in a legacy school in 2001 is not the same as the one-year dosage experienced by another student who was enrolled in a Cluster 2 school in 2008. Simply put, the selection model cannot be modelled accurately unless how and why students received different types and levels of City Connects dosage is fully understood. Therefore, the binary City Connects dose was estimated when generating PS weights in this study.

The existence of selection bias implies that the covariates are “imbalanced” between the treatment and the control groups. By applying PS weights, selection bias or covariate imbalance is eliminated. To evaluate the success of the PS weighting, two approaches were taken to examine covariate balancing.

The first approach is to examine standard bias statistics. Harder, Stuart, and Antony (2010) calculated standardized bias by dividing the difference in means of each covariate between the treatment and the comparison groups by the standard deviation. They argued that the choice of the standard deviation, whether using that of either one of



Note. Diagonal = City Connects was present ONLY with a health coordinator; vertical dashes = City Connects was present, no School Site Coordinator; black = school was not in operation; white = school was in operation, but no City Connects; and remainder = school had City Connects School Site Coordinator by wave

Figure 2.1. City Connects School Rollout History Timelines in BPS. From "City Connects BPS and SPS: School Rollout History Timelines", by T. Lee-St. John, 2013.

the treatment groups or that of both groups combined, does not matter as long as it is the same one used before and after weighting. In this study, the standard deviation of the City Connects group was used. The decision criterion was that a covariate was considered balanced if the value of the associated standardized bias was less than 0.25. However, Harder and his colleagues (2010) did emphasize that 0.25 is not a strict cut-off and they argued that when multiple PS techniques meet this cut-off, a stricter rule such as 0.10 may be applied.

The second approach is to run a series of weighted regression analyses to examine if there is any statistically significant difference in covariates between the two groups after weighting (Guo & Fraser, 2010). Each covariate was modeled as the outcome. Depending on whether the outcome was a continuous or a binary variable, either weighted linear regression or weighted logistic regression model was built with the treatment variable as the only predictor. Covariate imbalance was indicated if the p value associated with the treatment variable was smaller than 0.05.

2.4.2 Sensitivity Analysis to Measure Hidden Bias

All the statistical adjustments and research designs discussed in the previous sections, if applied appropriately, are able to remove overt bias in observational studies; however, hidden bias, as embodied by unobserved characteristics that are unintentionally omitted from the analytic models, may still remain. A sensitivity analysis was developed to measure sensitivity to hidden bias (i.e. introduced by U) such as “how much hidden bias would need to be present if hidden bias were to explain the differing outcomes in the treated and control groups” (Rosenbaum, 1991a, p. 901).

To start with, Rubin's potential outcomes model (Rubin, 1974) will be introduced for some notation and vocabulary. This model is based on the idea that each student *i* has two potential outcomes: $R_{i,1}$ if enrolled in a treatment group and $R_{i,0}$ if in a comparison group. The causal effect of the treatment for each student is the difference between these two outcomes (i.e., $R_{i,1} - R_{i,0}$); the average causal treatment effect is obtained by taking the average of these differences across all students. However, in reality one can only observe one of these two outcomes for each student. This problem can be treated as a missing data problem: some students are missing $R_{i,1}$ and other students are missing $R_{i,0}$. Although it is no longer possible to estimate individual causal effects, Rubin argued that if assignment to treatment is completely random, the treatment effect can still be estimated as the difference between the means of the treatment and the control groups.

When allocation is conditionally random on *X*, strong ignorability assumption holds, which is expressed as:

$$(R_{i,1}, R_{i,0}) \perp Z \mid X \tag{2.3}$$

where *Z* denotes the treatment assignment indicator variable;

and *X* denotes a set of observed covariates.

It means that assignment to treatment (*Z*) and the two potential outcomes ($R_{i,1}$, $R_{i,0}$) are independent of each other given the set of observed covariates. In other words, *X* includes all covariates that are both used to assign treatments and possibly related to the response. When strong ignorability assumption holds, treatment effects can be estimated without bias. However, if covariates that are related to both *Z* and ($R_{i,1}, R_{i,0}$) are omitted from *X*, this assumption is violated and the estimated causal effects are biased. Suppose that the actual relationship between *Z* and ($R_{i,1}, R_{i,0}$) is expressed as:

$$(R_{i,1}, R_{i,0}) \perp Z \mid (U, X) \qquad (2.4)$$

where U represents an unobserved variable that relates to both Z and $(R_{i,1}, R_{i,0})$.

It indicates that strong ignorability assumption holds when conditioning on both U and X. As a result, the treatment effects can be estimated without bias, given both the set of covariates X and the unobservable covariate U. A type of analysis was developed to measure sensitivity to hidden bias (i.e. introduced by U) such as "how much hidden bias would need to be present if hidden bias were to explain the differing outcomes in the treated and control groups" (Rosenbaum, 1991, p. 901).

There are different approaches to assess sensitivity to hidden bias depending on the types of unobserved covariate and the outcome, the statistical tests being used, as well as the number of treatment groups. For instance, Rosenbaum and Rubin (1983b) proposed a simple technique to estimate the average effect of a treatment (one treatment group and one comparison group) on a binary outcome after adjusting for observed categorical covariates and an unobserved binary covariate U. Using a maximum likelihood estimation procedure, the difference in probabilities of the expected outcome between the two treatment groups (the treatment effect) can be repeatedly estimated by altering assumptions about U. These assumptions include different values of the increase in the log odds of receiving the treatment associated with $U = 1$ rather than with $U = 0$; different values of the increase in the log odds of the expected outcome under one treatment associated with $U = 1$ rather than with $U = 0$; and different proportions of participants with $U = 0$. By examining the resulting different estimated treatment effects, one can infer how extreme the assumptions about the parameters governing U must be in order to meaningfully change the conclusions about the treatment effect.

With multiple control groups and treatment groups, more sophisticated techniques were developed by Rosenbaum (1988 and 1989); since then, efforts had been made to apply sensitivity analysis to permutation tests (Rosenbaum, 1987; Rosenbaum & Krieger, 1990); Rosenbaum also demonstrated how to perform sensitivity analysis in the context of multiple regression and matched case-control studies (1986 and 1991b).

Influenced by Rosenbaum's early work, with modifications suggested by Montgomery, Richards, & Braun (1986), Diaconu (2012) adopted in her dissertation a sensitivity analysis method assuming the existence of a binary unobserved variable that is related to both the binary treatment group assignment and a continuous outcome. The sensitivity analysis conducted in this study followed Diaconu's approach with minor changes to further investigate the impact of selection bias.

Hypothesizing the Unobserved Variable U. To start with, one needs to hypothesize a real but unobserved variable that bears some relationships with both the treatment assignment and the outcome. For the sake of easier interpretation, one would like to require both relationships to be positive: a higher value of U (i.e., $U = 1$) is more likely to be associated with $Z = 1$ (i.e., assignment to the treatment group) than with $Z = 0$ (i.e., assignment to the comparison group); a lower value of U (i.e., $U = 0$) is more likely to be associated with $Z = 0$ (i.e., assignment to the comparison group) than with $Z = 1$ (i.e., assignment to the treatment group); and a higher value of U (i.e., $U = 1$) is also more likely to be associated with higher values of the outcome than is a lower value of U (i.e., $U = 0$).

This is because the estimated positive treatment effect will be inflated if there is a confounding of treatment and selection. By including an unobserved variable that is positively related to the treatment assignment (selection), one will get smaller estimates of the treatment effect. Meanwhile, if this unobserved variable is also positively related to the outcome, it will further shrink the estimated treatment effect.

Given the hypothesized directions of the above relationships, one can simulate sets of U that satisfy the assumptions and include each set of U into the analytic model. Then the difference between the estimated treatment effects with and without one set of the unobserved variable U included in the model will estimate the hidden bias. With the aid of visual display, it becomes evident how varying the strength of the hypothesized relationship between U and Z (or that between U and the outcome) will impact the estimated treatment effect. As a result, how robust estimated treatment effects are to varying degrees of the violation of the strong ignorability assumption can be assessed.

In this study, *parental involvement*, a dummy variable indicating whether or not parents are involved with their children's education, was assumed to be the unmeasured variable U. This variable can be directly collected from a parent questionnaire asking whether or not parents believe they are involved with their children's education or from a teacher survey asking teachers about their impression of each student's parental involvement. Unfortunately, City Connects did not collect any family related data in BPS, so *parental involvement* is indeed unmeasured in this study.

Parental involvement was chosen because literature suggested that it has positive relationships with both school choice and academic achievement. It is necessary to mention that any existing yet unobserved variable that bears the assumed relationships could have been used. Epstein (1995) defined parental involvement as parents taking an active role in partnering with other family members, schools, and communities to form a caring environment for children. She further argued that an involved parent is one who consistently demonstrates good parenting skills, communicates with the school staff, volunteers in the school, helps their children learn at home, makes important school-related decisions for their children, and collaborates with the community. Under this definition, it is reasonable to believe that parents who are involved with their children's education will be those who directly help their children choose schools to attend. Involved parents are more likely to research on schools and know about school resources and services. They will be more likely to choose City Connects schools in which meaningful communications are frequent and a motivated and caring environment is formed. In other words, it is reasonable to believe that parental involvement is correlated with City Connects treatment membership: involved parents will be more likely to choose schools that implement the City Connects intervention on behalf of their children.

Both qualitative and quantitative studies found that parental involvement had a significant impact on academic achievement (Louks, 1992; Aronson, 1996; Columbo, 1995; Fan & Chen, 2001). Particularly in urban secondary schools, the setting similar to the one in this study, a meta-analysis including 52 studies were undertaken to examine the impact of parental involvement on educational outcomes such as standardized test results and GPA (Jeynes, 2007). The results showed that parental involvement had significantly positive effects on all academic achievement measures under study by about .5 to .55 of a standard deviation unit. Therefore, it is also reasonable to assume that parental involvement is correlated with academic outcomes: students with involved parents will be more likely to succeed academically.

General Procedure of Sensitivity Analysis. Random Samples of *U* can be drawn through Monte Carlo simulation, which utilizes random numbers in the simulation algorithm (Kennedy, 2003). The principle behind it is that "the behavior of a statistic in random samples can be assessed by the empirical process of actually drawing lots of random samples and observing this behavior" (Mooney, 1997, p.2). As described by Mooney (1997), Monte Carlo simulation is used to create a pseudo-population which possesses key mathematical properties that make it resemble samples of data drawn from the true population. Then multiple trials are drawn from the pseudo-population to conduct statistical analysis in order to investigate how the procedure behaves across trials. Simply put, for each trial, the treatment effect is estimated through a proposed statistical analysis with the simulated *U* included. Differences between the estimates with and without the inclusion of *U* are the hidden bias. By the aid of visual graphs, one can discern how the estimated treatment effects are robust to the presence of hidden bias when characteristics of *U* vary in magnitude under the specified assumptions.

2.5 Value-Added Models

2.5.1 Introduction of VAMs

The accountability system under NCLB has been widely criticized because it relies heavily on current status measures, which are merely snap-shots of students' academic performance for a given year. Many teachers, principals, researchers, and other educators argued that the status measures are inappropriate for judging education effectiveness, since assignment to different educational entities is non-random. Students and their parents self-select schools due to a variety of reasons such as family socioeconomic status, schools that siblings or friends go to, school locations, appealing new facilities and other resources, rigorous curricula and rich after-school programs, and school or teacher reputations. Teachers are usually assigned non-randomly by the principal to students based on their teaching experience, type of license, and in some cases, the prior academic or behavioral performances of their students. In some cases, schools choose their students as well. For instance, admission to one of the three exam schools in Boston is based entirely on students' academic performance and test scores.

All these factors can have an impact on student academic achievement and thus attributing student status achievement to their current school or teacher effectiveness is problematic. For example, one school may perform extremely well due to the fact that the school is located in an affluent and well-resourced neighborhood where the local community has developed a tradition of valuing education and facilitating learning. Students in this school generally perform well even if the school has offered little to enhance their achievement. By contrast, a school that serves a majority of low-achieving students will often fail to meet AYP, although it may be helping students progress toward the proficiency threshold at a faster rate. It is more likely that the former school will be rewarded and the latter will be punished under NCLB.

In the light of these concerns, more researchers and educators are seeking alternative measures. As a result, it has become popular to judge the effectiveness of an educational entity based on its contributions to student growth in achievement. Often referred as value-added modeling, the philosophy is to "hold schools and teachers accountable for the learning gains of students they serve" (Raudenbush, 2004, p.121). Statistically speaking, "using the longitudinal test scores of students as inputs, a value-added model (VAM) estimates as an output a numeric residual associated with a specific educational intervention" (Briggs, 2008, p.2).

2.5.2 Purposes and Types

VAMs have been proposed to serve four main purposes: "(1) school and teacher improvement, (2) school and teacher accountability, (3) program evaluation, and (4) research" (Briggs, 2008, p.4; McCaffrey & Lockwood, 2008, p.1). Generally speaking, there are four types of VAMs: ordinary linear regression models, random effect models, fixed-effects models, and layered random effect models (OECD, 2008; Tekwe, Carter, Ma, Algina, Lucas, Rush, Ariet, Fisher, & Resnick, 2004).

The linear regression models are simple linear regressions with measures of student prior achievement and contextual factors used for adjustment. One evident advantage of this approach is that simple regression analysis has been widely used and the mechanism is relatively straightforward and easy to explain to audience with limited statistical trainings. It may yield consistent estimates if the covariates included are not correlated with the error term ("the regression assumption"). However, in reality, this correlation may occur if the outcome causes at least one covariate, or if relevant covariates are intentionally or unintentionally omitted from the model, or if the covariates are measured with error. Additionally, this approach does not take into account the nested structure of educational data, so that students within one school are typically more like each other than like those in other schools as they are exposed to the same educational environment and peer influences. Although this problem can be alleviated by adding some contextual factors, not all the similarities between individuals within each school can be observed.

In view of the drawbacks of linear regression models, a multi-level approach has been developed. It not only facilitates taking into account the hierarchical structure of the data but also provides an opportunity to examine factors in all levels of the hierarchy. An application of random effects models is DVAAS (the Dallas Value-Added Accountability System), a two-stage VAM conducted in Dallas, Texas (Webster & Mendro, 1997).

In contrast with random effects models, fixed effects models treat macro-level (e.g., school-level) contributions as fixed parameters. One of the examples of using fixed effects models in estimating school effectiveness is the study conducted by Haegeland and Kirkeboen. They used Norwegian lower and upper secondary school data to examine the relationships between academic achievement, student prior knowledge and different types of SES variables (2008). The results showed that in a contextualized attainment model, which only included SES variables and school identifiers, adding more SES variables had large impacts on school performance indicators; however, if prior achievement was included in the model, the effects of adding more SES variables were limited. The authors concluded that "if one has to make a priority, (more) data on prior attainment should be preferred to (more) data on socioeconomic background" (Haegeland & Kirkeboen, 2008, p. 14).

Both random effects models and fixed effects models have unique features but also suffer some problems. Generally speaking, econometricians may prefer fixed effect models to examine individual fixed effects such as personal and family characteristics on achievement; while education researchers lean toward random effect models to examine school effectiveness (Todd, Wolpin, & Townsend, as cited in Clarke et al., 2010). The econometricians' preference is based on the fact that by modeling the macro-level units as fixed numbers (i.e., schools being included as N-1 dummy variables), a fixed effects model does not make "the random effects assumption", which states that macro-level residuals are independent of all the covariates in the model and the micro-level residuals (Raudenbush & Bryk, 2002; Allison, 2009). The downside is that a fixed effects model does not allow estimating macro-level characteristics and the estimates are unreliable for small schools. By contrast, a random effects model can model macro-level characteristics and get "shrinkage" estimates of random school effects (shrunken residuals), which weight schools based on their sizes (Clarke, Crawford, Steele, & Vignoles, 2010). These shrunken estimators generally lead to lower mean squared error (MSE) for all sample sizes, with more improvement for small ones, thus achieve more precision in estimation. In addition, Clarke and his colleagues argued that "if we have some knowledge about the school selection mechanism and can include measures of these factors in the model as 'controls', then we can also estimate the average treatment effect using the random effects model" (2010, p.13). In addition, Lockwood and McCaffrey (2007) argued that although mixed effect models lead to inconsistent estimates, they do not necessarily produce poor ones. In fact, if applied to longitudinal data with a large number of correlated measurements on each individual, mixed effect models can "provide nearly unbiased estimate even under relatively complex heterogeneity models involving multiple, unobserved individual-specific attributes whose relationship to the observed measurements varies across those measurements" (Lockwood & McCaffrey, 2007, p. 246).

It is worth mentioning that both fixed effects models and random effects models are still linear regression models so that they depend on regression assumptions. In a quasi-experiment, satisfying these assumptions is not guaranteed due to non-randomization.

The last type of VAMs is layered random effects models. Developed by Sanders, EVAAS (Education Value-Added Assessment System) is a multivariate, longitudinal, and mixed effects model for measuring student academic growth based on value-added estimates of teacher effects on student gain scores (Sanders, Saxton, & Horn, 1997). It collects student data in multiple subjects, grades, and years. EVAAS does not adjust for student characteristics and assumes that the teacher effect in one year will transfer intact (i.e., no attenuation) to the next year. Hence, we can keep adding years of data and this is where the name “layered model” came from. Another feature of EVAAS is that it handles missing data easily.

City Connects is a whole-school intervention that builds partnerships between schools, Boston College, and community resources in order to facilitate a positive school climate in which students can thrive. The primary interest of this study was to evaluate the effectiveness of City Connects elementary/middle schools and to estimate the extent to which such effectiveness could be accounted for by other school characteristics. In order to get statistically reasonable estimates for schools, even for small ones, and to have the capacity of estimating school-level characteristics, random effect models were a reasonable choice. Moreover, the relative standing of each elementary/middle City Connects school in terms of student academic performances was examined and compared across years.

2.5.3 Issues with VAMs

Although the concept of VAMs is appealing to educators, researchers, students, parents, and policymakers, its applications suffer a series of technical issues. Reardon and Raudenbush summarized six assumptions of VAMs needed to draw unbiased causal inferences about school effectiveness: 1) manipulability; 2) no interference between units; 3) interval scale metric; 4) homogeneity of effect; 5) strongly ignorable treatment assignment; and 6) functional form (2009, p.18).

First, manipulability means that each student can attend any school and each student has a non-zero probability of attending any school. In reality, this assumption is often violated. For instance, in BPS, student enrollment assignment is dependent on: 1) if schools are located in the zone in which students live; 2) if schools in other zones are within their walk zone; 3) or if schools are citywide K-8 and middle schools that are open to all students (Boston Public Schools, 2013). Additionally, some BPS schools require an interview or assessment to attend. Therefore, students cannot attend any schools they want.

Second, under this context, no interference between units means that there is no peer effect – the assignment of one student to one school is not dependent on the school assignment of other students. If student composition affects instructional practices and curricula and thus affects student learning; or if student composition affects school recruiting students and teachers, then this assumption is violated. The literature on peer effects suggested that this assumption typically does not hold (Hanushek, Kain, Markman, & Rivkin, 2003; Johnson, 2000; Zimmer & Toma, 2000; Zimmerman, 2003).

Third, interval-scale metric assumption requires the observed outcome to be on an interval scale. Unfortunately, state-wide standardized tests such as MCAS do not meet this assumption. Unlike physical quantities such as weights and heights, neither Mathematics nor ELA is a uni-dimensional construct that can be placed on a continuum with equal intervals and this dimensional mix changes from grade to grade. However, many researchers argued that interval scales can be constructed using Item Response Theory (IRT) models (de Ayala, 2009; Hambleton, Swaminathan, & Rogers, 1991; Yen, 1986).

Fourth, homogeneity means that school effect is constant across students who attend the given school. If a teacher or a school caters instructions to a special population of students, then this teacher or school will be more effective to such students than other teachers or schools without such a student composition.

Fifth, strongly ignorable treatment assignment stated that “any unobserved student characteristics that predict the potential outcomes are independent of school assignment once the observed pre-assignment characteristics of that student are taken into account” (Reardon & Raudenbush, 2009, p.8). In other words, if we control for all relevant covariates and exclude all the irrelevant ones, a student’s assignment to a particular school does not depend on his or her potential outcomes in that school. This assumption is hard to achieve and impossible to verify. Last, common support/ functional form assumption requires modeling the function form correctly for students who are not present in a given school.

Researchers have done extensive work to test these six assumptions. Assuming manipulability, no interference between units, and strong ignorable treatment assignment, Reardon and Raudenbush (2009) conducted a simulation study to examine the robustness of interval scale metrics, homogeneity, and function form assumptions. They emphasized the importance of modeling heterogeneous effects of schools when such effects existed. Failing to do so would make the violations to the other two assumptions even worse. More importantly, although the three key assumptions (manipulability, no interference between units, and strong ignorable treatment assignment) were accepted as facts in Reardon and Raudenbush's study, the violations to them might "substantially degrade results" and "the extent to which such violations are influential is a topic for future research" (p. 34).

Ballou (2008) examined the interval-scale metric assumption and challenged the common opinion that IRT scores are interval-scaled variables. He claimed that the interval-scale metric assumption requires "examinees and test items constitute, in the terminology of representational measurement theory, a conjoint structure" (p. ii). Although it is possible to detect strong departure from this hypothesis using statistical procedures, moderate inconsistency are hardly noticeable. Furthermore, even if the hypothesis holds to the norming sample used by test-developers to estimate item parameters and calibrate items, whether or not it holds for the empirical data obtained from end-users of the final instrument, is still questionable. Therefore, Ballou proposed using methods of ordinal data analysis instead. He argued that these methods reply on a weaker assumption that IRT scores are able to rank students, which is a more plausible assumption. However, he did admit that value-added estimates are more sensitive to the choice of ordinal methods than to conventional techniques.

Rothstein (2007) developed falsification tests to check strong ignorability assumption of teacher value-added models. These falsification tests make a simple assumption that “future treatments cannot have causal effects on current outcomes, and models that indicate such effects must be misspecified” (p.2). The results showed that 5th grade teachers had an impact on 5th grade student achievement gains, the magnitude of which was as large as their impact on 4th grade student achievement gains. Therefore, the assumption of strong ignorable teacher assignment was violated. However, Rothstein did not examine this assumption in the context of school value-added models.

In this study, as explained above, the assumptions of manipulability and no interference between units are implausible in the current school system. Item-level responses are required to examine the interval-scale metric assumption. Unfortunately, BPS did not provide such information. Homogeneity and function form assumptions were discussed in details in Reardon and Raudenbush’s 2009 study. Although Rothstein (2007) checked strong ignorability assumption for teacher value-added models, few researchers have examined this assumption for school value-added models, which was one of the focuses in this study. Therefore, due to its strong relevance to causal argument, as well as the availability of corresponding statistical testing techniques, the consequences of violations to the strongly ignorable treatment assignment assumption were assessed, assuming all the other five assumptions held.

Apart from the aforementioned technical issues, using VAMs to estimate causal school and teacher effects involves some conceptual issues. Raudenbush and Willms (1995) defined two conceptually distinctive types of school effects in school evaluation: A Type A effect is “the difference between a child’s actual performance and the performance that would have been expected if that child had attended a typical school" (p.309). It does not differentiate school practices from student composition and socioeconomic context in which the school is located. The Type A effect is usually the primary interest of parents when choosing schools for their children. By contrast, a Type B effect is "the difference between a child's performance in a particular school and the performance that would have been expected if that child had attended a school with identical context but with practice of average effectiveness" (p.310). It isolates the specific school-related practices from school context and includes "administrative leadership, curricular content, utilization of resources, and classroom instruction" (p.310). The Type B effect is the primary interest of district and state administrators when holding schools accountable for their students' academic performances. Raudenbush and Willms (1995) stated that it is possible to produce unbiased Type A effect estimates if assumptions are met; however, drawing causal inferences of Type B effects are much more problematic. This is because Type B effects are not defined and school practices are not differentiated from school context in the current accountability system. Consequently, Rubin, Stuart, and Zanutto claimed that VAMs "should not be seen as estimating causal effects of teachers or schools, but rather as providing descriptive measures" (2003, p.113).

Given all these issues, many researchers stated that VAMs should be used with caution, particularly when being used for school and teacher accountability. The National Research Council report (2010) stressed that solid evidence of reliability and validity of value-added results, which many researchers believe to be absent, is needed for high-stakes purposes. Briggs argued that "VAM residuals should not be the sole basis for high-stakes sanctions and rewards. They should be used in conjunction with direct observations of teacher and school practices" (2008, p. 14). Braun advocated using VAMs for low-stakes school or teacher improvement as "identifying schools that may be underperforming and should be audited to determine whether they are in need of specific kinds of assistance" (2005, p. 15). In this study, VAMs were used to evaluate program effectiveness of City Connects with the intention of informing improvement strategies, which is an appropriate use of VAMs.

CHAPTER 3. RESEARCH DESIGN

3.1 Research Questions

3.1.1 Research Question One

The first set of research questions concerns the impact of the City Connects elementary intervention on middle school achievement as measured by standardized MCAS scores and criterion-referenced GPA grades. Three sub questions are asked: first, after controlling for student characteristics and pre-existing academic achievement differences, does receiving the City Connects intervention in elementary school help students succeed in middle school and does this success persist through the entire middle school (denoted as RQ1a)? Second, do students who graduate from different City Connects elementary schools generate significantly different academic achievement patterns in middle school (denoted as RQ1b)? In other words, the author was interested in if the program impact varied among the participating elementary schools to produce significantly different academic results in middle school. Third, if City Connects effects are observed, to what extent can this be accounted for by both student and other school characteristics (denoted as RQ1c)?

The outcomes of interest were MCAS ELA and Mathematics scores and annual GPA grades in middle school. All the MCAS raw scores were converted into z scores by subject, by grade, and by school year using the means and standard deviations of the comparison group. Course letter grades were converted into a 0 to 4 scale with one more point added if the course was an honors or Advanced Placement (in Art or Music) course. The average GPA across all the courses that a student had taken during a given year was the second outcome of interest.

Chapter 3. Research Design

After matching students on baseline achievement and demographic characteristics[3], adjusting for differential middle school effectiveness, as well as taking into account the elementary school clusters (the most recent City Connects or comparison elementary schools that students attended), the overall treatment effect of the City Connects elementary intervention on student academic achievement in middle school was estimated. Then the relative standing of each City Connects elementary school in terms of such academic achievement was examined. This analysis was done separately for Grades 6, 7, and 8 so that general trends could be discerned. Finally, school-level covariates were added to the model to explain school differences.

The analytic sample included students who reached at least Grade 6 by the 2012-2013 academic year. The term “cohort year” was defined as the year when the students entered kindergarten. For instance, a cohort 2001 student was one who entered kindergarten during the 2001-2002 academic year. Therefore, this analysis consisted of students from cohort years 2000 to 2006. The current master file traced student records back to the 2001-2002 academic year. To be eligible for the analytic file, a student must have some baseline achievement measures. Cohort 2000 was the oldest cohort that could have had achievement scores in Grade 1 by the 2001-2002 academic year. Moreover, cohort 2006 was the youngest cohort that could have reached Grade 6 by the 2012-2013 academic year.

[3] Baseline achievement includes Grade 1 fall Report Card scores in Reading, Mathematics, Writing, Behavior, Work Habits, and Effort; while baseline demographic characteristics include gender, race, bilingual status, special education status, reduced or free lunch status, foreign born status, age when starting Grade 1, the number of school moves when starting Grade 1, and home distance to school in miles when starting Grade 1.

For the purpose of examining trends over time for each City Connects elementary school, student records were followed over the three years of middle school. Therefore, a second restriction was applied: only students who not only attended BPS elementary schools for at least one term but also attended Grade 6 in BPS were included in the analytic sample. To be specific, all students that reached Grade 6 in BPS by the 2012-2013 academic year were included in the Grade 6 models; among these students, those who continued enrolling in BPS middle schools and never switched schools were included in the Grade 7 models; and those who stayed in BPS and never switched schools for the entire three years of middle school were included in the Grade 8 models. The third restriction was that students in Special Education categories 4 and 5 (need substantially separate education or out of school or home program) were excluded. This is because the pedagogical treatments of the reported grades for students with severe special needs were so different from those of other students, they should be analyzed separately.

To accurately estimate City Connects treatment effects, it is important to take into account the cross-classified nested structure of the data: students attended different elementary schools, some of which were implementing the City Connects intervention; these students then progressed to different middle schools. As shown in Figure 3.1a, suppose students are enrolled in either City Connects (school Aor B) or comparison elementary schools (school C, D, or E), and all City Connects graduates go to one middle school (school X) and all comparison students go to another (school Y). In other words, students move to middle school in elementary school units; then it is reasonable to build a 3-level hierarchical model which takes into account students clustering in elementary schools and elementary schools clustering in middle schools. The logic is the same as

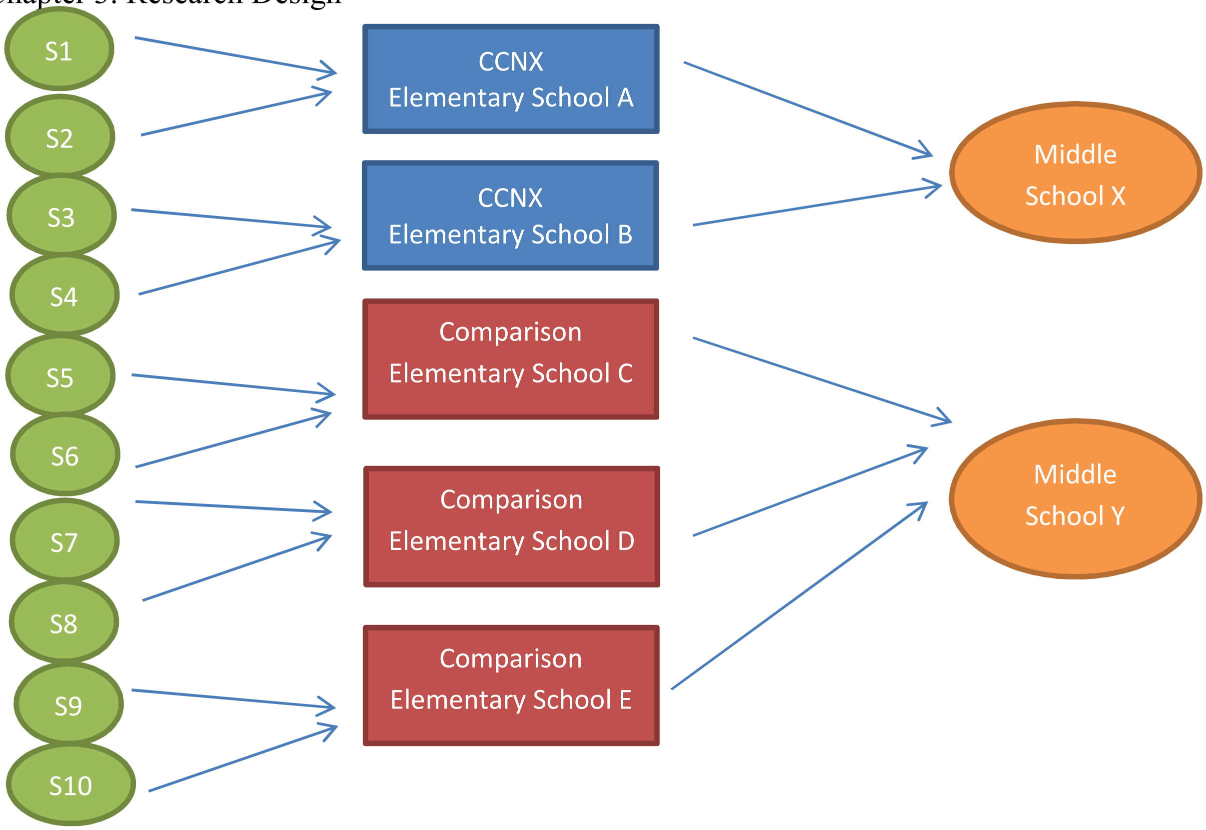


Figure 3.1a. Students' Progressing from Elementary Schools to Middle Schools (Ideal Sc

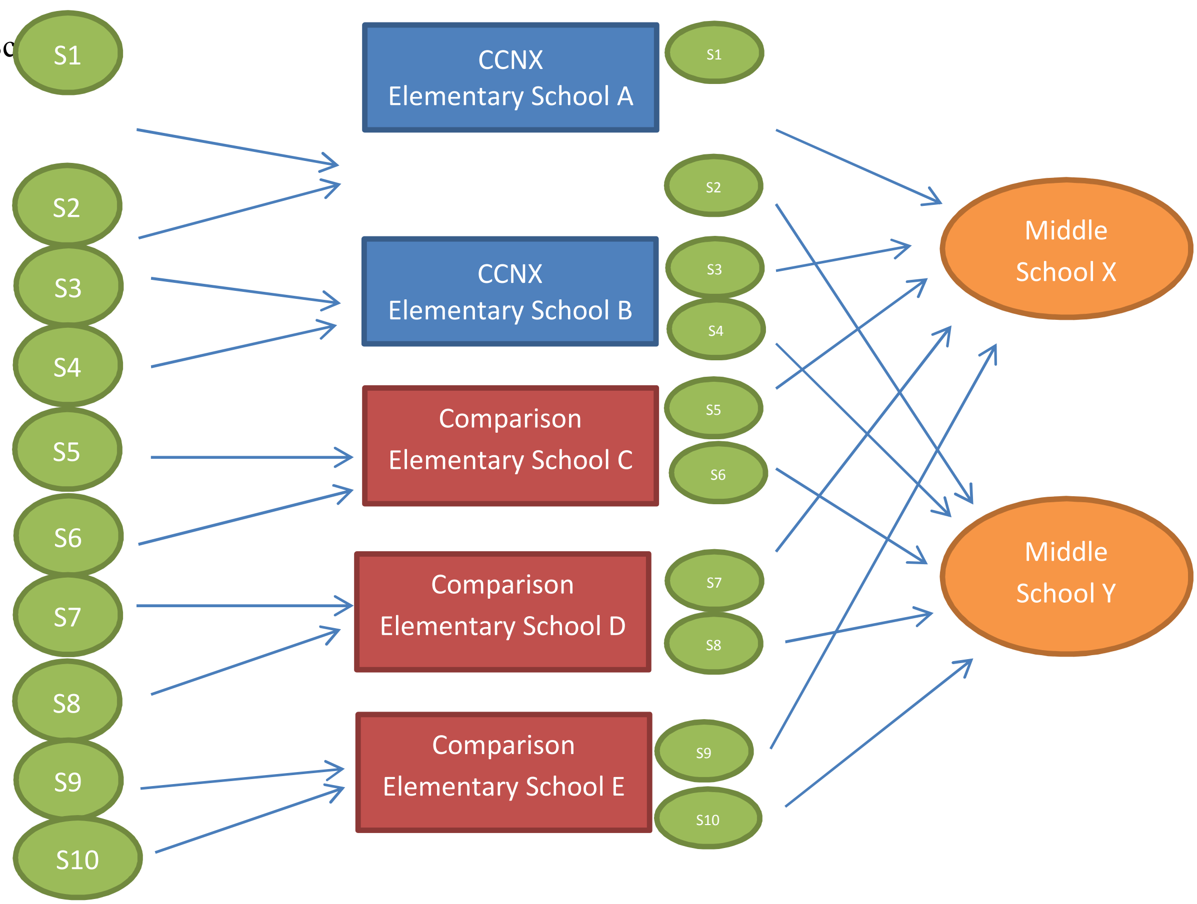


Figure 3.1b. Students' Progressing from Elementary Schools to Middle Schools (Actual Scenario)

building a typical 3-level model with students clustering in classrooms and classrooms clustering in schools. In this scenario, City Connects treatment effects will be estimated at the elementary school level (level 2).

However, in reality, students progress to middle schools as individuals. As shown in Figure 3.1b, both students 1 and 2 go to elementary school A; after graduation, student 1 goes to middle school X and student 2 goes to middle school Y. It is possible for both middle school X and Y to enroll students from all the five elementary schools. In this more realistic scenario, the typical 3-level model is infeasible because this cross-classified structure will make it difficult to partition variance components between level 2 and 3.

To avoid this complication, past City Connects middle school research simply ignored the middle school clustering (the third level) and built a two-level model with students clustering in most recent elementary schools. A major problem of this approach is that if City Connects is effective in boosting student achievement, then City Connects attendees may be more likely to go to a better middle school. Assuming middle school X is intrinsically more effective than middle school Y, then middle school X may recruit a proportionately larger population of City Connects students. Since the middle school clustering has not been taken into account, the estimated treatment effect is confounded with middle school effectiveness, leading to erroneous conclusions.

Model 1a. In order to adjust for differential middle school effectiveness, a two-phase analysis was conducted for each subject and in each grade to answer RQ1a. In the first phase, using comparison students only, a two-level model, which took into account middle school clusters, was built to predict the outcomes of interest. A reasonable concern is that if there is a positive City Connects elementary treatment effect on academic achievement, consequently graduates of City Connects elementary schools will be more likely to attend better middle schools. As a result, adjusting for middle school effectiveness using both the treatment and the comparison students will remove this positive effect undiscriminatingly, if it exists, and thus underestimate the City Connects elementary treatment effect in the outcome models. In other words, it is believed that City Connects elementary schools should take credits for sending their graduates to better middle schools and this effect should be reserved when adjusting middle school effectiveness in an attempt to differentiate the contribution of elementary schools on achievement from that of middle schools. To address this concern, only comparison students were used to estimate middle school effectiveness.

The model included prior achievement (Grade 5 MCAS ELA and Mathematics scores) and time-varying background variables in the outcome grade (the number of school moves, home distance to school to school in miles, and age), together with time-invariant student characteristics (gender, race, bilingual status, special education status, free or reduced lunch status, and foreign born status). They were denoted as *p* student-level covariates in Equation 3.1. $\hat{\beta}_{0j}$, the estimated adjusted mean of middle school *j*, was saved as middle school achievement adjustment score (denoted as Z^j).

Level 1 (Student Level):

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \dots + \beta_{pj} X_{pij} + r_{ij}$$

Level 2 (School Level):

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

...

$\beta_{pj} = \gamma_{p0}$ (3.1)

where i denotes students within middle schools, and j denotes middle schools;

Y_{ij} is the academic outcome measure (MCAS ELA, Mathematics, or annual GPA) for student i in middle school j in one of the middle school grades;

X_{1ij} to X_{pij} are p student-level covariates for student i in middle school j;

β_{0j} is the mean of the outcome measure for middle school j, adjusted for p covariates (X_{1ij} to X_{pij});

β_{1j} to β_{pj} are the regression coefficients for middle school j, associated with p covariates (X_{1ij} to X_{pij});

r_{ij} is the random error (or residual) at level 1, where $r_{ij} \sim N(0, \sigma^2)$ and σ^2 is the variance of the student-level residuals;

γ_{00} is the intercept at level 2, which is the grand mean of the adjusted means across all middle schools;

γ_{10} to γ_{p0} are constants indicating the means of the p regression coefficients across all middle schools;

and u_{0j} is the random error (or residual) at level 2, where $u_{0j} \sim N(0, \tau_0)$ and τ_0 is the variance of the school-level residuals for $\{\beta_{0j}\}_j$.

In the second phase, both City Connects and comparison students were used. The second phase included two stages. In the first stage, PS weights were estimated for each case. Referred as "the selection model", the probability of being categorized as a City Connects student (Dose) was estimated through a binary logistic regression. The selection model included baseline achievement (e.g., student Grade 1 fall Report Card scores in

Reading, Mathematics, Writing, Behavior, Work Habits, and Effort) and baseline demographic characteristics (e.g., gender, race, bilingual status, special education status, reduced or free lunch status, foreign born status, the number of school moves when starting Grade 1, home distance to school in miles when starting Grade 1, and age when starting Grade 1. They were denoted as q student-level covariates. For a City Connects student, the PS weight was the inverse of the predicted probability of getting into the treatment group; while for a comparison student, the weight was the inverse of one minus that predicted probability (Guo & Fraser, 2010).

In the second stage, referred as “the outcome model”, a two-level linear regression model was built for each subject in each outcome grade. The PS weights generated in the first stage were applied as the level-1 weights. The clustering variable was the most recent City Connects or comparison elementary schools that students attended. In addition, the same set of q covariates was included at the student level of the outcome model to control for pre-existing differences between the two groups even after the application of PS weights. The estimated middle school achievement adjustment score (Z^j) was subtracted from outcome scores to control for differential middle school effectiveness. A series of dummy variables indicating years spent with City Connects (City Connects dosage) were placed at level 1 (City Connects students with maximum years of dosage will serve as the reference group); and the City Connects treatment effect (the effect of City Connects Dose) was estimated at level 2. The full model was represented by Equation 3.2. The estimated treatment effect, $\hat{\gamma}_{01}$, was the answer to RQ1a.

Chapter 3. Research Design

Level 1 (Student Level):

$$Y_{ik}^{j} - Z_{ik}^{j} = \beta_{0k} + \beta_{1k}X_{1ik}^{j} \ldots + \beta_{qk}X_{qik}^{j} + \sum_{m=1}^{5}\theta_{mk}Dosage_{mik}^{j} + r_{ik}^{j}$$

Level 2 (School Level):

$$\beta_{0k} = \gamma_{00} + \gamma_{01}\, EDose_{k} + u_{0k}$$

$$\beta_{1k} = \gamma_{10} + u_{1k}$$

…

$$\beta_{qk} = \gamma_{q0} + u_{qk}$$

$$\sum_{m=1}^{5}\theta_{mk} = \sum_{m=1}^{5}(\varphi_{m0} + \psi_{mk}) \qquad (3.2)$$

where i denotes students within elementary schools, k denotes last elementary school attended, and j denotes middle schools;

Y_{ik}^{j} is the academic outcome measure (MCAS ELA, Mathematics, or annual GPA) in one of the middle school grades for student i in last elementary school k who then went to middle school j;

Z_{ik}^{j} is achievement adjustment score for middle school j attended by student i in elementary school k;

X_{1ik}^{j} to X_{qik}^{j} are q student-level covariates for student i in last elementary school k who then went to middle school j;

$\sum_{m=1}^{5} Dosage_{mik}^{j}$ represent a series of dummy variables indicating the number of years spent in City Connects elementary schools, where m = 1, 2, … 5[4];

[4] Since CCNX serves kindergarten to fifth grades in elementary school, the maximum years of the elementary CCNX one can receive is six years. Using students who received six years of CCNX as the reference group, the number of dosage dummy variables in the equation will be 6 – 1 = 5.

β_{0k} is the mean of the outcome measure for last elementary school k, adjusted for middle school achievement (Z_{ik}^{j}), q covariates (X_{1ik}^{j} to X_{qik}^{j}), and $\sum_{m=1}^{5} Dosage_{mik}^{j}$;

β_{1k} to β_{qk} are the regression coefficients for last elementary school k, associated with q covariates (X_{1ik}^{j} to X_{qik}^{j});

$\sum_{m=1}^{5} \theta_{mk}$ are regression coefficients for last elementary school k, associated with $\sum_{m=1}^{5} Dosage_{mik}^{j}$;

r_{ik}^{j} is the random error (or residual) at level 1, where $r_{ik}^{j} \sim \mathrm{N}(0, \sigma^2)$ and σ^2 is the variance of the student-level residuals;

$EDose_k$ is a dummy variable indicting treatment membership in elementary school, with 1 for treatment schools, and 0 for comparison schools;

γ_{01} is the estimated treatment effect;

γ_{00} is the intercept at level 2, which is the adjusted mean achievement for comparison elementary schools (i.e., when $EDose_k$ =0);

γ_{10} to γ_{q0} are constants indicating the means of the q regression coefficients across all last elementary schools;

$\sum_{m=1}^{5} \varphi_{m0}$ are constants indicating the mean values of $\sum_{m=1}^{5} \theta_{mk}$ across all last elementary schools;

u_{0k} to u_{qk} are random effects at level 2, where $u_{vk} \sim N(0, \tau_v)$ ($v = 0, 1, \ldots, q$) and τ_v is the variance of the school-level residuals for $\{\beta_{vk}\}_k$ ($v = 0, 1, \ldots, q$);;

and $\sum_{m=1}^{5} \psi_{mk}$ are random effects at level 2, where $\psi_{mk} \sim N(0, \tau'_m)$ ($m = 1, 2, \ldots, 5$).

The analytic sample included a total of 11051 students (1791 City Connects students and 9260 comparison students) with a complete set of baseline achievement and time-invariant demographic characteristics. They were enrolled in 15 City Connects and 78 comparison elementary schools and then went to one of 36 middle schools.

Model1b. To evaluate school effectiveness within the City Connects group when addressing RQ1b, the residual-based estimates ($\{\hat{u}_{0k}\}$) from Stage 2 of Model 1a were compared among City Connects schools and examined across the three grades to discern general trends. Furthermore, some measures of dispersion (i.e., standard deviation) were used to summarize $\{\hat{u}_{0k}\}$ of City Connects elementary schools to understand how much these schools varied from one another. The magnitude of such measures was compared with that of the estimated treatment effect ($\hat{\gamma}_{01}$) to demonstrate how City Connects elementary schools performed differently in terms of middle school academic achievement as compared to comparison schools. In addition, residual index statistics was calculated as a ratio between the average of $\{\hat{\delta}_k^2\}$[5] and the variance of $\{\hat{u}_{0k}\}$. The smaller the ratio, the better the model fit was.

Model 1c. To answer RQ1c, some school-level covariates (e.g., whether or not the school is a K-8 school, student /teacher ratio, school size, average class size, students per computer, and percentages of minority and low-income students) were added at level 2 to explain school differences (see Equation 3.3).

Level 1 (Student Level):

$$Y_{ik}^{j} - Z_{ik}^{j} = \beta_{0k} + \beta_{1k}X_{1ik}^{j} \ldots + \beta_{qk}X_{qik}^{j} + + \sum_{m=1}^{5} \theta_{mk} Dosage_{mik}^{j} + r_{ik}^{j}$$

Level 2 (School Level):

$$\beta_{0k} = \gamma_{00} + \gamma_{01}\, EDose_{k} + \gamma_{02}\, W_{2k} + \ldots + \gamma_{0s} W_{sk} + u_{0k}$$

---

[5] $\delta_k^2 = \frac{\sigma_k^2}{n_k}$, where σ_k^2 is the variance of the within-school residuals (level-1 residuals) for each elementary school *k* and n_k represents the number of students in each elementary school *k*.

$\beta_{1k} = \gamma_{10} + u_{1k}$

…

$\beta_{qk} = \gamma_{q0} + u_{qk}$

$$\sum_{m=1}^{5} \theta_{mk} = \sum_{m=1}^{5} (\varphi_{m0} + \psi_{mk}) \tag{3.3}$$

where W_{2k} to W_{sk} are s-1 school-level covariates for last elementary school k;

γ_{02} *to* γ_{0s} are the regression coefficients associated with school-level covariates W_{2k} to W_{sk};

γ_{00} is the intercept at level 2, which is the mean achievement for comparison elementary schools (i.e., when $EDose_k$ =0), adjusted for school-level covariates.

Note that the model presented here incorporates all available covariates and allows both the intercept and the slopes to vary randomly. The final model was much simpler based on model testing results. The q student-level covariates served as the standard set of demographic control that City Connects had been using for years, so they were kept intact in the model. The s -1 school-level covariates were subject to testing. They were added into the model one by one to predict the intercept; if insignificant, they were removed. Then the relationships between all the level-1 covariates and the outcome of interest were examined across institutions to determine whether or not the corresponding level-1 slopes should be fixed or allowed to vary: if there was no significant variation in the level-1 slopes across institutions, the level-1 slopes were fixed. Otherwise, they were allowed to vary.

Model 1a versus Model 1c. Given the quasi-experimental nature of the City Connects evaluation design, the City Connects effect may be confounded with student characteristics and pre-existing academic differences. Therefore, it is essential to statistically control for such factors at level 1 (as proposed in Model 1a). In addition, one may argue that certain school characteristics that contribute to academic success may be found more/less frequently in a City Connects school than in a comparison school. Then the apparent academic success of City Connects schools may be a result of these other school features rather than the strategies and practices of the intervention itself. That is why Model 1c was proposed to address the potential bias in the estimation of the treatment effect.

However, one may also argue that the treatment effect will be underestimated when these school characteristics are controlled for. The rationale of the City Connects intervention is that by sending SSC to schools to directly interact with students, parents, teachers, school staff, and local communities on a daily basis, a trustworthy, motivated and effective environment is created in which every single student can thrive. It is reasonable to believe that when being placed in such a positive environment one will try every means and bring in all possible resources (i.e. smaller class sizes, more computers) to help students improve. By controlling for these school characteristics one may mistakenly control for the very features/byproducts that make the intervention successful. On the contrary, one may also argue that most covariates at the school level are not malleable. Including these school characteristics could yield a larger estimate of the treatment effect.

Nevertheless, both models provide unique information yet serve different purposes. Model 1a is of primary interest to parents, teachers, and various stakeholders to address the effectiveness of the intervention; while Model 1c is of primary interest to policymakers in formulating policies to improve student achievement.

3.1.2 Research Question Two

Among the schools that received the City Connects intervention, three K-8 schools (SchoolA, SchoolB, and SchoolE) and two secondary schools (SchoolC and SchoolD) received City Connects in middle school grades. Some schools had received the City Connects intervention in all grades since they started working with City Connects (i.e., SchoolA and SchoolB); some schools had the City Connects elementary intervention for a long time and gradually extended to middle school grades (i.e., SchoolE); and others started freshly with the secondary intervention (i.e., SchoolC and SchoolD). Table 3.1 shows the history of City Connects serving Grade 6. As we can see, SchoolB had the longest history with City Connects in Grade 6 because it started in 2002, followed by SchoolA and SchoolC. SchoolD and SchoolE just started the intervention in 2012.

Table 3.1

Chronology of City Connects Schools that Served Grade 6

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
SchoolA												
SchoolB												
SchoolC												
SchoolD												
SchoolE												

These schools present a unique challenge in examining the effectiveness of the City Connects intervention on academic outcomes in middle school, due to the complications they bring to the clustering variable and to the estimation of the treatment effect.

If the clustering variable is the last City Connects/comparison schools that students attended, then the City Connects school clusters will be a mix of elementary and middle school ones, which makes it severely problematic to find corresponding comparison school clusters. Furthermore, within the City Connects group, a person may attend one of the City Connects elementary schools for a full term (from Kindergarten to Grade 5) and then attend one of the City Connects middle schools. In this case, to attribute this student's academic gains by the end of Grade 6 to the City Connects middle school that he or she attended (which he or she only attended for one year) is inappropriate.

Past City Connects analysis (prior to 2012) on middle school outcomes dealt with this problem by attributing the academic gains of such students to the elementary schools that they attended. Although inappropriate, due to small sample sizes (mostly of these students were SchoolA and SchoolB attendees and these two schools were known for small student enrollments) and the primary research focus back then on the average treatment effect, the results were not biased substantially. However, with the addition of SchoolC, SchoolD, and SchoolE in City Connects during the most recent years (SchoolC and SchoolD were large schools), together with the new interest in individual school effectiveness, it was necessary to tease out these students from RQ1 and develop a separate analysis for them.

The second set of research questions, then, address the impact of the City Connects middle school intervention on middle school achievement: First, after controlling for student characteristics and pre-existing academic achievement differences, are City Connects middle schools more effective in improving students' academic performances than other middle schools that are not receiving City Connects (referred as "non-City Connects middle schools") in Grade 6 (RQ2a)? Second, do students enrolled in different City Connects middle schools generate significantly different academic achievement patterns (RQ2b)? Third, to what extent are the observed City Connects middle school effects accounted for by both student and other school characteristics (RQ2c)? The outcomes of interest were the same as those for the set of RQ1.

The treatment group included students who were enrolled in one of the five City Connects schools during the years when these schools received the City Connects intervention in Grade 6. The comparison group included students who were enrolled in K-8 schools that were not receiving the City Connects middle school intervention over the same years. The five City Connects schools enrolled a total of 1037 students with a complete set of covariates and prior achievement scores. Among them, approximately 900 had outcome scores in Grade 6. Another 21 K-8 schools did not receive the City Connects intervention in Grade 6. These schools all together enrolled 7135 students (approximately 6000 of them had Grade 6 outcome scores).

This analysis was only conducted in Grade 6 because 1) people might attend one of the three exam schools in Grade 7 or higher. Based on past City Connects evaluation results, City Connects students were more likely to go to exam schools (An, Lee-St. John, Raczek, Walsh, & Madaus, 2014). Therefore, the City Connects sample in Grade 7 contained newly-added seventh/eighth graders and arguably lower-achieving City Connects students who did not qualify for exam schools; 2) the analytic sample sizes for the City Connects schools were smaller than 500 in Grade 7 and 8 because students must have both prior achievement data and the outcome.

Model 2a. In terms of the statistical procedure, the difference between Model 1a and Model 2a was that the latter did not need Phase 1. For each subject, the 2-stage procedure in Phase 2 was implemented: first, PS weights were generated using prior achievement (Grade 5 MCAS ELA and Mathematics scores) and a set of baseline demographic characteristics in Grade 5 (denoted as *t* set of student-level covariates). Second, a two-level hierarchical linear model which took into account the middle school clusters was built and PS weights were applied at level 1. The same *t* set of student-level covariates was included in the outcome models to adjust for pre-existing differences among students.

The City Connects middle school treatment effect was estimated at level 2 ($\hat{\gamma}_{01}$). Since it was possible for a non-City Connects middle school to receive students that had experienced the City Connects intervention in elementary school, a dummy variable indicating whether or not a student was a K-5 City Connects student was added at the student level. The general statistical form of the model is expressed in Equation 3.4.

Level 1 (Student Level):

$$Y_{ij} = \beta_{oj} + \beta_{1j} X_{1ij} + \ldots + \beta_{tj} X_{tij} + \beta_{(t+1)j} EDose_{ij} + r_{ij}$$

Level 2 (School Level):

$$\beta_{oj} = \gamma_{00} + \gamma_{01} MDose_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

…

$$\beta_{(t+1)j} = \gamma_{(t+1)0} + u_{(t+1)j} \tag{3.4}$$

where *i* denotes students within middle schools, and *j* denotes middle schools;

Y_{ij} is the academic outcome measure (MCAS ELA, math, or annual GPA) in Grade 6 for student i in middle school j;

X_{1ij} to X_{tij} are t student-level covariates for student i in middle school j;

$EDose_{ij}$ indicates whether or not a student had received the City Connects intervention in elementary school, with 1 for those who had and 0 for those who had not;

β_{oj} is the mean of the outcome measure for middle school j , adjusted for $t + 1$ covariates (X_{1ij} to $EDose_{ij}$);

β_{1j} to $\beta_{(t+1)j}$ are the regression coefficients for middle school j, associated with $t + 1$ covariates (X_{1ij} to $EDose_{ij}$);

r_{ij} is the random error (or residual) at level 1, where $r_{ij} \sim N(0, \sigma^2)$ and σ^2 is the variance of the student-level residuals;

$MDose_j$ is a dummy variable indicting middle school treatment membership, with 1 for treatment middle schools, and 0 for comparison middle schools;

γ_{01} is the estimated middle school treatment effect;

γ_{00} is the intercept at level 2, which is the adjusted mean achievement for comparison middle schools (i.e., when $MDose_j$ =0);

γ_{10} to $\gamma_{(t+1)0}$ are constants indicating the means of the $t + 1$ regression coefficients across all middle schools;

and u_{0j} to $u_{(t+1)j}$ are random effects at level 2, where $u_{(t+1)j} \sim N(0, \tau_v)$ ($v = 0, 1, \ldots, t+1$) and τ_v is the variance of the school-level residuals for $\{\beta_{vj}\}_j$ ($v = 0, 1, \ldots, t+1$).

In this model, City Connects dosage was not included because the City Connects group consisted of students who received City Connects since Kindergarten to Grade 6 and those who received City Connects only in Grade 6. For the former, the range of possible dosage was 1 to 7; while the latter group could only have a dosage of 1. Nevertheless, one of the concerns was that the one year dosage the latter group received was not the same as the one year dosage received by the former group. Therefore, given the concern that City Connects might have differential dosage effects among City Connects recipients, City Connects dosage was not used in the model as a treatment indicator.

Model 2b. Middle school effectiveness within the City Connects group was evaluated as in Model 1b.

Model 2c. School-level covariates that might account for school differences were included at level 2 of Model 2a. Resulting estimated treatment effects with and without school-level covariates were compared.

3.1.3 Research Question Three

The third research question (RQ3) asks whether or not the estimated treatment effects obtained from the first two analyses are robust to the violation of the ignorable treatment assignment assumption, the key assumption in making causal inferences. To provide a partial answer, a sensitivity analysis was conducted. As discussed in Chapter 2, *parental involvement* served as the real but unobserved variable U that represented hidden bias in this study.

To start with, two assumptions were made to define the conditions of the simulated U in this study. The first assumption dealt with how U was related to Z. Table 3.2 shows the conditional probabilities of *parental involvement* given the treatment assignment: the conditional probability of U taking any value of u given Z taking any value of z is expressed mathematically as $\Pr(U = u \mid Z = z)$, where $u = 1$ indicating high *parental involvement* and $u = 0$ indicating low *parental involvement* and $z = 1$ for the treatment group and $z = 0$ for the comparison group. As shown in the table, the conditional probability of $U = 0$ given $Z = 0$ is denoted as $\pi_{0|0}$; the conditional probability of $U = 1$ given $Z = 0$ is denoted as $\pi_{1|0}$; the conditional probability of $U = 0$ given $Z = 1$ is denoted as $\pi_{0|1}$; and the conditional probability of $U = 1$ given $Z = 1$ is denoted as $\pi_{1|1}$.

Table 3.2

The Conditional Probabilities of U Given Z

Conditional probability (π)		**Z**	
		0	*1*
U	*0*	$\pi_{0\|0}$	$\pi_{0\|1}$
	1	$\pi_{1\|0}$	$\pi_{1\|1}$

Note. From "Modeling Science Achievement Differences between Single-sex and Coeducational Schools: Analysis from Hong Kong, SAR and New Zealand from TIMSS 1995, 1999, and 2003", by D. V. Diaconu, 2012, (Doctoral dissertation), *ProQuest Dissertations and Theses*, (Accession Order No. [UMI3521765]), p.108. Copyright 2012 by Dana V. Diaconu. Reprinted with permission.

As discussed before, a reasonable hypothesized relationship between U (parental involvement) and Z (the treatment membership) is: first, higher parental involvement is associated with higher probability of attending City Connects schools than comparison schools. It is expressed mathematically as

$$\Pr(U = 1 \mid Z = 1) > \Pr(U = 1 \mid Z = 0)$$

or:

$$\pi_{1|1} > \pi_{1|0} \tag{3.5a}$$

Second, lower parental involvement is associated with higher probability of attending comparison schools than City Connects schools:

$$\Pr(U = 0 \mid Z = 0) > \Pr(U = 0 \mid Z = 1)$$

or:

$$\pi_{0|0} > \pi_{0|1} \tag{3.5b}$$

All these four parameters ($\pi_{1|1}$, $\pi_{1|0}$, $\pi_{0|0}$, and $\pi_{0|1}$) can take on any values between 0 to 1 (because they are probabilities so the range is 0 to 1) as long as they meet the requirement of Assumption 1. Large $\pi_{1|1} - \pi_{1|0}$ or $\pi_{0|0} - \pi_{0|1}$ corresponds to the existence of a strong selection bias since *U* is strongly associated with *Z*. On the contrary, small $\pi_{1|1} - \pi_{1|0}$ or $\pi_{0|0} - \pi_{0|1}$ indicates that the problem of selection bias may not be so severe.

Note that the conditional probability of *U* taking one value of *u* is dependent on the conditional probability of *U* taking the other value of *u* given *Z* taking the same value of *z*:

$$\Pr(U = 0 \mid Z = 0) + \Pr(U = 1 \mid Z = 0) = 1$$

or:

$$\pi_{0|0} + \pi_{1|0} = 1 \tag{3.6a}$$

AND

$$\Pr(U = 0 \mid Z = 1) + \Pr(U = 1 \mid Z = 1) = 1$$

or:

$$\pi_{0|1} + \pi_{1|1} = 1, \tag{3.6b}$$

Therefore, one can just focus on the relationship represented by one of the two inequalities ($\pi_{1|1} > \pi_{1|0}$ or $\pi_{0|0} > \pi_{0|1}$) . Assumption 1 is simplified as $\pi_{1|1} > \pi_{1|0}$.

The second assumption deals with how *U* is related to the outcome of interest. A reasonable hypothesized relationship between parental involvement and academic achievement is that they are positively related. In other words, the regression coefficient associated with *U* should be positive. Following Rosenbaum's approach (1986), values were set based on empirical results obtained from Model 1a and Model 2a. Possible values of the regression coefficient associated with *U* were: 1) the largest regression coefficients associated with student-level demographic covariates in the outcome models; and 2) the ones associated with prior or baseline achievement adjustments.

In this study, the key mathematical properties of *U* are the two assumptions made about *U*. As discussed above, the first assumption is that $\pi_{1|1} > \pi_{1|0}$. Following Diaconu's approach (2012), the range of $\pi_{1|0}$ and $\pi_{1|1}$ was set to be 0.2 to 0.8 with 0.15 as a basic incremental unit. To satisfy $\pi_{1|0} < \pi_{1|1}$, possible values of $\pi_{1|0}$ and $\pi_{1|1}$ are listed in Table 3.3. The first step of the sensitivity analysis for this study was to simulate *U* (u_1 to u_{10}) for each pair of the conditional probabilities of *U* given *Z* ($\pi_{1|0}$ and $\pi_{1|1}$).

Table 3.3

The Conditional Probabilities Selected in the Simulation of U

U	$\pi_{1\|0}$	$\pi_{1\|1}$
u_1	0.20	0.35
u_2	0.20	0.50
u_3	0.20	0.65
u_4	0.20	0.80
u_5	0.35	0.50
u_6	0.35	0.65
u_7	0.35	0.80
u_8	0.50	0.65
u_9	0.50	0.80
u_{10}	0.65	0.80

Note. From "Modeling Science Achievement Differences between Single-sex and Coeducational Schools: Analysis from Hong Kong, SAR and New Zealand from TIMSS 1995, 1999, and 2003", by D. V. Diaconu, 2012, (Doctoral dissertation), *ProQuest Dissertations and Theses*, (Accession Order No. [UMI3521765]), p.114.

The second step was to estimate the treatment effects with U included. For illustrative purpose, the model addressing RQ2a (as presented by Equation 4) was used as an example. Equation 9 is the same as Equation 4 except that it has one set of the newly-simulated U_{ij} and a pre-determined regression coefficient β_{Uj} adjusted to the outcome. The estimated treatment effect with U included is $\hat{\acute{\gamma}}_{01}$. The difference between $\hat{\acute{\gamma}}_{01}$(obtained from Equation 5) and $\hat{\gamma}_{01}$(obtained from Equation 4) is the hidden bias to the treatment estimate if U is not included in the model.

Level 1 (Student Level):

$Y_{ij} - \beta_{Uj}\, U_{ij} = \acute{\beta}_{0j} + \acute{\beta}_{1j}\, X_{1ij} + \ldots + \acute{\beta}_{tj}\, X_{tij} + \acute{\beta}_{(t+1)j} EDose_{ij} + \acute{r}_{ij}$

Level 2 (School Level):

$\acute{\beta}_{0j} = \acute{\gamma}_{00} + \acute{\gamma}_{01} MDose_{j} + \acute{u}_{0j}$

$\acute{\beta}_{1j} = \acute{\gamma}_{10} + \acute{u}_{1j}$

...

$$\acute{\beta}_{(t+1)j} = \acute{\gamma}_{(t+1)0} + \acute{u}_{(t+1)j} \tag{3.7}$$

where U_{ij} is one set of the simulated values for the unobserved variable U for student i in middle school j;

and β_{Uj} is one of the pre-determined regression coefficients associated with U_{ij} .

For each of the 10 pairs of conditional probabilities $\pi_{1|0}$ and $\pi_{1|1}$, the above two steps were repeated 100 times and an average estimated treatment effect across the 100 trials was calculated. Then the average treatment effects were plotted against pairs of $\pi_{1|0}$ and $\pi_{1|1}$ in a 3-dimensinal response surface to examine the extent of such bias. Figure 3.2a and 3.2b are examples of possible shapes of the response surface. The x-axis represents values of $\pi_{1|0}$; the y-axis represents values of $\pi_{1|1}$; and the z-axis represents the average estimated treatment effects. If the shape is as steep as in Figure 2a, it means that the estimated treatment effect will change dramatically when varying the strength of the relationship between *U* and *Z*. In other words, the estimated treatment effect is more sensitive to the presence of hidden bias. On the contrary, if the shape is a shallow surface as displayed in Figure 2b, the estimated treatment effect will not change considerably with the strength of the relationship varying. Therefore, the estimated treatment effect is less sensitive to the presence of hidden bias.

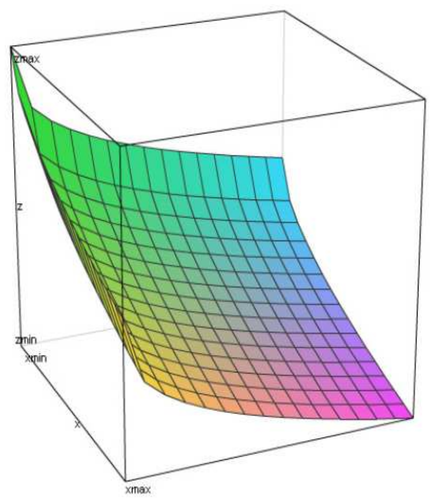


Figure 3.2a. An Example of a Steep Response Surface

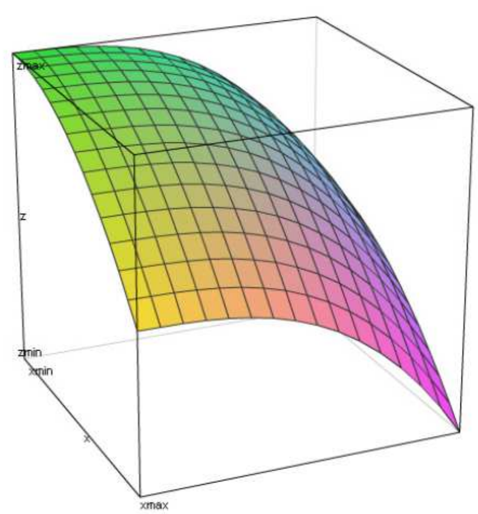


Figure 3.2b. An Example of a Shallow Response Surface

Note. From “Modeling Science Achievement Differences between Single-sex and Coeducational Schools: Analysis from Hong Kong, SAR and New Zealand from TIMSS 1995, 1999, and 2003”, by D. V. Diaconu, 2012, (Doctoral dissertation), *ProQuest Dissertations and Theses*, (Accession Order No. [UMI3521765]), pp.117-118. Copyright 2012 by Dana V. Diaconu. Reprinted with permission.

3.2 Variables

3.2.1 Outcome Variables

In the state of Massachusetts, the state–wide annual standardized test battery being used is the Massachusetts Comprehensive Assessment Systems (MCAS). Massachusetts Department of Elementary and Secondary Education (ESE) described MCAS as serving three purposes:

> (1) to inform and improve curriculum and instruction; (2) to evaluate student, school, and district performance according to the Massachusetts curriculum framework content standards and MCAS performance standards; and (3) to determine whether a student has met the state requirements for the Competency Determination (i.e., whether a student is eligible for a high school diploma. (2012, P.6)

It illustrated that MCAS was designed to evaluate the academic performance of each individual student; however, the MCAS test administration and reporting approaches were not designed to measure student growth from grade to grade.

MCAS tests use a variety of item types including multiple-choice, short-answer, short-response, open-response questions, and writing prompts, with each worth 1-10 points. The sum of the total points earned for a test is the raw score. At middle school grades, ELA is a 52-point test except for grade 7, which includes two writing prompts so the total is 72. The total points for math are 54, but the number of items for each item type has changed since 2008 (see Table 3.4). Using a “raw-score-to-theta equating procedure”, raw scores were converted into scaled scores, even-integer values with a range from 200 to 280 (ESE, 2007, p. 55). Fixed scale scores (220, 240, and 260) are then used to divide students into four distinct performance levels (advanced, proficient, needs improvement, and warning/failing) to be aligned with the Massachusetts curriculum framework content standards and MCAS performance standards. The emphasis is primarily on performance against achievement standards within a single grade rather than across grades.

Table 3.4

MCAS ELA and Math Item Types and Total Points during 2006-2012

Subject: ELA (2006-2012)	# of Items	# of Multiple-choice Items (1 point each)	# of Short Answers (1 point each)	# of Open-ended Responses (4 points each)	# of Writing Prompts (10 points each)	Total Points (raw scores)
Grade 6 & 8	40	36		4		52
Grade 7	42	36		4	2	72
Subject: Math (Grade 6-8)						
2006-2008	39	29	5	5		54
2008-2012	42	32	6	4		54

In this analysis, raw scores were used as outcomes because: 1) scaled scores involve a non-linear transformations and the range of possible scores are restricted; 2) raw scores have more variability (with a range of 0 up to 72) than scaled scores (with a range of 200 to 280 and only take even-integer values); 3) there were more raw scores than scaled scores in the analytic sample. To make scores interpretable across administration years and grades, all the raw scores were converted into *z* scores by subject, grade, and school year using the means and standard deviations of the sample.

The second type of outcomes was students' annual GPA in middle school. Applying the conversion rules listed in Table 3.5, letter grades were converted into scale scores ranging from 0 to 4. For honors and Advanced Placement (in Arts and Music) courses, an additional point was added. An annual GPA was the average score across all the courses a student had taken during a given year.

Table 3.5

Letter Grades and GPA Conversion Rules

Letter Grades	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
GPA	4	3.7	3.3	3	2.7	2.3	2	1.7	1.3	1	0.7	0

Note. AP course grades are increased by one point.

3.2.2 Student-level and School-level Covariates

Table 3.6 describes the student-level variables in details. Student-level covariates included time-invariant and time-varying demographic information, prior achievement, and treatment indicators. Time-invariant demographic variables (e.g., gender, race, free/reduced lunch status, bilingual status, special education status, and foreign born status) were used as a standard set of control variables for the first two research questions.

For RQ1, time-invariant demographic variables and time-varying ones in the outcome grades, together with prior achievement as measured by Grade 5 ELA and Mathematic scores, were used when adjusting for middle school effectiveness. For both the selection model (the PS weighting model) and the outcome model (the analytic model), in addition to the standard set of time-invariant demographic variables, Grade 1 Report Card scores in Reading, Mathematics, Writing, Work Habits, Behavior, and Effort and some other time-varying baseline characteristics (i.e., age, distance from school, the number of school moves in Grade 1) were used to account for pre-existing differences between the two groups.

Chapter 3. Research Design

Table 3.6

Descriptions of Student-level Covariates

Variables	Descriptions	Reference Group or Value Ranges
Time-invariant Demographic Information *(used in Equations 3.1 to 3.4)*		
Gender	The student is a male.	Female
Race/Ethnicity	Four dummy variables indicating a student's ethnicity: African American, Hispanic, Asian, and Others.	White
Free/Reduced Lunch Status	Two dummy variables indicating whether or not a student has ever received free- or reduced-priced lunch during his or her entire record in BPS.	Full-priced lunch status
Bilingual Status	The student has been enrolled in secondary English instruction.	Never enrolled
Special Education Status	Two dummy variables indicating whether or not a student has ever needed regular education with no more than 25% time out (SPED2) OR with no more than 60% time out (SPED3).	Never needed SPED or regular education with modifications
Foreign Born Status	The student is born outside of the U.S.A.	Born in the U.S.A.
Time-varying Demographic Information (*for Equation 3.1, the corresponding Grades 6/7/8 versions of the following variables will be used; for Equations 3.2 and 3.3, the Grade 1 version will be used; and for Equation 3.4, the Grade 5 version will be used*)		
Age	Student Age	6-18
Dist_fr_sch	Distance in miles from home to school	0-47
# of School Moves	Cumulative number of school moves	0-8
Prior Achievement		
RC_gr1 (*used in Equations 3.2 and 3.3*)	Reading, Mathematics, Writing, Work Habits, Behavior, and Effort Report Card scores converted into *z* scores	-3 to +3
MCAS _gr5 (*used in Equations 3.1 and 3.4*)	Grade 5 MCAS ELA and Math *z* raw scores	-3 to +3
Treatment Indicators (Student-level)		
Elementary City Connects Dose $EDose_{ij}$ (*used in Equation 3.4*)	Ever attended a City Connects elementary school in the grade where City Connects was implemented	Comparison students
Elementary City Connects Dosage (*used in Equations 3.2 and 3.3*)	Number of years spent in a City Connects elementary school	0-6

For RQ2, in addition to time-invariant demographic variables, students' prior achievement as measured by Grade 5 MCAS ELA and Mathematics scores and time-varying demographic characteristics in Grade 5 were included in the models. In addition, a student-level indicator of students' elementary City Connects dose was included and the interaction between this variable and the City Connects middle school dose indicator was tested.

Table 3.7 presents potential school-level variables. These variables came from two sources: the City Connects elementary and middle school treatment indicators were defined by the City Connects evaluation team; and school-level characteristics (including both the percentages of selected populations and school resource indicators) were directly extracted from the Massachusetts Department of Elementary and Secondary Education (ESE) website. Since multiple years of data were analyzed in this study, the values of the school-level characteristics were averaged across the years. Note that although all these school-level covariates were tested in the models, only a small number of them was included in the final models based on the results of significance testing and model fit statistics.

Chapter 3. Research Design

Table 3.7

Descriptions of School-level Covariates

Variables	Descriptions	Reference Group or Value Ranges
Treatment Indicators (School-level)		
Elementary City Connects Dose *EDose*$_k$ (*used in Equations 3.2 and 3.3*)	Ever attended an elementary school in the grade where City Connects was implemented	Comparison schools
Middle City Connects Dose *MDose*$_j$ (*used in Equation 3.4*)	Whether or not attended a City Connects middle school in Grade 6	Comparison schools
% of Selected Populations (*used in Equation 3.3*)		
% First Language not English	Percent of students whose first language is a language other than English in a given school	0-100%
% English Language Learner	Percent of students who are limited English proficient, meaning whose first language is a language other than English so that they are unable to perform ordinary classroom work in English	0-100%
% Low Income	Percent of students who are either eligible for free or reduced price lunch; or receives Transitional Aid to Families benefits; or are eligible for food stamps in a given school	0-100%
% Students with Disabilities	Percent of students who have an Individualized Education Program (IEP) in a given school	0-100%
School Resource Indicators (*used in Equation 3.3*)		
Student/teacher Ratio	Average student to teacher ratio during 2006-2012	
School Size	Average school size during 2006-2012	
Average Class Size	Average class size during 2006-2012	
Students per computer	Average students per computer during 2006-2012	

3.3 Centering Decision

The choice of placing predictors on different locations of their own distributions, such as no centering (raw metric), grand-mean centering, or group-mean centering, will affect the interpretations of both level-1 and level-2 parameters in a hierarchical linear model. As suggested by Raudenbush & Bryk, when estimating fixed level-1 coefficients, group-centering is preferred because it will produce unbiased estimate of β_w, "the level-1 relationship net of any group-membership effects" (2002, p. 135). However, when estimating level-2 effects while adjusting for level-1 covariates, grand-mean centering is recommended because the level-1 intercept becomes the mean outcome for each institution adjusted for differences between institutions in the means of level-1 covariates; whereas the intercept is the unadjusted mean of the outcome when group-mean centering is used. When compositional or contextual effect is of interest (β_c), which is defined as the effect of "the aggregate of a person-level characteristic, $\overline{X_{.j}}$" on the outcome, "even after controlling for the effect of the individual characteristic, X_{ij}", one can either calculate $\beta_c = \beta_b - \beta_w$ using group-mean centering or directly estimate β_c through grand-mean centering (Raudenbush & Bryk, 2002, p.135).

In this study, since the school-level treatment effect and the contributions of other institution-level characteristics were of primary interest, grand-mean centering of all the student-level covariates was applied in order to estimate school-level effects with adjustment for student-level covariates. This approach directly estimated contextual effects (aggregated student-level characteristics). Furthermore, for consistency and a clearer interpretation, all the dummy variables were grand-mean centered as well. Centering decision at level 2 is less critical. Grand-mean centering was applied to all the continuous variables at level 2 so that the value of zero is meaningful in interpretations.

3.4 Preliminary Analysis

3.4.1 Baseline Student Characteristics

Table 3.8a shows the baseline (i.e., at the beginning of Grade 1) student characteristics of the City Connects and the comparison groups in the analytic sample of

Chapter 3. Research Design

RQ1 before PS weighting: there were significantly more Asian students and fewer White, African American, and Hispanic students in the City Connects group. City Connects had significantly more bilingual and foreign born students than the comparison group. Compared to non-City Connects students, City Connects students lived significantly

Table 3.8a

Baseline (Grade 1) Student Characteristics by Group Membership for RQ1

Total: 11051	City Connects *N*=1791	Comparison *N*=9260
% Male	49.4%	48.4%
Race		
% White	9.6%	12.9%*
% African American	32.5%	41.6%*
% Asian	24.3%*	6.2%
% Hispanic	31.9%	37.7%*
% Other	1.7%	1.6%
% Bilingual	21.4%*	18.7%
Special Education		
% non SPED	81.1%	80.8%
% Regular Education with Modifications	0.5%	0.3%
% Regular Education with no more than 25% out (SPED2)	9.3%	9.8%
% Regular Education with no more than 60% out (SPED3)	9.0%	9.2%
Poverty Status		
% Receiving Full-price Lunch	5.8%	7.1%
% Receiving Reduced-price Lunch	3.1%	2.5%
% Receiving Free Lunch	91.1%	90.4%
% Foreign Born	14.2%*	10.1%
	Mean (SD)	Mean (SD)
RC_Reading_gr1	0.05(0.94)	0.17(0.99)*
RC_Math_gr1	0.02(0.89)	0.16(0.96)*
RC_Writing_gr1	0.03(0.90)	0.17(1.00)*
RC_WorkHabits_gr1	0.06(0.90)	0.16(0.96)*
RC_Behavior_gr1	-0.01(0.90)	0.14(0.95)*
RC_Effort_gr1	0.11(0.91)	0.15(0.97)
Age_gr1	6.73(0.33)	6.74(0.33)
Distance from School_gr1	2.56(2.54)*	1.56(1.83)
# of School Moves_gr1	0.22(0.42)	0.20(0.40)

Note. *Statistically significantly more/higher than the other group at $p<0.05$

further away from school when entering elementary school. Moreover, they performed significantly lower than their comparison counterparts in all Report Card measures except for Report Card Effort. However, there were no statistically significantly differences between the two groups in terms of gender, age and number of school changes when entering elementary school, percent of students who received free- or reduced-priced lunch, and percent of students who needed some type of special education.

Similarly, Table 3.8b shows the baseline (i.e., at the end of Grade 5) student characteristics of the City Connects and the comparison groups in the analytic sample of RQ2 before PS weighting: there were more Asian students and fewer White and African American students in the City Connects group. City Connects had more bilingual and foreign born students than the comparison group. Significantly fewer City Connects students received full-price lunch. Compared to non-City Connects students, City Connects students usually lived further away from school before entering middle school. Moreover, there were no statistical differences in terms of prior academic achievement in elementary school between the two groups.

To sum up, these comparisons on student characteristics at the baseline reveal that City Connects started with a relatively more disadvantaged population of students than the comparison group: a larger proportion of them were bilingual or born outside of the U.S.A; they usually lived farther away from school; and more of them struggled with poverty and suffered low achievement. In terms of ethnic composition, it is worth mentioning that the City Connects group did attract a larger population of Asian students. All these imbalances are expected to be removed by the PS weighting procedure.

Table 3.8b

Baseline (Grade 5) Student Characteristics by Group Membership for RQ2

Total: 8172	City Connects *N*=1037	Comparison *N*=7135
% Male	48.3%	50.4%
Race		
% White	15.9%	19.8%*
% African American	32.5%	37.4%*
% Asian	17.3%*	6.5%
% Hispanic	32.2%	34.7%
% Other	2.1%	1.6%
% Bilingual	21.2%*	18.1%
Special Education		
% non SPED	82.7%	80.6%
% Regular Education with Modifications	0.5%	0.3%
% Regular Education with no more than 25% out (SPED2)	9.9%	10.2%
% Regular Education with no more than 60% out (SPED3)	6.8%	8.9%
Poverty Status		
% Receiving Full-price Lunch	10.6%	14.4%*
% Receiving Reduced-price Lunch	3.7%	3.6%
% Receiving Free Lunch	85.7%	82.0%
% Foreign Born	16.3%*	13.3%
	Mean (SD)	Mean (SD)
MCAS_ELA_gr5	0.26(0.95)	0.26(0.90)
MCAS_Math_gr5	0.29(0.96)	0.23(0.92)
Age_gr5	11.49(0.49)	11.51(0.49)
Distance from School_gr5	2.03(1.89)*	1.76(1.98)
# of School Moves_gr5	0.81(0.92)	0.85(0.85)

Note. *Statistically significantly more/higher than the other group at $p<0.05$

3.4.2 Covariate Balancing Statistics

The covariate balancing statistics using both standardized bias and *p*-value methods are presented in Table 3.9a and 3.9b for RQ1 and in Table 3.10a and 3.10b for RQ2. For instance, in Table 3.9a, for the outcome measure of ELA in Grade 6, the standard bias value for “Distance from School_gr1” is 0.395 before weighting and becomes 0.001 after weighting. Since 0.001 is smaller than the threshold of 0.25, this covariate is considered balanced after applying the PS weights. One will draw the same conclusion using the *p* value method: As shown in Table 3.9b, the *p* value associated with the same covariate for the same outcome is 0.991, which is larger than 0.05; therefore, the null hypothesis is retained and the covariate of "Distance from School_gr1" is considered balanced after applying the PS weights.

According to Table 3.9 and Table 3.10, all the standardized bias values are smaller than 0.25, and even 0.10, after weighting, meaning that the City Connects and the comparison groups achieved balance in terms of the observed covariates; all the *p*-values are larger than 0.05, indicating that there was no statistically significant covariance imbalance left after applying the PS weights. Based on both statistics, PS weighting did successfully balance pre-existing observable differences between the two groups.

Table 3.9a

Covariate Balance (standardized bias method) for RQ1

ELA	Grade 6		Grade 7		Grade 8	
	before weighting	after weighting	before weighting	after weighting	before weighting	after weighting
Male	0.008	-0.009	0.004	0.024	-0.053	0.012
Is_Black	-0.208	-0.013	-0.075	-0.009	-0.050	-0.017
Is_Asian	0.463	-0.005	0.339	0.003	0.336	0.007
Is_Hispanic	-0.163	-0.002	-0.172	-0.001	-0.208	-0.010
Is_Other	0.004	0.011	0.029	0.016	0.024	-0.002
Bilingual	0.072	-0.054	-0.057	-0.026	-0.102	-0.039
SPED2	-0.036	0.007	-0.022	0.031	0.016	0.033
SPED3	0.002	0.019	0.072	0.008	0.080	0.011
Reduced Lunch	0.037	0.015	0.017	-0.007	-0.009	-0.014
Free Lunch	0.028	-0.027	0.006	0.011	0.007	0.006
Foreign Born	0.126	0.008	0.064	0.053	0.059	0.055
RC_Reading_gr1	-0.117	0.027	-0.132	-0.030	-0.084	-0.039
RC_Math_gr1	-0.162	0.035	-0.134	0.004	-0.163	0.006
RC_Writing_gr1	-0.169	0.023	-0.144	-0.036	-0.158	-0.038
RC_WorkHabits_gr1	-0.103	-0.001	-0.130	-0.054	-0.124	-0.059

RC_Behavior_gr1	-0.150	0.012	-0.201	-0.024	-0.222	-0.040
RC_Effort_gr1	-0.045	0.020	-0.055	-0.012	-0.057	-0.011
Age_gr1	-0.038	0.009	-0.028	-0.001	-0.026	-0.005
Distance from School_gr1	0.427	-0.028	0.482	-0.035	0.458	-0.049
# School Moves_gr1	0.050	0.014	0.094	0.024	0.112	0.042
Math	Grade 6		Grade 7		Grade 8	
	before weighting	after weighting	before weighting	after weighting	before weighting	after weighting
Male	0.007	-0.008	0.004	0.019	-0.050	0.013
Is_Black	-0.211	-0.016	-0.081	-0.013	-0.044	-0.019
Is_Asian	0.465	-0.005	0.340	0.004	0.335	0.006
Is_Hispanic	-0.164	0.000	-0.165	0.001	-0.213	-0.009
Is_Other	0.005	0.011	0.031	0.020	0.023	-0.002
Bilingual	0.073	-0.053	-0.053	-0.023	-0.106	-0.040
SPED2	-0.033	0.009	-0.025	0.026	0.024	0.036
SPED3	-0.005	0.018	0.064	0.002	0.076	0.010
Reduced Lunch	0.038	0.016	0.018	-0.006	-0.010	-0.014
Free Lunch	0.027	-0.027	0.004	0.009	0.006	0.003
Foreign Born	0.129	0.010	0.064	0.055	0.057	0.056
RC_Reading_gr1	-0.112	0.030	-0.132	-0.034	-0.082	-0.040
RC_Math_gr1	-0.158	0.038	-0.133	0.000	-0.154	0.009
RC_Writing_gr1	-0.164	0.026	-0.143	-0.038	-0.151	-0.037
RC_WorkHabits_gr1	-0.098	0.002	-0.129	-0.057	-0.116	-0.059
RC_Behavior_gr1	-0.146	0.015	-0.197	-0.023	-0.216	-0.038
RC_Effort_gr1	-0.041	0.023	-0.054	-0.015	-0.048	-0.009
Age_gr1	-0.036	0.013	-0.028	-0.001	-0.030	-0.002
Distance from School_gr1	0.424	-0.031	0.484	-0.033	0.464	-0.047
# School Moves_gr1	0.049	0.015	0.089	0.022	0.116	0.045
GPA	Grade 6		Grade 7		Grade 8	
	before weighting	after weighting	before weighting	after weighting	before weighting	after weighting
Male	0.005	-0.014	0.002	0.020	-0.053	0.006
Is_Black	-0.195	0.001	-0.062	0.001	-0.045	-0.009
Is_Asian	0.461	-0.005	0.342	0.005	0.330	0.005
Is_Hispanic	-0.170	-0.011	-0.187	-0.013	-0.205	-0.017
Is_Other	0.003	0.012	0.025	0.017	0.035	-0.003
Bilingual	0.068	-0.061	-0.063	-0.038	-0.117	-0.048
SPED2	-0.035	0.005	-0.011	0.033	0.025	0.036
SPED3	0.003	0.025	0.059	0.000	0.085	0.008
Reduced Lunch	0.042	0.018	0.020	-0.006	-0.005	-0.010

Free Lunch	0.035	-0.023	0.004	0.007	0.003	-0.001
Foreign Born	0.130	0.008	0.065	0.055	0.057	0.057
RC_Reading_gr1	-0.116	0.025	-0.129	-0.027	-0.081	-0.034
RC_Math_gr1	-0.165	0.029	-0.134	0.003	-0.166	0.010
RC_Writing_gr1	-0.170	0.016	-0.146	-0.040	-0.166	-0.041
RC_WorkHabits_gr1	-0.103	-0.005	-0.122	-0.053	-0.123	-0.057
RC_Behavior_gr1	-0.150	0.013	-0.189	-0.023	-0.212	-0.038
RC_Effort_gr1	-0.050	0.017	-0.050	-0.010	-0.067	-0.009
Age_gr1	-0.041	0.005	-0.028	-0.003	-0.040	-0.005
Distance from School_gr1	0.432	-0.033	0.482	-0.037	0.460	-0.052
# School Moves_gr1	0.057	0.021	0.082	0.023	0.100	0.037

Table 3.9b

Covariate Balance (p-value method) for RQ1

	ELA			Math			GPA		
	Gr6	Gr7	Gr8	Gr6	Gr7	Gr8	Gr6	Gr7	Gr8
Male	0.790	0.580	0.813	0.800	0.624	0.828	0.676	0.507	0.784
Is_Black	0.919	0.948	0.906	0.921	0.952	0.908	0.931	0.946	0.910
Is_Asian	0.971	0.979	0.953	0.971	0.979	0.953	0.976	0.985	0.953
Is_Hispanic	0.990	0.994	0.951	0.993	0.983	0.954	0.979	0.998	0.928
Is_Other	0.720	0.721	0.956	0.727	0.769	0.971	0.719	0.743	0.713
Bilingual	0.622	0.854	0.790	0.618	0.867	0.777	0.645	0.862	0.792
SPED2	0.828	0.557	0.525	0.819	0.564	0.594	0.789	0.534	0.529
SPED3	0.681	0.866	0.846	0.641	0.888	0.857	0.636	0.877	0.878
Reduced Lunch	0.636	0.849	0.814	0.649	0.843	0.837	0.659	0.914	0.863
Free Lunch	0.796	0.845	0.939	0.800	0.838	0.897	0.804	0.862	0.926
Foreign Born	0.890	0.553	0.455	0.899	0.536	0.459	0.871	0.560	0.422
RC_Reading_gr1	0.739	0.638	0.592	0.742	0.645	0.620	0.745	0.671	0.583
RC_Math_gr1	0.666	0.959	0.925	0.668	0.965	0.921	0.707	0.964	0.874
RC_Writing_gr1	0.779	0.507	0.582	0.786	0.526	0.607	0.805	0.490	0.548
RC_WorkHabits_gr1	0.986	0.485	0.438	0.977	0.483	0.442	0.980	0.488	0.443
RC_Behavior_gr1	0.857	0.656	0.533	0.843	0.665	0.508	0.910	0.635	0.539
RC_Effort_gr1	0.808	0.887	0.900	0.808	0.887	0.917	0.825	0.880	0.906
Age_gr1	0.758	0.986	0.944	0.741	0.993	0.960	0.927	0.921	0.928
Distance from School_gr1	0.649	0.611	0.510	0.664	0.596	0.478	0.673	0.607	0.486
# School Moves_gr1	0.867	0.759	0.576	0.850	0.763	0.608	0.864	0.775	0.603

Table 3.10a

Covariate Balance (standardized bias method) for RQ2

Grade 6	ELA		Math		GPA	
	before weighting	after weighting	before weighting	after weighting	before weighting	after weighting
Male	-0.064	-0.001	-0.062	0.000	-0.069	-0.015
Is_Black	-0.079	0.009	-0.074	0.013	-0.083	0.022
Is_Asian	0.291	-0.006	0.291	-0.005	0.330	0.019
Is_Hispanic	-0.079	0.008	-0.086	0.000	-0.072	0.015
Is_Other	0.031	-0.002	0.031	-0.002	0.038	0.011
Bilingual	0.076	-0.041	0.071	-0.045	0.116	-0.004
SPED2	-0.027	0.000	-0.029	-0.006	-0.040	-0.019
SPED3	-0.076	-0.006	-0.080	-0.010	-0.090	-0.014
Reduced Lunch	0.000	0.007	0.001	0.008	-0.049	-0.049
Free Lunch	0.125	0.003	0.124	0.001	0.239	0.126
Foreign Born	0.073	-0.018	0.068	-0.023	0.102	0.014
MCAS_ELA_gr5	0.002	0.003	0.004	0.005	-0.006	-0.004
MCAS_Math_gr5	0.071	-0.009	0.074	-0.008	0.118	0.038
Age_gr5	-0.054	-0.009	-0.060	-0.014	-0.047	0.007
Distance from School_gr5	0.148	0.014	0.146	0.015	0.144	0.005
# School Moves_gr5	-0.054	0.034	-0.053	0.036	-0.027	0.064

Table 3.10b

Covariate Balance (p-value method) for RQ2

	Gr6		
	ELA	Math	GPA
Male	0.985	0.984	0.934
Is_Black	0.958	0.956	0.995
Is_Asian	0.968	0.970	0.976
Is_Hispanic	0.955	0.957	0.953
Is_Other	0.932	0.937	0.912
Bilingual	0.737	0.729	0.682
SPED2	0.994	0.983	0.953
SPED3	0.930	0.913	0.842
Reduced Lunch	0.911	0.906	0.977
Free Lunch	0.990	0.992	0.990
Foreign Born	0.741	0.716	0.769
MCAS_ELA_gr5	0.983	0.991	0.982
MCAS_Math_gr5	0.954	0.953	0.980

Age_gr5	0.916	0.910	0.868
Distance from School_gr5	0.904	0.907	0.981
# School Moves_gr5	0.751	0.739	0.790

CHAPTER 4. RESULTS

4.1 Research Question One

The first set of research questions aims at estimating the impact of the City Connects elementary school intervention on middle school achievement as measured by standardized MCAS scores and criterion-referenced GPA grades. The associated three sub-questions will be answered one by one in the following sections.

4.1.1 RQ1a

RQ1a asks whether receiving the City Connects intervention in elementary school helps students succeed in middle school, as measured by standardized MCAS scores and criterion-referenced GPA grades, and if this success persists through all middle school grades, after controlling for student characteristics and pre-existing academic achievement differences. A series of hierarchical linear models were built for each subject and in each grade via a two-phase analysis.

As discussed in Chapter 3, in order to differentiate the estimated treatment effect of City Connects elementary schools from general middle school effectiveness, statistical adjustments need to be applied. In the first phase, the middle school adjustment score, the average predicted score for each middle school, was obtained from two-level linear models predicting the outcomes of interest with the current middle schools, employing data from the comparison students attended those schools. Then the middle school adjustment score was subtracted from every student's score on each of the outcomes of interest. These middle school adjustment models took account of students' demographic characteristics and their baseline academic differences (MCAS ELA and Mathematics scores in Grade 5). The unconditional intra-class correlation coefficient (ICC), which is the ratio between between-group variance and the total variance (i.e., $\frac{\tau_{00}}{\tau_{00} + \sigma^2}$), together with the total percent of variance explained by the model (i.e., $\frac{(\tau_{00} - \hat{\tau}_{00_{resid}} + \sigma^2 - \hat{\sigma}_{resid}{}^2)}{\tau_{00} + \sigma^2}$) (the conditional ICC), are reported in Table 4.1 for each subject and in each grade. Detailed HLM results and the corresponding statistical models can be found in Appendix A.

Table 4.1

Variance Partitions for RQ1a: Middle School Adjustment Models

	Grade 6		Grade 7		Grade 8	
	ICC	% Total Variance Explained	ICC	% Total Variance Explained	ICC	% Total Variance Explained
MCAS ELA	15.2%	63.7%	9.4%	46.1%	4.5%	45.3%
MCAS Mathematics	16.4%	68.2%	15.6%	51.6%	10.2%	47.2%
Weighted GPA	17.0%	44.7%	11.1%	21.3%	9.2%	19.6%

According to the table, all the outcome models have moderate between-school variance (9.2% to 17.0%, with one exception of 4.5% for the Grade 8 ELA outcome model). It means that approximately 10% to 20% of the total variance for each of the outcome measures was due to middle school differences, which justified the application of hierarchical linear models to adjust for the clustering effect. In addition, the inclusion of students' demographic characteristics and their prior achievement explained about half or even two thirds of the total variation (44.7% to 68.2%) among outcomes except for the Grade 7 and Grade 8 GPA outcome model, for which 21.3% and 19.6% of the total variance were explained by the models.

Table 4.2 presents the basic measures of central tendency and variability for the middle school adjustment score for each subject and in each grade. For MCAS measures, the outcome scores were converted into *z* scores by subject, grade, and school year using the means and the standard deviations of the sample. For GPA, letter grades were first converted into scale scores ranging from 0 to 4 with an additional point added for honors and Advanced Placement (in Arts and Music) courses, and then the converted GPA scores were averaged across all the courses that students had taken during a given year to get the weighted GPA scores.

Table 4.2

Basic Descriptive Statistics of Middle School Adjustment Scores

	N	*Minimum*	*Maximum*	*Mean*	*SD*
Grade 6					
MCAS ELA	44	-1.39	1.24	0.10	0.72
MCAS Mathematics	44	-1.62	1.20	0.05	0.78
Weighted GPA	44	1.16	3.91	2.62	0.60
Grade 7					
MCAS ELA	41	-1.19	0.83	-0.01	0.52
MCAS Mathematics	41	-1.58	0.96	-0.15	0.64
Weighted GPA	41	1.43	3.71	2.46	0.49
Grade 8					
MCAS ELA	40	-1.55	0.94	-0.01	0.60
MCAS Mathematics	40	-1.67	1.13	-0.20	0.69
Weighted GPA	40	1.57	3.76	2.54	0.49

As shown in the table, the estimated academic achievement did vary to a certain extent by middle school. If the apparent effect of students' attending different middle schools on middle school achievement is not adjusted for, one will get biased estimates of the effect of elementary schools: estimated elementary school effects on middle school outcomes may not be due solely to students attending different elementary schools; instead, it may be a result of these students attending different middle schools. Empirical evidence of differential middle school effectiveness justifies the application of middle school adjustment.

The second phase comprises two stages. The PS weights were generated in the first stage. According to the preliminary analysis results presented at the end of Chapter 3, applying such weights did significantly reduce overt selection bias. In other words, the treatment and the comparison groups were nearly statistically equivalent in terms of all the explanatory variables in the outcome models after applying the PS weights (see Section 3.4.2 for details).

A series of two-level linear regression models were built for each subject and in each grade in the second stage. These models did adjust for students' demographic characteristics and their prior achievement differences at the baseline grade at the student level and estimated the City Connects elementary treatment effect on adjusted middle school outcomes at the school level. School Clusters were defined as the last City Connects or comparison elementary schools that students attended. Furthermore, to maximize the difference between the treatment and the comparison groups, a series of dummy variables indicating years spent with City Connects were included at the student level with the maximum years of City Connects (six years of City Connects) as the reference group. By doing so, the estimated treatment effect at the school level can be interpreted as the average difference between the comparison students and the City Connects students who received the maximum "dosage" of City Connects.

Variance partitions of the final outcome models are reported in Table 4.3. The unconditional ICCs (i.e., $\frac{\tau_{00}}{\tau_{00}+\sigma^2}$) indicate that approximately 7% to 16% of the total variability among each of the (adjusted) outcome measures was due to students' attending different elementary schools. The between-school variation decreased as students progressed through grades (e.g., from 14.8% to 7.3% as students progressing from grade 6 to grade 8 for the MCAS ELA outcome models), which is reasonable because variation associated with elementary schools attended should fade with more distal outcomes.

In addition, the conditional ICC, the percent of variance explained by the outcome model at each level (i.e., $\frac{(\sigma^2-\hat{\sigma}_{resid}{}^2)}{\sigma^2}$ at Level 1 and $\frac{(\tau_{00}-\hat{\tau}_{00resid})}{\tau_{00}}$ at Level 2), was also reported for each subject and in each grade: approximately 20% to 30% of the variation among outcomes of interest at Level 1 was explained by the student-level demographic characteristics and students' prior achievement measures.

The lowest percent of variance explained at Level 2 for the MCAS measures is in Grade 7: the City Connects treatment indicator only explained 2.1% of the Level-2 variation for MCAS ELA and 2.6% for MCAS Mathematics. For weighted GPA, the lowest percent of variance explained at Level 2 is in Grade 8: 7.5% of the Level-2 variation was explained by the City Connects treatment indicator. All the other percent values explained at Level 2 are large in magnitude: the City Connects elementary school treatment indictor, the only predictor at Level 2, explained 10.2% to 18.0% of the total variance at the school level for MCAS measures in Grades 6 and 8 and 25.4% to 27.6% for the weighted GPA in Grades 6 and 7. To conclude, it seems that whether or not students' last elementary schools attended were City Connects schools was generally an important factor in explaining between-elementary school differences in terms of academic achievement in middle school.

Chapter 4. Results

Table 4.3

Variance Partitions for RQ1a: Final Outcome Models

	Grade 6			Grade 7			Grade 8		
	ICC	% Variance Explained at Each Level		ICC	% Variance Explained at Each Level		ICC	% Variance Explained at Each Level	
		Level 1	Level 2*		Level 1	Level 2		Level 1	Level 2
MCAS ELA	14.8%	25.7%	12.2%	9.6%	18.7%	2.1%	7.3%	19.1%	13.3%
MCAS Mathematics	16.2%	27.5%	18.0%	10.6%	18.3%	2.6%	12.0%	19.9%	10.2%
Weighted GPA	14.8%	27.6%	27.6%	9.2%	18.4%	25.4%	7.4%	17.7%	7.5%

Note. *% Variance Explained at Level 2 reports the additional variance explained by including the treatment variables into the corresponding outcome models.

Table 4.4 summarizes the results of the estimated treatment effects of the City Connects elementary intervention on middle school academic outcomes obtained from the final models of RQ1a. Fixed effects at Level 1 and random effects at Level 2 are not reported here since they are not of interest. Detailed HLM results and the corresponding statistical models can be found in Appendix B.

Generally speaking, City Connects elementary school graduates outperformed their counterparts who graduated from comparison elementary schools on grade 6 academic achievement measures, after taking into account available student characteristics and pre-existing academic achievement differences. The strong positive effect did persist to Grade 7 and Grade 8.

To be specific, in Grade 6, students who received the City Connects treatment in elementary school scored statistically significantly higher on all the three subjects (0.43, $t(86) = 2.462$, $p = 0.016$ for MCAS ELA; 0.67, $t(86) = 2.695$, $p = 0.009$ for MCAS Mathematics; and 0.64, $t(86) = 8.622$, $p < 0.001$ for weighted GPA) when taking into account available students' demographic and academic differences at the baseline.

Chapter 4. Results

Table 4.4

Results of RQ1a: Estimated City Connects Elementary Treatment Effects (γ_{01}) in Middle School

	Coef.	s.e.	p-value	Effect Size
MCAS ELA				
Grade 6	0.43	0.17	**0.016**	0.41
Grade 7	0.25	0.08	**0.004**	0.29
Grade 8	0.38	0.09	**0.000**	0.44
MCAS Mathematics				
Grade 6	0.67	0.25	**0.009**	0.62
Grade 7	0.38	0.16	**0.022**	0.42
Grade 8	0.63	0.16	**0.000**	0.67
Weighted GPA				
Grade 6	0.64	0.07	**0.000**	0.67
Grade 7	0.38	0.06	**0.000**	0.44
Grade 8	0.34	0.06	**0.000**	0.40

Note. Bolded values are statistically significant at .05.

Although still being statistically significant (0.25, $t(84) = 3.019$, $p = 0.004$ for MCAS ELA; 0.38, $t(84) = 2.328$, $p = 0.022$ for MCAS Mathematics; and 0.38, $t(84) = 6.023$, $p < 0.001$ for weighted GPA), the magnitude of the positive effect of City Connects dropped a little bit in Grade 7 (from 0.43 to 0.25 for MCAS ELA; from 0.67 to 0.38 for MCAS Mathematics; and from 0.64 to 0.38 for weighted GPA).

The positive effect of the City Connects elementary intervention became strong again for all the three outcomes in Grade 8 (0.38, $t(84) = 4.285$, $p < 0.001$ for MCAS ELA; 0.63, $t(84) = 3.894$, $p < 0.001$ for MCAS Mathematics; and 0.34, $t(84) = 5.648$, $p < 0.001$ for weighted GPA).

Additionally, following the approach suggested by the WWC (2011, F.9), effect sizes were computed as Hedges' g, the ratio of the estimated treatment effect to the unadjusted pooled within-group SD using the formula below:

Chapter 4. Results

$$g = \frac{\omega\gamma}{\sqrt{\frac{(n_i-1)s_i^2+(n_c-1)s_c^2}{n_i+n_c-2}}} \tag{4.1}$$

where γ is the HLM coefficient for the intervention's effect;

ω is the small-sample bias corrector calculated as 1 - 3/(4*df* - 1) (Hedges, 1981, p.114), with *df* being the number of degree of freedom used to estimate the standard deviation;

s_i and s_c are the unadjusted standard deviations of the treatment and the comparison groups, separately;

and n_i and n_c are the sample sizes of the two groups.

As shown in Table 4.4, the magnitude of effect sizes is quite large with a range of 0.29 to 0.67 for MCAS measures and a range of 0.40 to 0.67 for weighted GPA, indicating that the estimated treatment effects of the City Connects elementary intervention on middle school academic outcomes are not only statistically significant but also practically significant: the City Connects group outperformed the comparison group by at least one third of a standard deviation (i.e., for the outcome of MCAS ELA in Grade 7) and at most two thirds of a standard deviation (i.e., for the outcome of MCAS Mathematics in Grade 8 and the outcome of weighted GPA in Grade 6).

Furthermore, since the final outcome models included both a binary treatment indicator of City Connects elementary schools (City Connects EDose) and a series of dummy variables indicating years spent with City Connects elementary schools (City Connects Dosage), the regression coefficients (γ_{01}) associated with EDose represent the estimated differences in outcomes between City Connects students who received the maximum years of City Connects (six years) and comparison students. Meanwhile, the regression coefficients (β_{1k} *to* β_{5k}) associated with the dosage dummy variables represent the estimated differences in outcomes between City Connects students who received the corresponding years of City Connects and those who received maximum years of City Connects.

As reported in Table 4.5, except for some minor fluctuations, the coefficients associated with City Connects dosage variables are mostly negative values decreasing in absolute value with increasing years of exposure to the treatment; that is, as years spent with City Connects elementary schools increase, the average estimated differences in outcomes between those who received maximum years of City Connects and those who received fewer than maximum years of City Connects become smaller. In other words,

Table 4.5

Results of RQ1a: Estimated City Connects Elementary Dosage Effects (β_{1k} to β_{5k}) in Middle School

	Grade 6			Grade 7			Grade 8		
Fixed Effects	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
MCAS ELA									
1-Year vs. 6-Year Dosage	-0.33	0.17	**0.048**	-0.19	0.06	**0.002**	-0.29	0.10	**0.004**
2-Year vs. 6-Year Dosage	-0.15	0.18	0.388	-0.25	0.07	**0.001**	-0.46	0.12	**0.000**
3-Year vs. 6-Year Dosage	-0.08	0.20	0.675	-0.06	0.13	0.609	-0.10	0.15	0.537
4-Year vs. 6-Year Dosage	-0.13	0.14	0.335	-0.04	0.08	0.635	-0.28	0.13	**0.032**
5-Year vs. 6-Year Dosage	0.15	0.08	0.060	0.07	0.14	0.604	-0.11	0.21	0.605
MCAS Mathematics									
1-Year vs. 6-Year Dosage	-0.54	0.29	0.065	-0.29	0.21	0.163	-0.47	0.18	**0.011**
2-Year vs. 6-Year Dosage	-0.36	0.24	0.135	-0.30	0.19	0.111	-0.55	0.16	**0.001**
3-Year vs. 6-Year Dosage	-0.26	0.23	0.245	-0.25	0.14	0.066	-0.26	0.11	**0.019**
4-Year vs. 6-Year Dosage	-0.33	0.16	**0.043**	-0.18	0.11	0.105	-0.24	0.12	**0.034**
5-Year vs. 6-Year Dosage	-0.12	0.15	0.421	-0.12	0.16	0.448	-0.30	0.18	0.102
Weighted GPA									
1-Year vs. 6-Year Dosage	-0.52	0.11	**0.000**	-0.18	0.10	0.069	-0.27	0.06	**0.000**
2-Year vs. 6-Year Dosage	-0.32	0.07	**0.000**	-0.31	0.09	**0.001**	-0.40	0.06	**0.000**
3-Year vs. 6-Year Dosage	-0.33	0.09	**0.000**	-0.10	0.08	0.200	-0.06	0.09	0.523
4-Year vs. 6-Year Dosage	-0.16	0.07	**0.024**	0.02	0.10	0.854	-0.11	0.05	**0.013**
5-Year vs. 6-Year Dosage	-0.13	0.09	0.115	-0.05	0.14	0.727	-0.11	0.16	0.476

Note. Bolded values are statistically significant at .05.

the more years one receives the City Connects elementary intervention, the higher his or her academic achievement likely will be in middle school grades.

4.1.2 RQ1b

RQ1b focuses on relative school effectiveness among the City Connects schools. Residual-based estimates associated with City Connects elementary schools were obtained from the outcome models addressing RQ1a.

In order to summarize how much each of the City Connects elementary schools differed from one another, Table 4.6 shows the standard deviation of the residual-based estimates for the City Connects elementary schools and compares it with the estimated treatment effect for each subject and in each grade. As we see, the variation among City Connects elementary schools is generally smaller than the magnitude of the difference between City Connects and comparison elementary schools except for Grade 7 ELA and Grade 8 weighted GPA (the corresponding standard deviations of the residual-based estimates of City Connects elementary schools are 0.34 and 0.37, respectively, which are larger than the estimated treatment effects of 0.25 and 0.35). It indicates that City Connects elementary schools generally outperformed comparison schools and their contributions to students' academic improvement were not driven by some exceptional schools: City Connects elementary schools did not differ more from each other than they were different from the comparison schools in terms of their students' academic achievement in middle school.

Chapter 4. Results

Table 4.6

Standard Deviations of the Residual-based Estimates of City Connects Elementary Schools versus the Estimated Treatment Effects

	MCAS ELA			MCAS Math			Weighted GPA		
	Treatment VAM *SD* (y)	Estimated Treatment Effect (x)	Ratio (x/y)	Treatment VAM *SD* (y)	Estimated Treatment Effect (x)	Ratio (x/y)	Treatment VAM *SD* (y)	Estimated Treatment Effect (x)	Ratio (x/y)
Grade 6	.42	.43	1.02	.60	.67	1.11	.40	.64	1.59
Grade 7	.34	.25	0.74	.34	.38	1.12	.24	.38	1.56
Grade 8	.25	.38	1.53	.44	.63	1.44	.37	.35	0.94

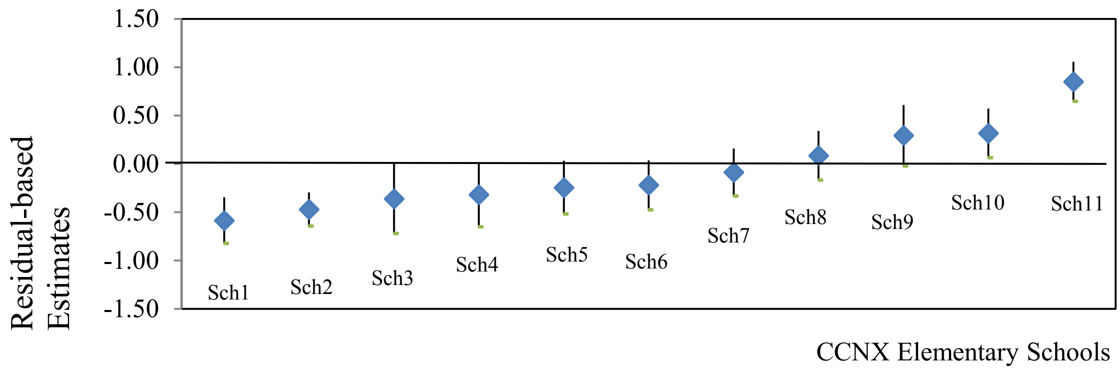
Meanwhile, Figures 4.1 to 4.3 display both point estimates of school-specific residuals for City Connects elementary schools (denoted as blue diamonds in the figures) and the associated 95% confidence intervals (shown as the vertical bars centered at the point estimates of school-specific residuals). The reference line at 0 represents the average effect for all sampled schools (both City Connects and comparison elementary schools) after adjusting for the treatment effect in the outcome models[6].

In general, Sch4, Sch9, and Sch5 have relatively wide confidence intervals. It primarily reflects their relatively small numbers of students who took the tests (shown in Table 4.7): fifty or less students from these schools took the three tests in Grade 6, about half of them took the tests in Grade 7, and then about half of the remaining students took the tests in Grade 8[7]. The confidence intervals around the estimates for Sch6 are narrow

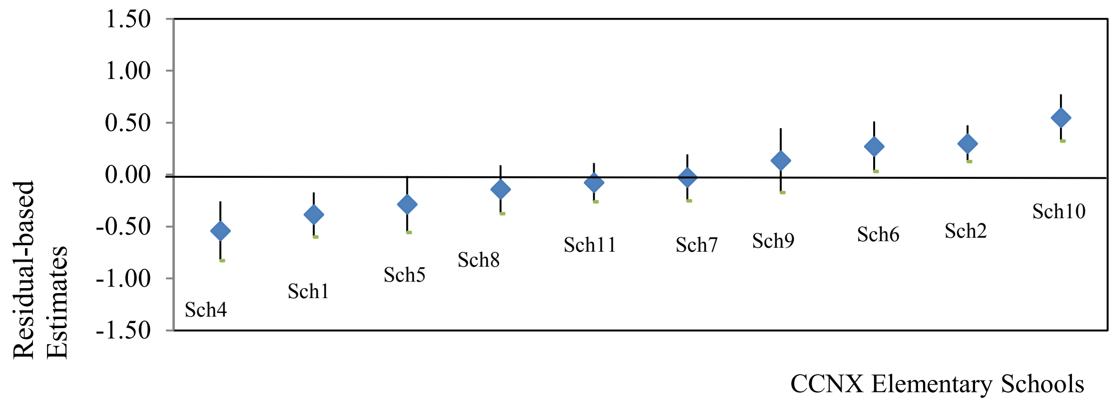
[6] Based on the results to RQ1a, one may expect a majority of the point estimates of CCNX schools to be located above the reference line since the treatment effect on average is strongly positive. However, the treatment effect (Dose) was included in the outcome models, which means the positive effect of CCNX has already been removed from/centered among the residuals. Therefore, it makes sense to have the residuals varying around the zero reference line.

[7] This degree of attrition was partly due to losing one cohort per grade given the design of the study and partly due to student mobility, which might be an issue and will be discussed in Chapter 5.

Grade 6



Grade 7



Grade 8

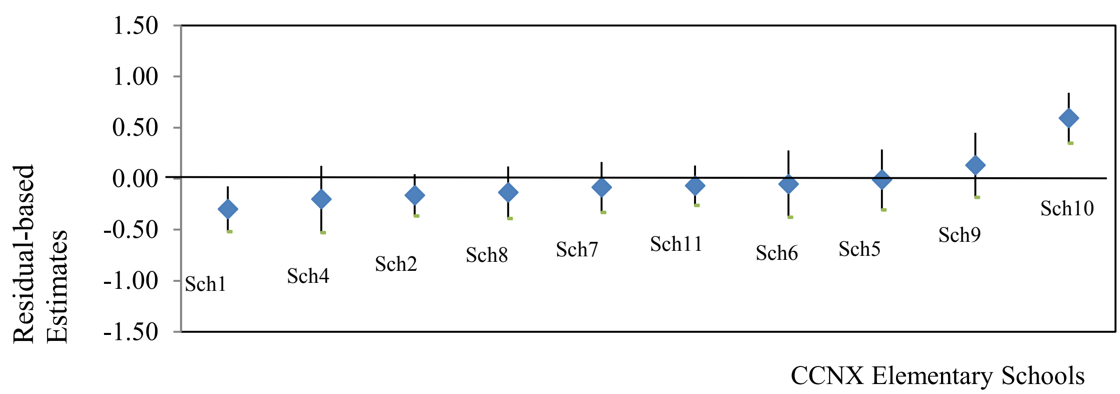


Figure 4.1. Residual-based Estimates and 95% Confidence Intervals of City Connects Elementary Schools Obtained from MCAS ELA Outcome Models

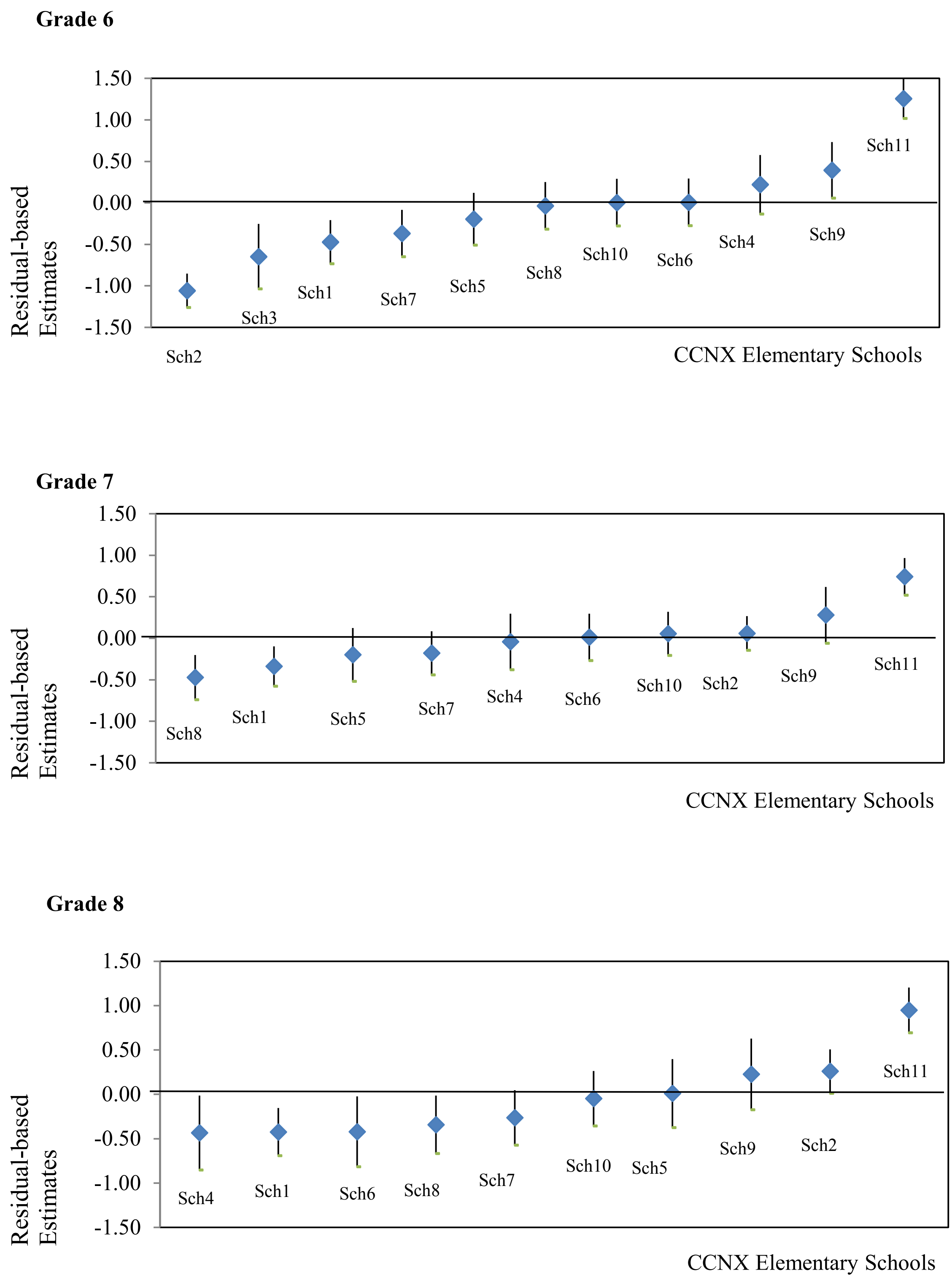
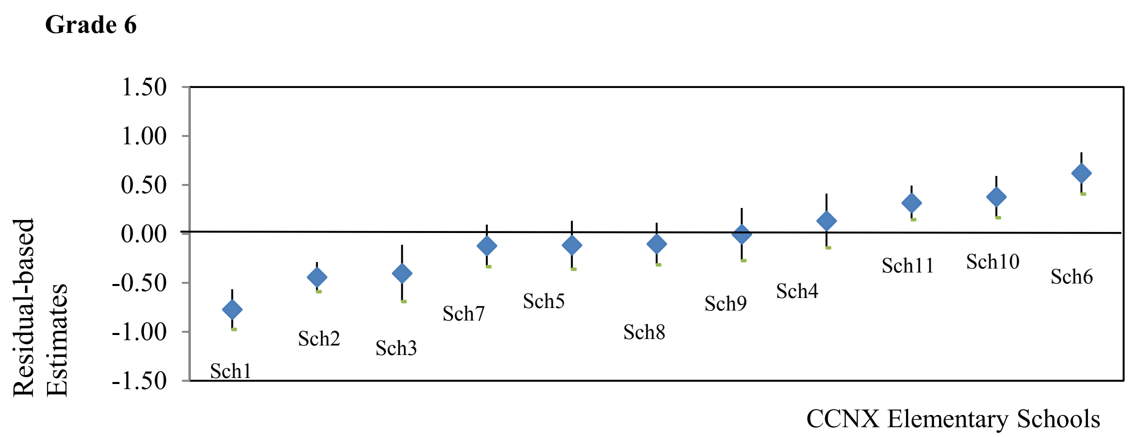
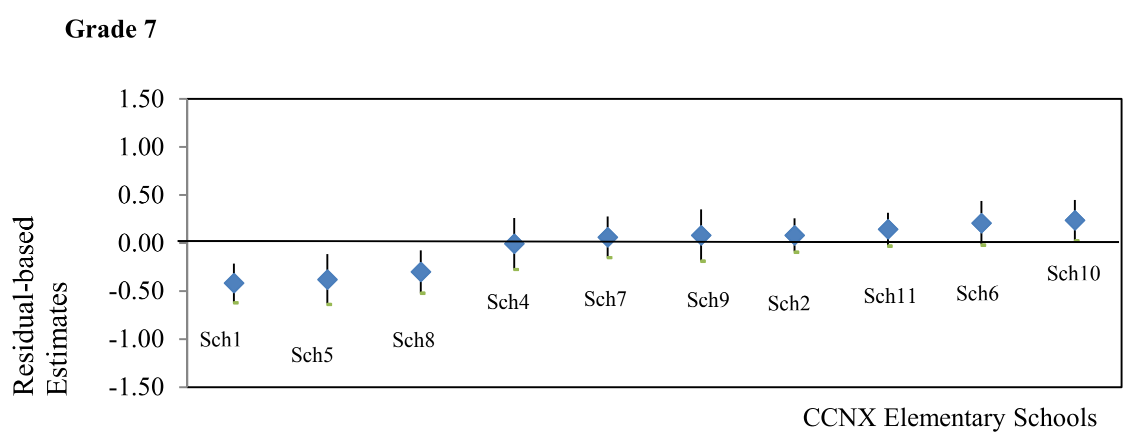


Figure 4.2. Residual-based Estimates and 95% Confidence Intervals of City Connects Elementary Schools Obtained from MCAS Mathematics Outcome Models





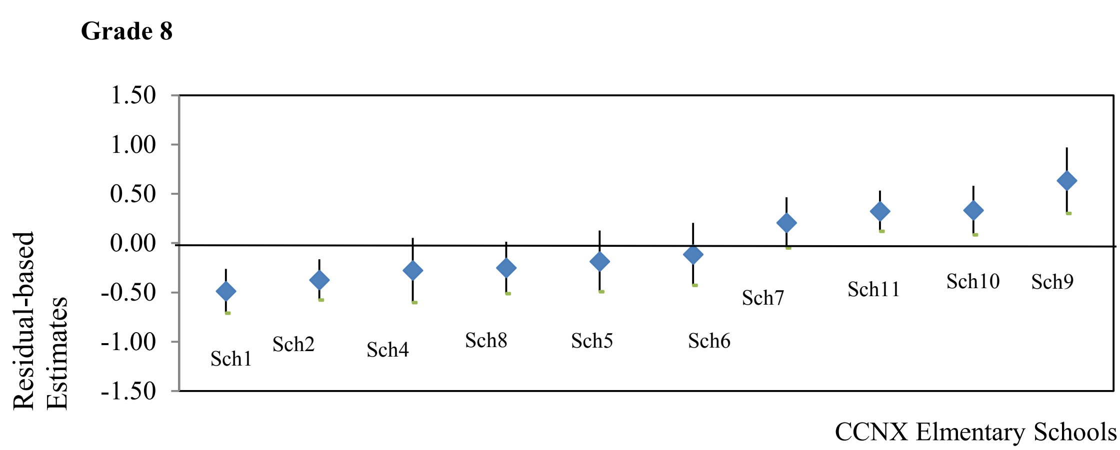


Figure 4.3. Residual-based Estimates and 95% Confidence Intervals of City Connects Elementary Schools Obtained from Weighted GPA Outcome Models

for all the subjects in Grades 6 and 7 but are widened in Grade 8 due to sample size changes: approximately 100 students who came from Sch6 took the three tests in Grade 6; like other schools, the number were cut in half in Grade 7. However, the sample size dropped rapidly to only about 10 students for all the subjects in Grade 8.

Table 4.7

*RQ1b: Analytic Sample Sizes by City Connects Elementary Schools**

	Grade 6			Grade 7			Grade 8		
	ELA	Math	GPA	ELA	Math	GPA	ELA	Math	GPA
Sch3	33	33	34						
Sch4	51	51	51	31	31	31	15	16	16
Sch9	51	51	46	22	21	20	13	13	12
Sch10	116	116	119	79	78	80	49	49	51
Sch1	100	100	100	67	67	66	54	55	54
Sch7	109	107	111	64	63	64	43	43	45
Sch6	102	102	103	37	37	38	10	11	12
Sch5	44	44	46	22	22	23	15	15	15
Sch8	102	100	104	53	53	54	33	33	34
Sch11	365	362	372	180	179	182	132	131	132
Sch2	392	391	393	139	138	138	88	88	89

Note. Two schools were dropped from the analytic samples due to less than 10 students took the tests in Grades 6

In terms of the magnitude of point estimates of school-specific residuals, some schools' performances, as measured by the "added values" to their students' test scores, are quite consistent across MCAS subjects and grades. For instance, Sch9 placed consistently near the top among City Connects schools for both MCAS measures and in all grades. Sch1, on the other hand, remained near the bottom for both MCAS measures and in all three grades.

Regarding the top-performing school for each MCAS subject, Sch11 ranked first for MCAS Mathematics in all the three grades; however, for MCAS ELA, although it took the first place in Grade 6, it became an average-performing school in later grades. In contrast, Sch10 ranked the top or near the top for MCAS ELA in all the three grades;

however, its' scores on MCAS Mathematics were consistently at the average across the grades.

Furthermore, some schools made noticeable progress over the three years of middle school for one of the MCAS subjects while other schools failed gradually. Sch2, for example, started from the bottom in Grade 6 and slowly climbed to the top in later grades for MCAS Mathematics. On the contrary, Sch6 ranked the 4th in Grade 6, the 5th in Grade 7, and dropped to the 8th place in Grade 8 for the same subject. A similar trend was observed for Sch4, which started near the top in Grade 6 and gradually lagged behind and ended up being the last in Grade 8 in terms of MCAS Mathematics.

At last, for weighted GPA, Sch10 and Sch11 performed the best in all grades; while Sch1 performed the worst. Sch9 started "near expected" in Grade 6 and gradually climbed to the top in Grade 8; whereas Sch4 started at a similar position but gradually dropped to the bottom. A similar trend was observed for Sch6 as well: it started near the top and gradually dropped to "near expected" in Grade 8. Schools such as Sch7, Sch5, and Sch8, had overlapping confidence intervals and therefore their differences were not easily distinguishable in all subjects and in all grades.

Finally, as reported in Table 4.8, a residual index statistic was calculated as the ratio of the estimated variance of within-school residuals to that of between-school residuals (for detailed formula, please see footnote 5 on page 49 in Chapter 3). The variance of within-school residuals is random variation that the outcome models failed to explain; while that of between-school residuals is due to differences in the estimated school effects. The smaller the ratio the better the outcome models.

Generally speaking, the values of the residual index statistics are the smallest for all the subjects in Grade 6 and they are greater for higher grades. Therefore, Grade 6 outcome models have the best model fit among those in all middle school grades.

Table 4.8

Residual Index Statistics of the Relative Effects of City Connects Elementary Schools

	MCAS ELA	MCAS Math	Weighted GPA
Grade 6	0.065	0.068	0.085
Grade 7	0.186	0.146	0.329
Grade 8	0.435	0.214	0.437

4.1.3 RQ1c

RQ1c asks to what extent the estimated treatment effects can be accounted for by school characteristics. To answer this question, same models as RQ1a were used, with school-level covariates added at Level 2. These school-level indicators were extracted directly from the DESE website and were averaged over the years (if available) that schools were included in the sample. Some covariates were only collected for a certain period of time: for instance, DESE has stopped reporting “students per computer” since the 2011-2012 academic years; and “average class size” has only been available since the 2010-2011 academic years. If a characteristic of interest was not available for the sampled school years, data extracted from the nearest year were used.

Variance partitions for the final models addressing RQ1c are reported in Table 4.9. For each grade, the first column reports the percent of variance explained by the City Connects treatment indictor (obtained from Table 4.3); and the second column reports the additional percent of variance explained by the newly-added school-level covariates. Note that since none of the school-level covariates was statistically significant for MCAS

ELA outcome models in Grades 7 and 8, the percent variance explained by these covariates is 0% in the corresponding cells.

Table 4.9

Variance Partitions at Level 2 for RQ1c: Final Outcome Models with School-level Covariates

	Grade 6		Grade 7		Grade 8	
	EDose	School Covariates	EDose	School Covariates	EDose	School Covariates
MCAS ELA	12.2%	2.7%	2.1%	0%	13.3%	0%
MCAS Mathematics	18.0%	7.0%	2.6%	13.6%	10.2%	7.0%
Weighted GPA	27.6%	4.5%	25.4%	3.7%	7.5%	5.3%

The range of the additional percent of variance explained by school-level characteristics is from 2.7% to 13.6% for the MCAS outcomes and from 3.7% to 5.3% for the weighted GPA. In general, the combined contribution of school-level characteristics to explain school-level academic differences is quite small when compared to that of the City Connects treatment indicator except for Grade 7 MCAS Mathematics: whether or not the school was a City Connects elementary school only explained 2.6% of school-level variation; while school-level characteristics explained an additional 13.6%.

The results of the estimated treatment effects when school characteristics were added, together with the regression coefficient associated with each of these school-level covariates, are reported in Table 4.10. Detailed HLM results and the corresponding statistical models can be found in Appendix C. To compare, the estimated treatment effects before school characteristics were added (obtained from Table 4.4) are also reported in italic type.

Chapter 4. Results

Table 4.10

Results of RQ1c: Estimated City Connects Elementary Treatment Effects and the Effects of School Characteristics on Academic Achievement Indicators in Middle School

Note. Bolded values are statistically significant at .05.

	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
MCAS ELA									
% Low Income	-0.01	0.00	**0.011**						
% Students with Disabilities	-0.03	0.01	**0.005**						
Student/Teacher Ratio	-0.08	0.03	**0.017**						
Dose (Ever City Connects)	0.42	0.16	**0.011**	0.25	0.08	0.004	0.38	0.09	0.000
Dose from Table 4.4	*0.43*	*0.17*	*0.016*	*0.25*	*0.08*	*0.004*	*0.38*	*0.09*	*0.000*
MCAS Mathematics									
Average Class Size	-0.07	0.02	**0.001**	-0.04	0.02	**0.029**	-0.05	0.02	**0.011**
% Low Income	-0.01	0.00	0.066						
% Free Lunch				-0.01	0.00	**0.020**			
% Reduced Lunch				-0.04	0.02	**0.021**			
Dose (Ever City Connects)	0.73	0.24	**0.003**	0.43	0.15	**0.005**	0.59	0.15	**0.000**
Dose from Table 4.4	*0.67*	*0.25*	*0.009*	*0.38*	*0.16*	*0.022*	*0.63*	*0.16*	*0.000*
Weighted GPA									
Average Class Size							-0.03	0.02	0.095
% Free Lunch	-0.00	0.00	0.091	-0.01	0.00	0.116			
% Reduced Lunch	-0.05	0.02	**0.001**	-0.03	0.01	**0.030**			
Dose (Ever City Connects)	0.72	0.07	**0.000**	0.42	0.06	**0.000**	0.33	0.06	**0.000**
Dose from Table 4.4	*0.64*	*0.07*	*0.000*	*0.38*	*0.06*	*0.000*	*0.34*	*0.06*	*0.000*

According to the table, the estimated treatment effects and the associated standard errors did not differ too much from each other with or without school characteristics. It indicates that the observed school characteristics did not explain differences in apparent effectiveness of City Connects schools. In other words, the City Connects elementary intervention did have a significantly positive impact on students' academic achievement, even after taking into account some major differences in school compositions and resources.

The final set of school-level covariates for each outcome model was chosen based on the results of significance testing. Extra attention should be paid to the covariates in the final models. Major influential school characteristics that had statistically significant associations with students' academic achievement are average class size and percent of students who are eligible for free- or reduced-priced lunch: small class size was associated with high academic achievement, which is consistent with previous findings (Glass, Cahen, Smith, & Filby, 1982; Schanzenbach, 2014). Additionally, schools with higher percent of students coming from lower socio-economic backgrounds, as manifested by having higher percent of students eligible for free- or reduced-priced lunch, performed worse academically. Finally, it is worth mentioning that both the percent of students who had an IEP in a given school and the student to teacher ratio are strongly associated with MCAS ELA scores in Grade 6: schools with higher percent of students of disabilities tested lower on MCAS ELA; so did schools with higher student to teacher ratios.

4.2 Research Question Two

The second set of research questions aims at estimating the impact of the City Connects middle school intervention on middle school achievement as measured by standardized MCAS scores and criterion-referenced GPA grades. The associated three sub-questions will be answered one by one in the following sections.

4.2.1 RQ2a

RQ2a asks whether or not City Connects middle schools are more effective in improving students' academic performances than non-City Connects middle schools in Grade 6, after controlling for student characteristics and pre-existing academic achievement differences at the end of elementary school. A two-level linear regression model was built for each subject. These models adjusted for demographic characteristics and prior achievement in elementary school (students' academic records by the end of spring semester when attending Grade 5) at the student level, and placed the middle school City Connects treatment indicator at the school level. School Clusters were defined as current City Connects or non-City Connects middle schools that students attended. Again, PS weights were applied at the student level.

The unconditional ICCs reported in Table 4.11 indicate that approximately 10% of the total variability in each of the outcome measures was due to students' attending different middle schools. Moreover, the first column of the conditional ICCs shows that approximately 50% to 70% of the variation among outcomes of interest at Level 1 was explained by student-level covariates, which is reasonable given prior achievement by the end of elementary school and students' demographic characteristics were included in the outcome models.

Table 4.11

Variance Partitions for RQ2a: Final Outcome Models

	Grade 6		
	ICC	% Variance Explained at Each Level	
		Level 1	Level 2*
MCAS ELA	10.5%	67.1%	1.3%
MCAS Mathematics	10.2%	70.2%	0%
Weighted GPA	10.4%	49.6%	0%

Note. *% Variance Explained at Level 2 reports the additional variance explained by including the treatment variable into the corresponding outcome models

The second column of the conditional ICCs reports the additional percent of variance explained by the inclusion of the treatment indicator at Level 2: the City

Connects middle school treatment indicator only explained 1.3% of the Level-2 variation in MCAS ELA and close to 0% in MCAS Mathematics and weighted GPA. To conclude, it seems that whether or not students were currently attending one of the City Connects middle schools was not a key factor in explaining between-middle school differences in terms of academic achievement in Grade 6.

Table 4.12 summarizes the results of the estimated treatment effects of the City Connects middle school intervention on middle school academic outcomes obtained from the final models of RQ2a. Fixed effects at Level 1 and random effects at Level 2 are not reported here since they are not of interest. Detailed HLM results and the statistical models can be found in Appendix D.1.

Table 4.12

Results of RQ2a: Estimated Treatment Effects (γ_{01}) of Middle School City Connects on Middle School Outcomes

	Coef.	s.e.	p-value	Effect Size
Grade 6				
MCAS ELA	-0.06	0.03	0.104	-0.07
MCAS Mathematics	0.08	0.05	0.086	0.09
Weighted GPA	0.06	0.06	0.342	0.07

Note. Bolded values are statistically significant at .05.

Generally speaking, the City Connects middle school intervention did not have a strong statistical association with students' academic achievement measures in Grade 6, after taking into account available student characteristics and pre-existing academic achievement differences in elementary school. This perhaps is not a surprising result because the positive effect of attending City Connects elementary schools, if it exists, was removed from the analysis through the statistical control of academic differences at the end of Grade 5; and students can only receive the City Connects middle school intervention for one year by the end of Grade 6. It may take several years for the intervention to "ferment" and make an observable difference.

Furthermore, like the elementary school counterpart, the middle school intervention seems to show stronger associations with Mathematics than with ELA. The estimated treatment effect of City Connects middle schools is -0.06 ($t(23) = -1.691$, $p = 0.104$) for MCAS ELA and 0.08 ($t(23) = 1.791$, $p = 0.086$) for MCAS Mathematics. The latter estimate, although is still not statistically significant at the 0.05 level, approaches the threshold. Finally, although the impact of City Connects middle schools on weighted GPA is not statistically significant ($t(22) = 0.972$, $p = 0.342$), the direction of this impact is promising: the positive number 0.06 indicates that on average students who attended City Connects middle schools performed higher than those who attended non-City Connects middle schools.

Within the City Connects group (1037 students), 528 of them attended City Connects elementary schools before they were enrolled in City Connects middle schools and the rest (509 students) did not. A dummy variable indicating whether or not a student had attended one of the City Connects elementary schools was added at the student level to further differentiate the contribution of different sectors of City Connects on academic outcomes. Figure 4.4 displays the sample breakdown by City Connects elementary and middle school attendance.

In the figure, the City Connects middle school indicator was denoted as "MDose" and the City Connects elementary school indicator was denoted as "EDose". Note that students who were not currently attending City Connects middle schools but had attended

City Connects elementary schools (MDose = 0 & EDose = 1) were the target treatment sample for RQ1; therefore, they were not part of the sample for RQ2.

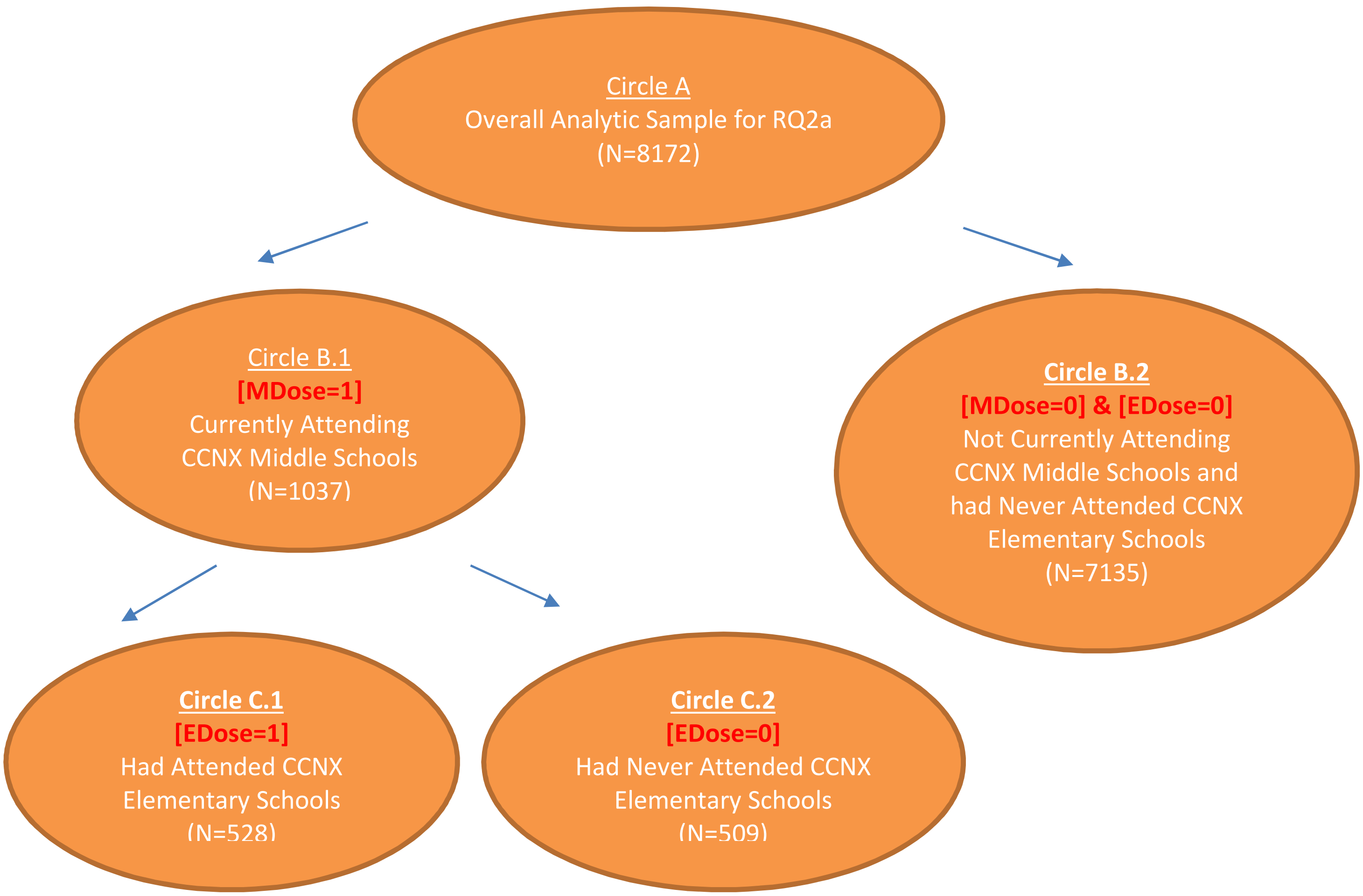


Figure 4.4 Sample breakdown by City Connects elementary and middle school attendance for RQ2a

By including EDose into the outcome models, the contribution of City Connects middle schools on academic achievement in Grade 6 can be estimated without confounding it with that of City Connects elementary schools. Table 4.13 summarizes the results (for detailed HLM results and the corresponding statistical models, please consult Appendix D.2).

Chapter 4. Results

Table 4.13

Results for RQ2a: Estimated Treatment Effects (γ_{01}) of Middle School City Connects on Middle School Outcomes with the Treatment Indicator of Elementary School City Connects (β_{1j}) Included

	MDose (γ_{01})			EDose (β_{1j})		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Grade 6						
MCAS ELA	-0.05	0.03	0.175	-0.02	0.03	0.558
MCAS Mathematics	0.04	0.05	0.419	0.05	0.03	0.120
Weighted GPA	0.00	0.05	0.970	0.09	0.04	**0.035**

Note. Bolded values are statistically significant at .05.

When including EDose in the outcome model for each subject, the regression coefficient (γ_{01}) associated with MDose no longer represents the average difference in Grade 6 outcomes between students who attended one of the City Connects middle schools and those who attended one of the comparison middle schools but had never attended City Connects elementary schools (Circle B.1 versus Circle B.2 in Figure 4.4). Instead, it represents the average difference in outcomes between those who attended City Connects middle schools but never attended City Connects elementary schools and the comparison students (Circle C.2 versus Circle B.2 in Figure 4.4). This is because γ_{01} represents the average difference in outcomes between students who had EDose = 0 & MDose = 1 and those who had EDose = 0 & MDose = 0. As discussed earlier, when MDose equals 0, EDose will always equal 0 since students with MDose = 0 & EDose = 1 were removed from the analysis and served as the target treatment sample for RQ1.

Following the same logic, β_{1j} represents the difference in outcomes between students who had EDose = 1 & MDose = 1 and those who had EDose = 0 & MDose = 1. In other words, it represents the difference between students who attended both City

Connects elementary and middle schools and those who only attended City Connects middle schools (Circle C.1 versus Circle C.2 in Figure 4.4)

Take the outcome model for MCAS Mathematics in Grade 6 as an example: γ_{01} is 0.04 and β_{1j} is 0.05, both of which are not statistically significant at the 0.05 level. It means that on average students who attended one of the City Connects middle schools but had never attended any City Connects elementary schools scored 0.04 points higher on MCAS Mathematics in Grade 6 than those who attended one of the comparison middle schools and also had never attended any City Connects elementary schools. Meanwhile, those students who attended both City Connects middle and elementary schools scored 0.05 points higher than those who only attended City Connects middle schools. The inequality is expressed as follows:

City Connects middle school attendants who had attended City Connects elementary schools

$$\underset{(0.05)}{>}$$

City Connects middle school attendants who had NOT attended City Connects elementary schools

$$\underset{(0.04)}{>}$$

Non-City Connects middle school attendants who had NOT attended City Connects elementary schools (4.2)

As shown in the table, the values of γ_{01} are all smaller than those reported in Table 4.12, for which EDose was not included in the outcome models. This is because the City Connects group used to estimate γ_{01} in Table 4.12 comprised both students who only attended City Connects middle schools and those who attended both City Connects middle and elementary schools (Circle B.1); whereas in Table 4.13, γ_{01} was estimated using City Connects students who only attended City Connects middle schools (Circle C.2). The fact that all these regression coefficients are small and statistically insignificant indicates that only attending City Connects middle schools did not make a noticeable impact on students' academic achievement in Grade 6.

Furthermore, β_{1j} being positive values indicates that students who attended both City Connects middle and elementary schools (Circle C.1) generally performed better than those who only attended City Connects middle schools (Circle C.2). In the case of Weighted GPA, the coefficient is even statistically significant (i.e., the associated p-value is 0.035), meaning that students who attended both scored significantly higher than those who only attended City Connects middle schools.

In theory, it is quite unlikely since by including prior academic achievement, the main effect of the City Connects elementary intervention has already been taken into account. However, since this main effect was adjusted for using students' prior achievement scores on MCAS measures, it could be argued that weighted GPA, although being strongly correlated with MCAS scores, is still a different measure that the main effect on GPA scores could not be completely controlled for by the prior performance on state-mandated standardized assessments. In addition, if the City Connects elementary intervention creates a positive environment that addresses non-cognitive factors and eventually fosters learning, then it is reasonable to believe that these non-cognitive boosters might not be fully captured by controlling for prior standaradized academic measures.

4.2.2 RQ2b

RQ2b evaluates school effectiveness within the City Connects group. Residual-based estimates of City Connects middle school effects were obtained from the outcome models addressing RQ2a. Table 4.14 reports the analytic sample size of each of the City Connects middle schools for each subject. SchoolC is the largest City Connects middle school with approximately 500 students in the analytic samples and SchoolE is smallest with only about 40 students. SchoolA, SchoolB, and SchoolD are similar in size with about 100 to 130 students each.

Table 4.14

RQ2b: Analytic Sample Sizes by City Connects Middle Schools

	MCAS ELA	MCAS Mathematics	Weighted GPA
SchoolA	105	104	107
SchoolB	118	118	-*
SchoolC	528	528	539
SchoolD	128	128	130
SchoolE	35	33	36

Note. * Only 2 students enrolled in School B had valid scores on weighted GPA in Grade 6; therefore, these two cases were dropped from the analytic samples

For each subject, the standard deviation of the residual-based estimates of these five City Connects middle schools is shown in Table 4.15. The estimated treatment effects of City Connects middle school intervention on these three outcomes (obtained from Table 4.12) are also listed for reference. As we see, the variation among City Connects middle schools is substantially larger than the magnitude of the average difference between City Connects and comparison middle schools for each subject. In other words, City Connects middle schools differed more from each other than they were different from the comparison schools in terms of their students' academic achievement in Grade 6.

Table 4.15

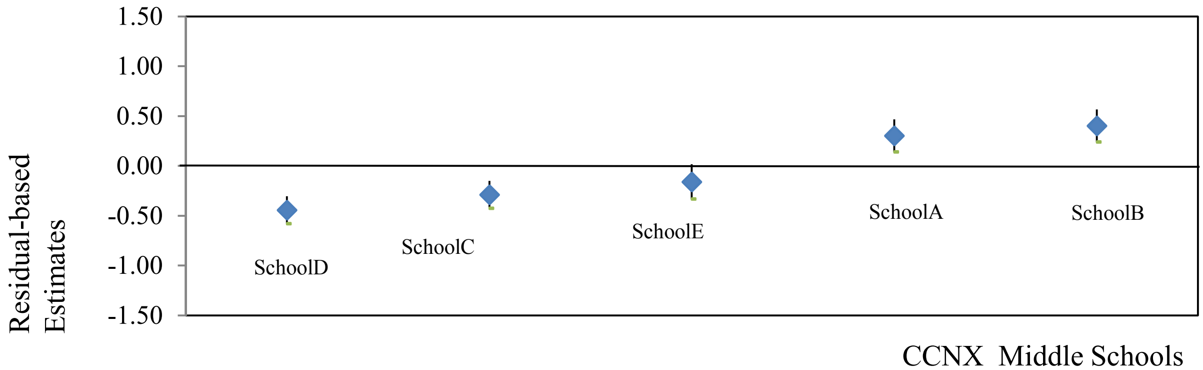
Standard Deviations of the Residual-based Estimates of City Connects Middle Schools versus the Estimated Treatment Effects

	MCAS ELA		MCAS Math		Weight GPA	
	Treatment VAM *SD*	Estimated Treatment Effect	Treatment VAM *SD*	Estimated Treatment Effect	Treatment VAM *SD*	Estimated Treatment Effect
Grade 6	.37	-.06	.37	.08	.44	.06

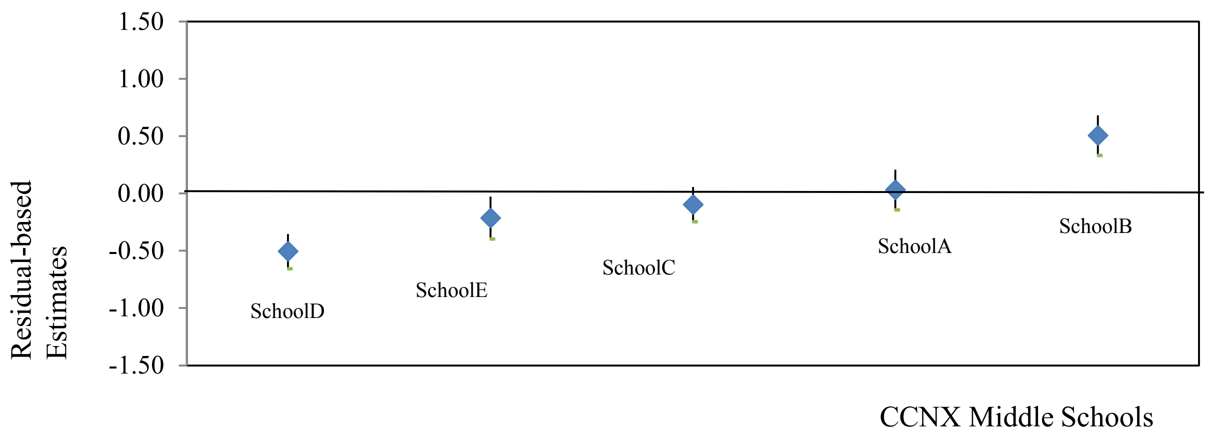
Figures 4.5 displays both point estimates of school-specific residuals for City Connects middle schools and the associated 95% confidence intervals. The reference line at 0 indicates the average performance of all middle schools (both City Connects and comparison ones). SchoolB performed the best among the five City Connects K-8 schools in terms of its students' scores on MCAS measures. Both point estimates and confidence intervals indicate that SchoolB performed significantly above expected on both MCAS ELA and Mathematics (0.40 with 95% CI [0.24, 0.57] and 0.51 with 95% CI [0.33, 0.68], separately).

In contrast, both point estimates and confidence intervals indicate that SchoolD performed the worst among the five City Connects K-8 schools for all the three subjects (-0.44 with 95% CI [-0.58, -0.30] for MCAS ELA; -0.51 with 95% CI [-0.66, -0.35] for MCAS Mathematics, and -0.68 with 95% CI [-0.86, -0.51] for weighted GPA). SchoolA took the second place next to SchoolB for both MCAS ELA and Mathematics with point estimates of 0.30 and 0.03 respectively. It ranked first for weighted GPA with a point estimate of 0.38 after SchoolB was dropped out of the sample due to missing too

MCAS ELA



MCAS Mathematics



Weighted GPA

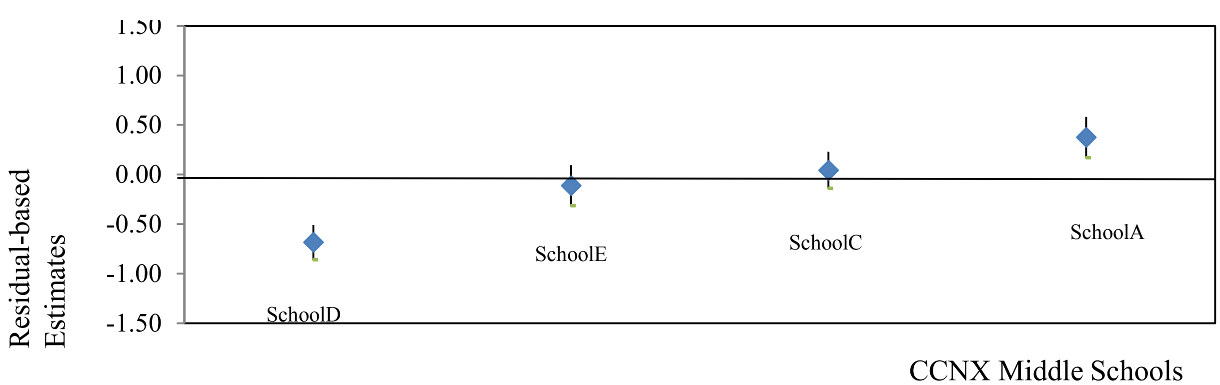


Figure 4.5 Residual-based Estimates and 95% Confidence Intervals of City Connects Middle Schools Obtained from Middle School Outcome Models

many cases. It scored above expected for MCAS ELA (95% CI [0.14, 0.47]) and weighted GPA (95% CI [0.17, 0.58]) but performed relatively mediocre for MCAS Mathematics (95% CI [-0.14, 0.21]). The rest two schools, SchoolE and SchoolC, scored near the reference line for all the three subjects (SchoolE: -0.16 with 95% CI [-0.33, 0.02] for MCAS ELA; -0.21 with 95% CI [-0.40, -0.03] for MCAS Mathematics; and -0.11 with 95% CI [-0.32, 0.09] for weighted GPA. SchoolC: -0.29 with 95% CI [-0.43, -0.15] for MCAS ELA; -0.10 with 95% CI [-0.25, 0.06] for MCAS Mathematics; and 0.04 with 95% CI [-0.14, 0.23] for weighted GPA) and their performances are indistinguishable.

Finally, the residual index statistics is reported for each subject in Table 4.16. The values of these statistics are small and similar to each other in magnitude (0.028 for MCAS ELA; 0.024 for MCAS Mathematics; and 0.027 for weighted GPA), which indicates that the residual-based estimates obtained from the outcome models to answer RQ2a are quite reliable.

Table 4.16

Residual Index Statistics of the Relative Effects of City Connects Middle Schools

	MCAS ELA	MCAS Math	Weighted GPA
Grade 6	0.028	0.024	0.027

4.2.3 RQ2c

RQ2c asks the same question as RQ1c: To what extent can the estimated treatment effects be accounted for by school characteristics? Again, school-level indicators were added to the outcome models and a summary of the key results is shown in Table 4.17 (for more details, see Appendix E). Additionally, the estimated treatment effects before school characteristics were added (obtained from Table 4.12) are reported in italic type for comparison.

According to the table, with the inclusion of school characteristics, the estimated treatment effects became statistically significant for some outcome measures (i.e., the associated regression coefficient is 0.13, $t(21) = 5.169$, $p < 0.001$ for MCAS Mathematics and the associated regression coefficient is 0.14, $t(20) = 3.462$, $p = 0.003$ for weighted GPA). It indicates that the City Connects middle school intervention did have a significantly positive impact on MCAS Mathematics and weighted GPA scores in Grade 6, if some differences in school composition and resources are taken into account.

Table 4.17

Results of RQ2c: Estimated City Connects Middle School Treatment Effects and the Effects of School Characteristics on Academic Achievement Indicators in Grade 6

	MCAS ELA			MCAS Mathematics			Weighted GPA		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	0.50	0.12	0.001	0.27	0.03	0.000	2.74	0.09	0.000
% Foreign Language Not English				-0.01	0.00	0.002	-0.01	0.00	0.001
Students per Computer	0.06	0.01	0.001	0.07	0.01	0.000	0.05	0.01	0.004
Student/Teacher Ratio	-0.03	0.01	0.011						
MDose (Ever City Connects)	-0.04	0.03	0.144	0.13	0.02	**0.000**	0.14	0.04	**0.003**
MDose from Table 4.12	*-0.06*	*0.03*	*0.104*	*0.08*	*0.05*	*0.086*	*0.06*	*0.06*	*0.342*

Note. Bolded values are statistically significant at .05.

Based on the results of significance testing, students per computer was included in the final models due to its significantly positive relationships with all the three outcome measures (the associated regression coefficient is 0.06, $t(21) = 4.120$, $p = 0.001$ for MCAS ELA; 0.07, $t(21) = 5.314$, $p < 0.001$ for MCAS Mathematics; and 0.05, $t(20) = 3.277$, $p = 0.004$ for weighted GPA): the more students shared one computer, the higher was these students' achievement.

One reasonable explanation is that the number of computers at school does not matter that much; instead, the quality of computer use does. It might be the case that for schools that possess relatively small number of computers, they struggle to use such limited resources as efficiently as possible: they may carefully design their classroom practices and after-class assignments that are in need of the assistance of computers. As a result, their students benefit more from using computers. Based on this explanation, students per computer will serve as a proxy for efficient computer use. However, more information about computer use at school (e.g., how often computers are used; how computers are used; who are assisting or monitoring computer use?) needs to be collected and a more representative sample of middle schools needs to be selected to test this hypothesis.

Moreover, the percent of students whose first language was not English was a strong predictor of MCAS Mathematics (-0.01, $t(21) = -3.657$, $p = 0.002$) and weighted GPA (-0.01, $t(20) = -3.884$, $p = 0.001$) in Grade 6: the higher the percent of students whose first language was not English that a given school had, the lower was its average score on MCAS Mathematics.

Finally, the student to teacher ratio was a statistically significant predictor for MCAS ELA in Grade 6 (-0.03, $t(21) = -2.821$, $p = 0.011$): a higher student/teacher ratio was associated with lower test scores. In other words, more students per teacher that a given school had, the lower was its average score on MCAS ELA.

4.3 Research Question Three

The last research question asks whether the estimated treatment effects obtained in answering the first two research questions are robust to the presence of hidden selection bias. To partially answer this question, a sensitivity analysis was conducted. Since the results of RQ1 demonstrated strong and positive treatment effects on all the outcomes and in all the grades, a sensitivity analysis was only conducted for the outcome models answering RQ1. Additionally, given the heavy computational requirements of sensitivity analysis and limited computer capacity, only outcome models addressing MCAS ELA and Mathematics scores were investigated.

As described in Chapter 3, two assumptions are made about the unobserved variable *U*, which represents hidden bias in this study. The first assumption requires *U* to be probabilistically related to *Z* (the binary treatment indicator) in a particular manner: to be specific, the conditional probabilities of U given Z need to satisfy $\pi_{1|1} > \pi_{1|0}$. The second assumption requires *U* to be strongly positively related to the outcomes of interest.

To satisfy the first assumption about *U*, 10 representative pairs of conditional probabilities of *U* given *Z* were chosen to simulate *U*. Each simulated value of *U*, multiplied by the corresponding pre-determined regression coefficient, was adjusted from the outcome scores. The treatment effect was then re-estimated using the adjusted outcome scores instead. These steps were repeated 100 times for each pair of the conditional probabilities and the treatment effect estimates were averaged across the 100 trials.

The pre-determined regression coefficient associated with each set of the simulated *U* was set equal to 0.3 based on the empirical results of RQ1: Among all the positive (estimated) regression coefficients indicating the relationships between the corresponding variables and the outcomes (see Appendix B), 0.3 is considered relatively strong and influential. Using the outcome model predicting Grade 6 MCAS ELA as an example, the relationship between "Foreign Born" and "Grade 6 ELA" is 0.26, which is the largest positive regression coefficient in that model. The largest regression coefficient associated with prior achievement is 0.12 between "RC_Reading_gr1" and the outcome. It means that if the regression coefficient of the hypothesized U has a value of 0.30, this U would be more influential than Foreign Born status and any of the prior achievement measures in predicting the outcome. Whether or not such a strongly influential U exists in reality is questionable. However, this conservative approach of assuming a relatively strong relationship will result in reductions of the estimated treatment effects. If the re-estimated effects are still significant, then this will strengthen the argument that City Connects indeed produced consistent and positive impacts on the outcomes of interest, despite the presence of selection bias.

4.3.1 Software Comparisons and Model Simplification

Ideally, the procedure of the proposed sensitivity analysis should be conducted using the same software as the one used to answer the first two RQs (i.e., HLM 6.0) so that the results are consistent and comparable. However, HLM does not allow a looping process, which makes it nearly impossible to manually repeat the steps for a total of 6000 times (100 repetitions per pair by 10 pairs of conditional probabilities of U given Z by 6 outcomes (MCAS ELA and Mathematics in Grades 6 to 8)) within a reasonable time period. Therefore, Stata, an alternative data analysis and statistical software package that allows looping, was chosen. To be specific, Stata 13 MP (StataCorp, 2013b) was used due to its capacity to run on multiprocessor and multicore computers to produce results faster and its capacity to work with matrices that are larger than 800 in size (StataCorp, 2013c).

Unfortunately, different software systems yield distinctly different results as shown in Table 4.18, which compares results of RQ1 in predicting Grade 6 MCAS ELA produced by HLM and Stata (the corresponding columns with PS weights). One of the possible explanations is that HLM and Stata may handle sampling weights differently. Another possible explanation is that the method of estimation in HLM used to analyze the first two RQs is restricted maximum likelihood (REML). However, Stata does not support REML with weights; so Stata fits models via the maximum likelihood (MLE) estimation method if weights are applied. Unfortunately, similar results cannot be produced if Stata is used instead of HLM when weights are applied. Therefore, it is necessary to return to less complicated models for which the two software systems are able to produce similar results.

As shown in Table 4.18, random effects models without PS weights were chosen. By comparing the results with and without weights in HLM, it seems that the estimates did not deviate too much when the weights were removed from the models. Then the random effects models without PS weights produced by HLM were compared with the ones produced by Stata. The differences were negligible. Therefore, random effect models without PS weights run by Stata were chosen for the sensitivity analysis. The obtained results with *U* included were compared with the ones without *U*, with the understanding that the latter are still a little bit different from the original results of RQ1, since PS weights were not used.

Chapter 4. Results

Table 4.18

Software Comparison: Results of RQ1 (Outcome Models Predicting MCAS ELA in Grade 6)

Fixed Effects	HLM Results				Stata Results			
	With PS Weights		Without PS Weights		With PS Weights		Without PS Weights	
	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.*
Intercept	-0.07	0.05	-0.05	0.05	0.33	0.28	0.69	0.19
EDose (Ever City Connects)	0.42	0.17	0.39	0.18	0.32	0.21	0.40	0.14
1-Year vs. 6-Year Dosage	-0.33	0.17	-0.38	0.18	-0.37	0.17	-0.39	0.08
2-Year vs. 6-Year Dosage	-0.16	0.18	-0.17	0.15	-0.17	0.18	-0.17	0.08
3-Year vs. 6-Year Dosage	-0.08	0.20	-0.13	0.19	-0.10	0.20	-0.13	0.08
4-Year vs. 6-Year Dosage	-0.13	0.13	-0.15	0.10	-0.14	0.13	-0.16	0.08
5-Year vs. 6-Year Dosage	0.15	0.08	0.07	0.05	0.12	0.09	0.06	0.09
Male	-0.15	0.02	-0.17	0.02	-0.16	0.02	-0.17	0.02
is_Black	-0.20	0.04	-0.14	0.04	-0.18	0.04	-0.14	0.03
is_Asian	0.16	0.04	0.19	0.04	0.15	0.04	0.19	0.04
is_Hispanic	-0.13	0.04	-0.05	0.04	-0.13	0.04	-0.06	0.03
is_Other	-0.05	0.13	-0.03	0.07	-0.03	0.13	-0.03	0.08
Bilingual	-0.15	0.03	-0.16	0.03	-0.16	0.03	-0.16	0.03
Special Needs 2	-0.28	0.07	-0.23	0.04	-0.25	0.05	-0.24	0.04
Special Needs 3	-0.63	0.06	-0.67	0.05	-0.66	0.05	-0.65	0.04
Reduced Lunch	-0.04	0.06	-0.15	0.06	-0.05	0.05	-0.14	0.07
Free Lunch	-0.30	0.06	-0.32	0.04	-0.28	0.06	-0.32	0.04
Foreign Born	0.26	0.03	0.23	0.03	0.27	0.03	0.23	0.03
RC_Reading_gr1	0.12	0.03	0.11	0.02	0.12	0.02	0.11	0.02
RC_Math_gr1	0.03	0.02	0.05	0.01	0.03	0.02	0.05	0.02
RC_Writing_gr1	0.03	0.02	0.04	0.02	0.03	0.02	0.03	0.02
RC_WorkHabit_gr1	0.07	0.02	0.07	0.02	0.07	0.03	0.07	0.02
RC_Behavior_gr1	-0.02	0.03	-0.02	0.01	-0.02	0.02	-0.02	0.01
RC_Effort_gr1	0.03	0.04	0.02	0.02	0.03	0.03	0.02	0.02
Age_gr1	0.05	0.04	-0.03	0.03	0.02	0.04	-0.03	0.03
Distance from School_gr1	-0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01
# School Moves_gr1	0.03	0.03	0.01	0.03	0.03	0.03	0.01	0.02

Note. * Standard errors reported in this column are not robust standard errors like the other ones, which makes them relatively smaller.

4.3.2 Computer Capacity and Analysis Setup

In order to handle the heavy computations, it was advised by Dr. Matt Gregas to use the Linux Cluster at Boston College (personal communication, September 4, 2014), which uses multiple computers and multiple storage devices to form a single highly available system. The cluster adopts a queuing system and implements a fair-share scheduling policy to all faculty members and research teams on campus. Given the time when heavy computations were needed, the availability on the Linux Cluster to run all the repetitions was limited. Therefore, five additional home-owned high-performing desktop computers were used to share the workload.

It should be noted that Stata employs a pseudorandom-number generator function. This is a deterministic algorithm that always produces the same sequence of values given a certain input (StataCorp, 2013a). The underlying sequence is determined by the seed, a random number that serve as the initial value to start the sequence. With the same seed, Stata's random-number generation functions will always return to the same sequence of numbers every time Stata is launched. For this study, two seeds were chosen for each of the 100-trial job to break the job in half to make it manageable[8].

4.3.3 Results Predicting MCAS Mathematics

Table 4.19 presents the estimated treatment effects of the City Connects elementary intervention on MCAS Mathematics in Grades 6 to 8 (γ'_{01}) when the outcome

[8] It is recommended to choose a random seed and set the seed as infrequently as possible (StataCorp, 2013a). An ideal scenario would be to set a random seed and use it as the starting point to repeat the 100 trials for each pair of the 10 conditional probabilities. However, with limited number of computers and processors in hand, depending on the complexity of the models, the running time for each trial ranges from 2 minutes to 1 hour. By setting one seed per 100-trial job, the computers need to run non-stop for up to a week to finish just one job. To prevent overheating the computers, for each of the 100-trial job, two seeds were set to break it into two jobs with 50 trials each. The seeds chosen were 90907 and 21624, both of which were randomly generated by Excel.

models included different sets of *U* (the chosen 10 pairs of *U* based on the conditional probabilities of *U* given *Z* as listed in Table 3.3). In addition, the ones estimated from models that excluded *U* ($*\ \gamma_{01}$), together with the corresponding standard errors and the 95% confidence interval, are shown at the bottom of the table for comparison[9].

Table 4.19

The Estimated Treatment Effects (γ'_{01}) of RQ1 with Different Sets of U Included in the Outcome Models Predicting MCAS Mathematics

U	$\pi_{1\|0}$	$\pi_{1\|1}$	γ'_{01}		
			Grade 6	Grade 7	Grade 8
u_1	0.20	0.35	0.60	0.34	0.51
u_2	0.20	0.50	0.55	0.29	0.47
u_3	0.20	0.65	0.51	0.25	0.43
u_4	0.20	0.80	0.46	0.20	0.38
u_5	0.35	0.50	0.60	0.34	0.51
u_6	0.35	0.65	0.56	0.29	0.47
u_7	0.35	0.80	0.51	0.25	0.43
u_8	0.50	0.65	0.60	0.34	0.52
u_9	0.50	0.80	0.55	0.29	0.47
u_{10}	0.65	0.80	0.60	0.34	0.52
		$*\ \gamma_{01}$	0.65	0.38	0.56
		s.e.	0.15	0.14	0.15
		One-sided 90% CI	[0.39, 0.65]	[0.16, 0.38]	[0.31, 0.56]

As shown in the table, the original estimated treatment effects are 0.65, 0.38, and 0.56 in Grades 6 to 8, respectively. By including the unobservable *U*, the estimated treatment effects shrunk slightly: the ranges of these estimates are from 0.46 to 0.60,

[9] $*\ \gamma_{01}$ was obtained from random effects models without PS weights and was produced by Stata.

from 0.20 to 0.34, and from 0.38 to 0.52 in Grades 6 to 8, respectively. However, since all these estimates fall within the one- sided 90% confidence intervals of the original ones, one can conclude that the estimated treatment effects are reasonably robust to the presence of the type of hidden bias specified in this study.

Following Diaconu's (2012) approach, the ranges of $\pi_{1|0}$ and $\pi_{1|1}$ was set to be 0.20 to 0.80 with 0.15 as a basic incremental unit. Initially, it was assumed that the magnitude of these conditional probabilities represented the severity of selection bias. For instance, u_1 ($\pi_{1|0} = 0.20$ and $\pi_{1|1} = 0.35$)represented relatively mild selection bias; while u_{10} ($\pi_{1|0} = 0.65$ and $\pi_{1|1} = 0.80$) represented relatively strong selection bias. However, the empirical results suggested that it was the difference between the values of each pair of conditional probabilities that determined the severity of such bias.

As shown in Table 4.19, the estimated treatment effects are approximately the same in each grade if the difference in value between the corresponding conditional probabilities is the same. For instance, γ'_{01} is approximately 0.60, 0.34, and 0.51 in Grades 6 to 8, respectively, for u_1 ($\pi_{1|0} = 0.20$ and $\pi_{1|1} = 0.35$), u_5 ($\pi_{1|0} = 0.35$ and $\pi_{1|1} = 0.50$), and u_{10} ($\pi_{1|0} = 0.65$ and $\pi_{1|1} = 0.80$) because the difference in value between the two conditional probabilities is 0.15 for all these three pairs. Thus u_4 ($\pi_{1|0} = 0.20$ and $\pi_{1|1} = 0.80$) represents relatively strong selection bias, and u_1 ($\pi_{1|0} = 0.20$ and $\pi_{1|1} = 0.35$), u_5 ($\pi_{1|0} = 0.35$ and $\pi_{1|1} = 0.50$), and u_{10} ($\pi_{1|0} = 0.65$ and $\pi_{1|1} = 0.80$) represent relatively mild selection bias of the same degree.

To further establish this point, Table 4.20 presents the detailed results of the outcome models predicting MCAS Mathematics in Grade 6 when the unobservable *U* was included. *U* was simulated based on two pairs of conditional probabilities of *U* given *Z*:

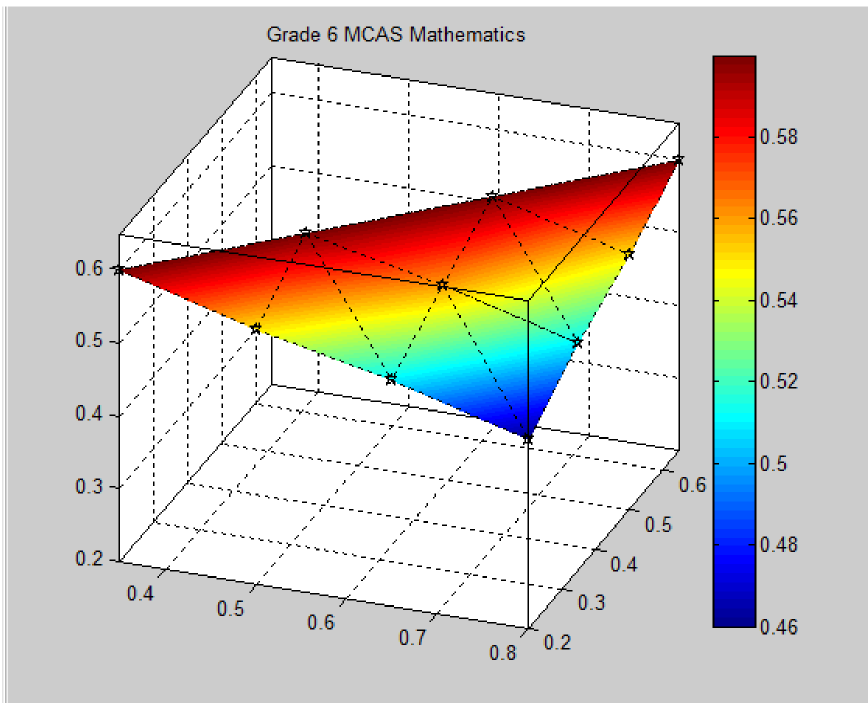
u_1 ($\pi_{1|0}$ = 0.20 and $\pi_{1|1}$ = 0.35) and u_{10} ($\pi_{1|0}$ = 0.65 and $\pi_{1|1}$ = 0.80). As we see, the results obtained using u_1 and u_{10} are nearly identical.

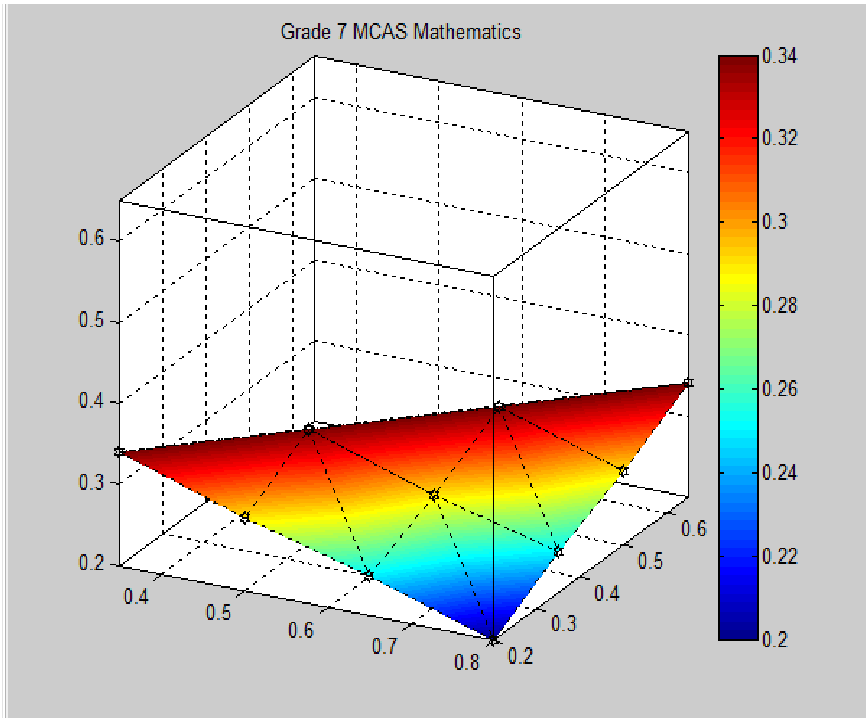
Table 4.20

Results Comparison Using Different Pairs of Conditional Probabilities of U given Z: Results of RQ1 (Outcome Models Predicting MCAS Mathematics in Grade 6 with the Unobserved U Included)

	u_1			**u_{10}**		
Fixed Effects	$\pi_{1\|0}$ = 0.20 and $\pi_{1\|1}$ = 0.35			$\pi_{1\|0}$ = 0.65 and $\pi_{1\|1}$ = 0.80		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	1.30	0.20	0.000	1.11	0.20	0.000
EDose (Ever City Connects)	0.60	0.15	0.000	0.61	0.15	0.000
1-Year vs. 6-Year Dosage	-0.55	0.08	0.000	-0.58	0.08	0.000
2-Year vs. 6-Year Dosage	-0.35	0.08	0.000	-0.36	0.08	0.000
3-Year vs. 6-Year Dosage	-0.29	0.09	0.001	-0.31	0.09	0.000
4-Year vs. 6-Year Dosage	-0.36	0.08	0.000	-0.37	0.08	0.000
5-Year vs. 6-Year Dosage	-0.13	0.09	0.167	-0.16	0.09	0.077
Male	0.03	0.02	0.184	0.02	0.02	0.377
is_Black	-0.19	0.04	0.000	-0.19	0.04	0.000
is_Asian	0.37	0.05	0.000	0.37	0.05	0.000
is_Hispanic	-0.07	0.04	0.081	-0.07	0.04	0.101
is_Other	-0.12	0.08	0.124	-0.11	0.08	0.177
Bilingual	-0.05	0.03	0.059	-0.05	0.03	0.087
Special Needs 2	-0.20	0.04	0.000	-0.20	0.04	0.000
Special Needs 3	-0.59	0.04	0.000	-0.59	0.04	0.000
Reduced Lunch	-0.16	0.07	0.020	-0.17	0.07	0.017
Free Lunch	-0.26	0.04	0.000	-0.25	0.04	0.000
Foreign Born	0.15	0.04	0.000	0.15	0.04	0.000
RC_Reading_gr1	0.06	0.02	0.002	0.06	0.02	0.002
RC_Math_gr1	0.17	0.02	0.000	0.17	0.02	0.000
RC_Writing_gr1	0.04	0.02	0.040	0.04	0.02	0.036
RC_WorkHabit_gr1	0.09	0.02	0.000	0.09	0.02	0.000
RC_Behavior_gr1	-0.01	0.02	0.588	-0.01	0.02	0.429
RC_Effort_gr1	0.00	0.02	0.957	0.00	0.02	0.906
Age_gr1	-0.16	0.03	0.000	-0.15	0.03	0.000
Distance from School_gr1	0.00	0.01	0.490	0.00	0.01	0.544
# School Moves_gr1	0.00	0.02	0.849	0.02	0.02	0.474

Furthermore, using MATLAB 8.4 (The Mathworks Inc., 2014), Figure 4.6 displays the 3-dimentional response surface graphing the 10 pairs of conditional probabilities on the x and y axes (the x axis represents $\boldsymbol{\pi_{1|1}}$ and the y axis represents $\boldsymbol{\pi_{1|0}}$) against the estimated treatment effects (γ'_{01}) on the z axis for each of the outcome models predicting MCAS Mathematics in Grades 6 to 8, given $\beta_{Uk} = 0.3$.





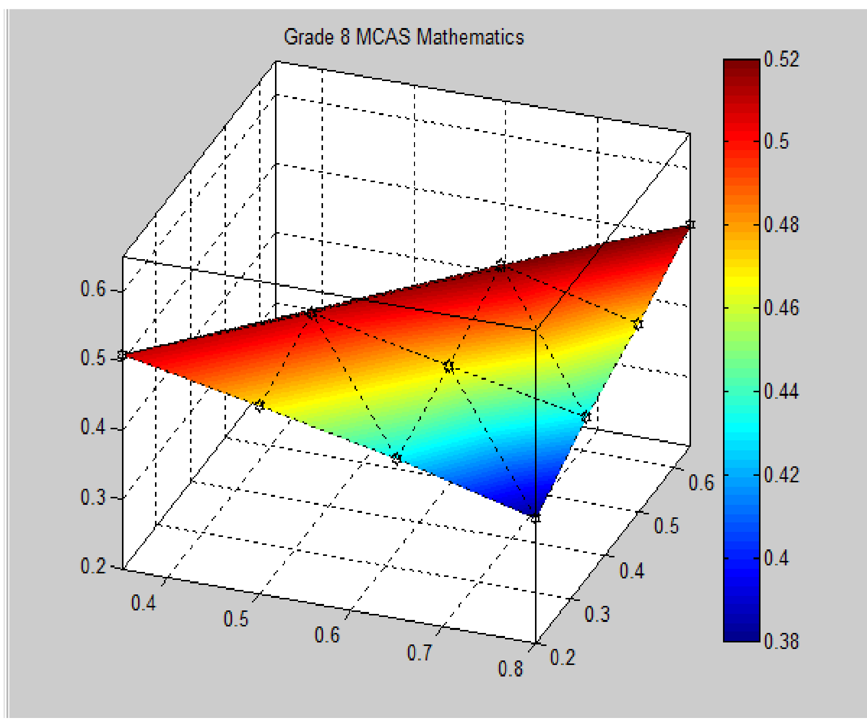


Figure 4.6 Sensitivity analysis results for the outcome models predicting MCAS Mathematics in Grades 6 to 8: response surface of the estimated treatment effects (γ_{01}') as a function of conditional probabilities, for $\beta_{Uk} = 0.3$

Generally speaking, the shape of the response surfaces in Figure 4.6 can be considered shallow: it is not surprising to see that when altering the relationship between $\pi_{1|0}$ and $\pi_{1|1}$ to the maximum (e.g., $\pi_{1|0}$ = 0.20 and $\pi_{1|1}$ = 0.80) to represent relatively strong selection bias, the estimated treatment effects were reduced the most (i.e., to 0.46, 0.20, and 0.38 in each grade); whereas if the relationship between the two parameters is set to the minimum (i.e., $\pi_{1|0}$ = 0.20 and $\pi_{1|1}$ = 0.35 or $\pi_{1|0}$ = 0.65 and $\pi_{1|1}$ = 0.80) to represent relatively mild selection bias, the estimated treatment effects were reduced the least (to 0.60, 0.34, and 0.51 in each grade for the pair of $\pi_{1|0}$ = 0.20 and $\pi_{1|1}$ = 0.35; and 0.60, 0.34, and 0.52 in each grade for the pair of $\pi_{1|0}$ = 0.65 and $\pi_{1|1}$ = 0.80). Nevertheless, the estimates did not deviate too much from the original ones (γ_{01}). As a result, when considering the surfaces in Figure 4.6, it is reasonable to consider them being shallow.

As discussed in Chapter 3, a shallow surface indicates mild sensitivity of the estimated treatment effects to the presence of hidden bias. Therefore, one can conclude that the estimated treatment effects of the City Connects elementary intervention on middle school academic achievement as measured by standardized MCAS Mathematics scores are robust to the presence of some forms of hidden bias.

4.3.4 Results Predicting MCAS ELA

Table 4.21 presents the estimated treatment effects of the City Connects elementary intervention on MCAS ELA in Grades 6 to 8 (γ'_{01}) when the outcome models included different sets of *U* (the chosen 10 pairs of *U* based on the conditional probabilities of *U* given *Z* as listed in Table 3.3). In addition, the ones estimated from models that excluded U ($*\gamma_{01}$[10]), together with the corresponding standard errors and the 95% confidence interval, are shown at the bottom of the table for comparison.

Table 4.21

The Estimated Treatment Effects (γ'_{01}) of RQ1 with Different Sets of U Included in the Outcome Models Predicting MCAS ELA

U	$\boldsymbol{\pi_{1\|0}}$	$\boldsymbol{\pi_{1\|1}}$	γ'_{01}		
			Grade 6	Grade 7	Grade 8
u_1	0.20	0.35	0.35	0.16	0.31
u_2	0.20	0.50	0.31	0.11	0.27
u_3	0.20	0.65	0.26	0.07	0.23
u_4	0.20	0.80	0.22	0.02	0.18
u_5	0.35	0.50	0.35	0.16	0.32
u_6	0.35	0.65	0.31	0.11	0.27
u_7	0.35	0.80	0.26	0.07	0.23
u_8	0.50	0.65	0.35	0.16	0.32
u_9	0.50	0.80	0.31	0.11	0.27
u_{10}	0.65	0.80	0.35	0.16	0.32
		$*\gamma_{01}$	0.40	0.20	0.36
		s.e.	0.14	0.11	0.13
		One-sided 90% CI	[0.17, 0.40]	[0.01, 0.20]	[0.15, 0.36]

As shown in the table, the original estimated treatment effects are 0.40, 0.20, and 0.36 in Grades 6 to 8, respectively. By including the unobservable U the estimated treatment effects shrunk slightly: the ranges of these estimates are from 0.22 to 0.35, from 0.02 to 0.16, and from 0.18 to 0.32 in Grades 6 to 8, respectively. The lowest estimated treatment effect of City Connects with U included is 0.02 for the outcome

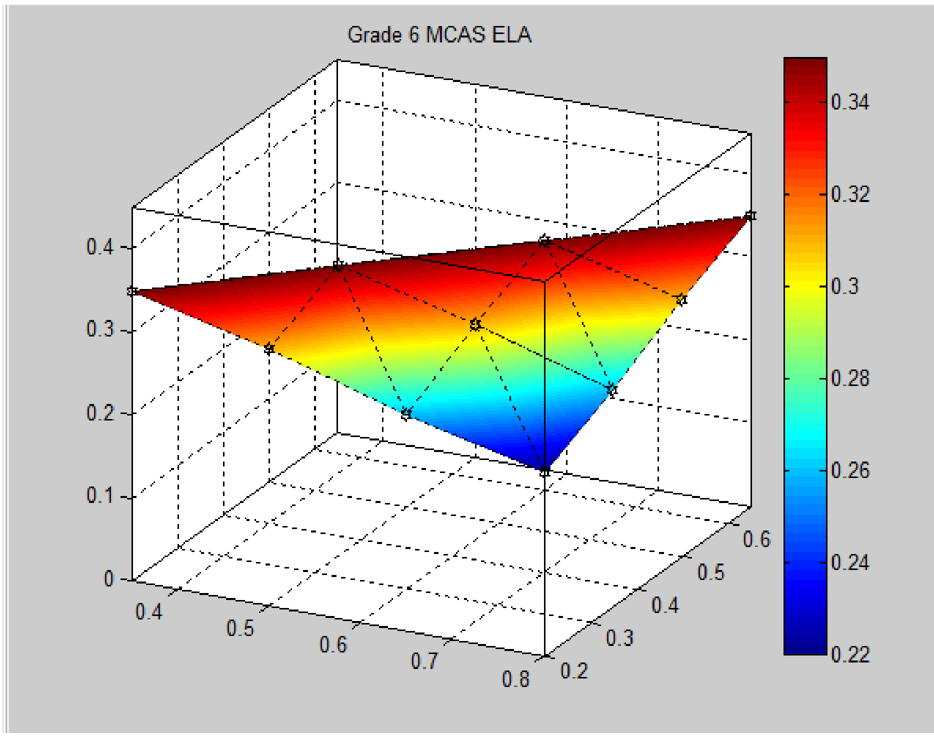
[10] $*\gamma_{01}$ was obtained from random effects models without PS weights and was produced by Stata.

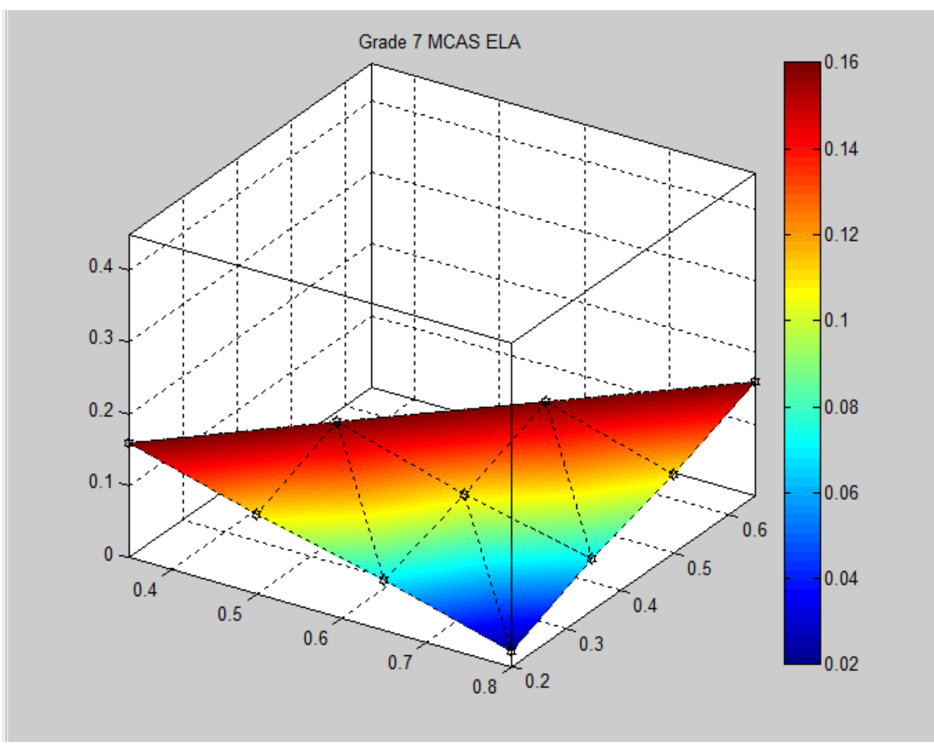
model predicting Grade 7 ELA with $\pi_{1|0}$ = 0.20 and $\pi_{1|1}$ = 0.80 (the presence of a relatively strong selection bias), which is quite small. However, since all these ranges fall within the one-sided 90% confidence intervals of the original estimates, one can conclude that the estimated treatment effects are robust to the presence of hidden bias specified in this study.

Furthermore, Figure 4.7 displays the 3-dimentional response surface graphing the 10 pairs of conditional probabilities on the x and y axes (the x axis represents $\pi_{1|1}$ and the y axis represents $\pi_{1|0}$) against the estimated treatment effects (γ'_{01}) on the z axis for each of the outcome models predicting MCAS ELA in Grades 6 to 8, given $\beta_{Uk} = 0.3$.

Generally speaking, the shape of the response surfaces in Figure 4.7 can be considered shallow: it is not surprising to see that when altering the relationship between $\pi_{1|0}$ and $\pi_{1|1}$ to the maximum (e.g., $\pi_{1|0}$ = 0.20 and $\pi_{1|1}$ = 0.80) to represent relatively strong selection bias, the estimated treatment effects were shrunk the most (i.e., to 0.22, 0.02, and 0.18 in each grade); whereas if the relationship between the two parameters is set to the minimum (i.e., $\pi_{1|0}$ = 0.20 and $\pi_{1|1}$ = 0.35 or $\pi_{1|0}$ = 0.65 and $\pi_{1|1}$ = 0.80) to represent relatively mild selection bias, the estimated treatment effects were shrunk the least (to 0.35, 0.16, and 0.31 in each grade for the pair of $\pi_{1|0}$ = 0.20 and $\pi_{1|1}$ = 0.35; and 0.35, 0.16, and 0.32 in each grade for the pair of $\pi_{1|0}$ = 0.65 and $\pi_{1|1}$ = 0.80). Nevertheless, the estimates did not deviate too much from the original ones (γ_{01}). As a result, when considering the surfaces in Figure 4.7, it is reasonable to consider them being shallow.

Therefore, one can conclude that the estimated treatment effects of the City Connects elementary intervention on middle school academic achievement as measured by both standardized MCAS Mathematics and ELA scores are robust to the presence of some forms of hidden bias.





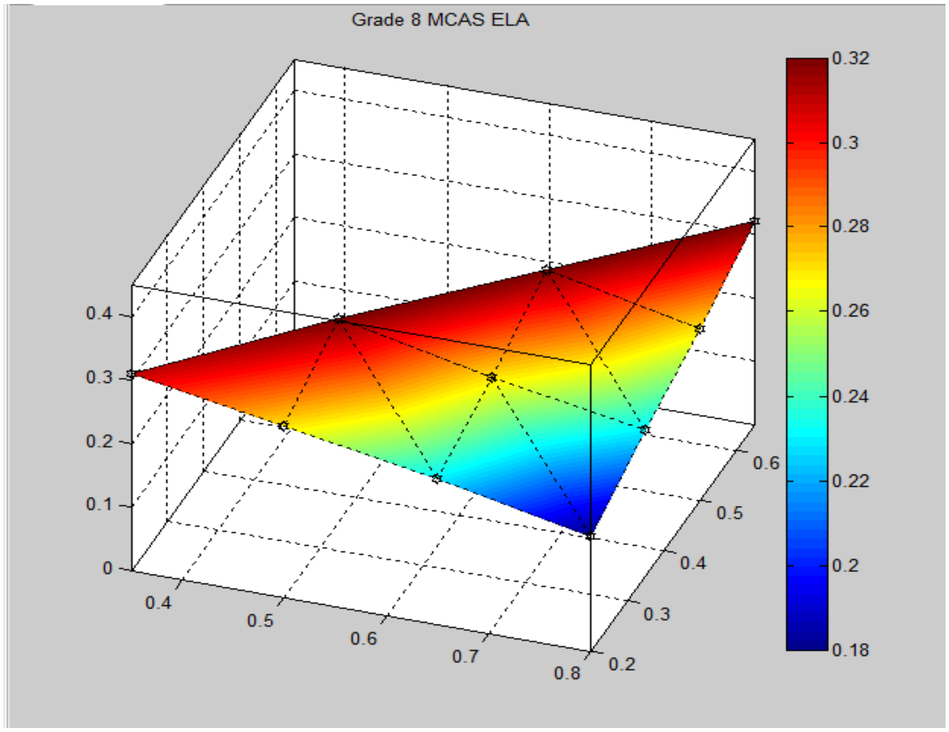


Figure 4.7 Sensitivity analysis results for the outcome models predicting MCAS ELA in Grades 6 to 8: response surface of the estimated treatment effects (γ'_{01}) as a function of conditional probabilities, for $\beta_{Uk} = 0.3$

CHAPTER 5. CONCLUSIONS

5.1 Summary of Findings

In order to provide a comprehensive evaluation of City Connects treatment effects on middle school academic performance using student longitudinal records, parallel analyses were conducted to evaluate both City Connects elementary and middle school interventions.

The results for RQ1 show that students who were exposed to the City Connects elementary intervention outperformed their counterparts, who graduated from the comparison elementary schools, on academic achievement in all middle school grades. It is believed that by immediately and continually addressing each student's strengths and needs, City Connects is a long-term intervention that over time provides sufficient student support and creates a positive environment at school that is conducive to the improvement of students' academic achievement. The process is gradual and time-consuming. With several years of exposure to City Connects, significant impact on academic achievement can be expected and the impact does not fade easily. In the case of RQ2, since all City Connects students only received a maximum of one year[11] of City Connects middle school intervention, it is still too soon to expect any significant changes. As students' exposure to City Connects in middle schools lengthens, it is plausible that the impact of City Connects will manifest itself more strongly.

Furthermore, the robustness of the estimated treatment effects of the City Connects elementary schools on middle school academic outcomes to the presence of

[11] CCNX treatment dosage was cumulated based on semesters and it was possible for a student to have less than 1 year of dosage if the student joined a CCNX middle school in the middle of an academic year or transferred to a non-CCNX middle school after attending a CCNX middle school for one semester.

unobserved selection bias was examined through a sensitivity analysis. The results indicated that the estimated treatment effects of the City Connects elementary intervention on middle school academic achievement as measured by both standardized MCAS ELA and Mathematics scores are only mildly sensitive to the presence of some forms of hidden bias.

5.1.1 RQ1

In order to estimate the unique contributions of the City Connects elementary schools on middle school outcomes, overall differences in academic achievement among middle school effects had to be taken into account. For every sampled student, the outcome score for a grade-specific subject was adjusted based on the middle school he or she was currently enrolled in. These adjustments were estimated through two-level linear regression models predicting outcomes of interest with current middle schools as clusters. Only data from comparison students were used in these adjustment models so that if receiving City Connects in elementary school helped prepare a student to get into a better middle school, this effect would not be removed by adjusting the middle school effectiveness. Meanwhile, in order to address overt selection bias, the PS weighting method was used to remove the observed differences with respect to academic outcomes and key demographic characteristics at the baseline grade. Finally, two-level linear regression models with last City Connects or comparison elementary schools that students attended as clusters were built to examine the effectiveness of the City Connects elementary school intervention on middle school academic outcomes.

The total percent of variance explained by each middle school adjustment model is generally quite large (ranging from 45.3% to 68.2% for MCAS measures and from

19.6% to 44.7% for weighted GPA). However, it is noticeable that the total percent of variance explained is quite a bit lower for weighted GPA than for MCAS scores. A plausible explanation is that GPAs are norm-referenced measures and are graded by subject teachers within each school; therefore, they are not as strongly correlated with different middle school attendance as MCAS measures are.

The results of the outcome models showed that students who received the City Connects treatment in elementary school scored significantly higher on all three outcomes in all middle school grades. The magnitude of such positive effects was the largest in Grade 6, dropped a little bit in Grade 7, and became larger again in Grade 8. The corresponding effect sizes were also large, ranging from 0.29 to 0.67 for MCAS measures and from 0.40 to 0.67 for weighted GPA scores, indicating that the estimated treatment effects of the City Connects elementary intervention on middle school academic outcomes were not only statistically significant but also practically important.

In addition, residual-based estimates for the City Connects elementary schools were saved and graphed to examine the relative standing of each City Connects elementary school in comparison to all the elementary schools in the sample. It is really important for City Connects to investigate the relative performance of each City Connects school so that targeted enhancement plans can be made to help low-performing City Connects schools improve and best practices of high-performing City Connects schools can be studied further.

Standard deviations of these residual-based estimates of City Connects elementary schools were compared with the magnitude of the estimated treatment effects. Generally speaking, the former were smaller than the latter. In other words, City

Connects elementary schools did not differ more from each other than they were different from the comparison schools. Therefore, it was concluded that City Connects elementary schools' overall positive treatment effects on academic achievement in middle school were not driven by some exceptional schools.

From a general perspective, it is relatively easy to differentiate extremely-low performing and extremely-high performing schools and their relative standings across grades are, in most cases, quite stable. However, average-performing schools are quite indistinguishable, which is still a major challenge in using VAMs to evaluate school effectiveness.

Additionally, the residual index statistics was calculated as a measure of "reliability" to evaluate how large the within-school variance was as relative to the variance of the estimated school effects. It indicated that Grade 6 outcome models had the best model fit, followed by the ones in Grade 7 and 8.

In the last phase of analysis, school-level covariates were included in the outcome models to examine the extent to which the estimated treatment effects could be accounted for by school characteristics. Variance partitions at the school level indicated that the incremental contribution of school-level covariates was typically smaller (from 2.7% to 13.6% for MCAS measures and from 3.7% to 5.3% for the weighted GPA) than that of attending City Connects elementary schools (from 2.1% to 18% for MCAS measures and from 7.5% to 27.6% for the weighted GPA). Furthermore, in terms of specific influential school-level covariates, smaller class size and lower percent of students who were eligible for free-or reduced-priced lunch were associated with higher achievement.

Chapter 5. Conclusions

5.1.2 RQ2

Similarly, the City Connects middle school effects on Grade 6 academic outcomes were estimated through two-level linear regression models with the current middle schools that students attended as clusters. PS weights were applied to remove overt selection bias at the end of elementary school. Baseline achievement and key student characteristics were included as covariates. It was not surprising to see that the City Connects middle school intervention did not have a great impact on students' academic achievement measures in Grade 6 because City Connects students who only received the City Connects middle school intervention could receive a maximum of one year of City Connects by the end of Grade 6 and those City Connects students who received both the City Connects elementary and middle school interventions had their prior achievement gains by attending City Connects elementary schools removed through statistical adjustments.

To further differentiate the unique contribution of attending City Connects middle schools alone and that of attending both City Connects elementary and middle schools, a binary variable indicating whether or not a City Connects middle school student had also attended City Connects elementary schools was included in the outcome models. The results showed that those who received only the City Connects middle school intervention scored higher than the comparison group. Furthermore, those who received both City Connects elementary and middle school interventions scored slightly higher than those who received the latter only. However, in most of the cases none of these City Connects effects were statistically significant.

Standard deviations of the residual-based estimates were generally larger than the magnitude of the estimated treatment effects, which indicated that City Connects middle schools differed as much or more from each other than they did from the comparison schools in terms of their students' academic achievement in Grade 6. Moreover, the outcome models to answer RQ2a were considered to have good model fit based on the residual index statistics.

Finally, school-level covariates were included in the outcome models. Influential indicators at the school level that were positively associated with achievement included: more students per computer, lower percent of students whose first language was not English, and lower student/teacher ratio. Interestingly, with the inclusion of such characteristics, the estimated treatment effects became statistically significant for some outcome measures (i.e., for MCAS Mathematics and weighted GPA). It seems that the City Connects middle school intervention may have had an effect on achievement, once observed differences in school composition and resources were taken into account. Generally speaking, City Connects schools enrolled significantly more ELL and foreign-born students in Grade 6 (as shown in Table 3.8b) and had slightly fewer students per computer. When these disadvantaged differences were removed by statistical adjustments, a positive City Connects effect on academic achievement started to occur.

5.1.3 RQ3

A sensitivity analysis was conducted to assess the robustness of the estimated treatment effects (obtained for RQ1a) to the presence of hidden selection bias. Sets of binary variable U that met two key assumptions were randomly generated using Monte Carlo simulation. The first assumption dealt with how U was related to Z, the indicator of treatment assignment: by altering the magnitude of the conditional probabilities of *U* given *Z*, 10 sets of such probabilities were chosen. The second assumption pre-determined the magnitude of the relationship between *U* and the outcomes. Based on empirical results of RQ1a, 0.3 was chosen as the magnitude of the corresponding regression coefficient, because it was the largest regression coefficient associated with (observed) covariates across subjects and grade levels. The newly generated sets of *U* were then added into the outcome models to examine how the estimated treatment effects would be affected by the inclusion of *U*. This procedure was repeated 100 times and the resulting treatment effect estimates were averaged over the 100 trials.

The results showed that the estimated treatment effects for both MCAS Mathematics and ELA were reduced slightly with the inclusion of *U*; however, the fact that they still fell within the one-sided 90% confidence intervals of the original ones indicated only a mild sensitivity to hidden bias. In addition, the higher the strength of the selection bias, as partly indicated by the mathematical difference between each pair of the conditional probabilities of *U* given *Z*, the smaller the estimated treatment effects.

5.2 Limitations and Future Research

5.2.1 Limitations on Statistical Models Used in Estimating School Effectiveness

For RQ1, both the selection models (models that generated PS weights) and the outcome models (models to estimate treatment effects) have some features that can be further improved. To start with, the selection models have issues with both the baseline time point chosen and the application of PS weights. First, the baseline time point chosen is the beginning of Grade 1; however, City Connects serves kindergarten as well, so the baseline is not a true pre-intervention time point[12]. Suppose City Connects has a positive effect on its' students in kindergarten, then by making students of the treatment and the comparison groups statistically equivalent on observed covariates at the beginning of Grade 1, the City Connects treatment effect was underestimated. Slight underestimation of the treatment effect seems not to be a serious issue since the study has already demonstrated a strong positive effect of City Connects; however, although quite unlikely, if City Connects has somehow lowered students' achievement in kindergarten, then the treatment effect was overestimated, which is problematic. A study on students' performance on kindergarten academic outcomes, no matter how small the sample size is and how limited the assessment tools are, should be examined closely to at least rule out such possibility. City Connects has initiated such an analysis in the past, but it is an on-going investigation.

Second, schools participated in City Connects as units, but the current PS weights were generated based on student characteristics. Research on estimating PS in a multilevel model and then applying them to estimate treatment effects has just started to mushroom. The approach not only reduces selection bias as the traditional PS generated from a single-level model do, but also addresses the bias associated with random effects across units. For example, in order to estimate the effectiveness of kindergarten retention policy on kindergarten reading achievement, Hong and Raudenbush (2005) estimated PS through a multi-level logistic regression model and then used them to stratify the analytic sample. The treatment effect was then estimated within each stratum and the results were summarized. In addition, Xiang and Wang (2013) applied the same approach to estimate

[12] Approximately 30% of the total sample of the City Connects students started the intervention in Kindergarten.

the effectiveness of charter schools on student achievement and growth in Grade 6. In view of such studies, future research should be considered to generate and apply PS through a more appropriate approach that takes into account the clustered data structure.

For outcome models of RQ1, school clusters were defined as the last City Connects or comparison schools students attended: this definition confounds the unique contribution of the last elementary school that a student attended with that of all elementary schools that he or she attended. For instance, a student may spend most of his or her elementary school years in one City Connects or comparison school and then transfer to another City Connects school. In this case, based on the definition of City Connects, he or she is categorized as a City Connects student and his or her last City Connects school attended will be the cluster that is held accountable for the student's achievement. The contribution of the last elementary school, together with that of all the schools this student had attended during the elementary school years, are attributed to one school, which is certainly problematic when the focus is on evaluating the effectiveness of each school. It is worth mentioning that the outcome models in this study did include the number of years in City Connects (Dosage) as a predictor; however, the cumulative number of years in City Connects did not differentiate the number of years in one City Connects school versus that in another City Connects school. Therefore, Dosage does not directly address this problem.

In terms of the residual-based estimate of each school, comparisons across grades can be visualized; however, both MCAS scores and GPA points across grades are not vertically scaled, so the comparisons are merely suggestive.

Finally, to discern longitudinal patterns, the study followed the same students progressing through middle school grades. Table 4.7 showed that the analytic sample size for each outcome measure generally reduced by one third when progressing to the next grade. This is partly due to losing one more cohort per grade in Grade 7 and 8 and partly due to students transferring to other schools. The former is a design issue and may not be a serious threat. The latter, however, demands some further study.

For instance, the Grade 6 analytic sample comprised cohorts 2000-2006, but the Grade 7 analytic sample comprised cohorts 2000-2005 because cohort 2006 has not yet matriculated to Grade 7 by the end of 2012-2013 academic school year. In the same vein, the Grade 8 sample only comprised cohorts 2000-2004. Furthermore, to ensure comparability across grades in terms of the effectiveness of each elementary school, students who transferred to another school in later grades were subsequently dropped from the sample. If there are systematic differences on academic achievement for students who transferred and who stayed, the estimated treatment effects in later grades maybe biased.

For RQ2, the study showed that exposure to the City Connects middle school intervention alone did not have a significant effect on middle school outcomes. However, it should be noted that only Grade 6 outcomes were examined and by the end of Grade 6 students can only receive up to one year of City Connects. With the City Connects intervention gradually extending to latter middle school grades and the number of middle schools that joining City Connects grows, these research questions need to be revisited.

5.2.2 Future Research on Sensitivity Analysis

For RQ3, more studies are needed to improve upon the sensitivity analysis conducted in this study: first, in addition to simply drawing descriptive graphs as in Figures 4.6 and 4.7, inferential statistics to measure the severity of hidden bias should be developed.

Second, the current study only simulated 100 trials for each of the pre-determined 10 sets of unobserved U; with higher computer capacities available, more trials should be conducted to get more accurate estimates.

Third, the resulting estimated treatment effects with U included displayed a repetitive pattern due to the way the sets of U were chosen (i.e., using 0.15 as an incremental unit). In the future, random pairs of U should be generated as long as they meet the corresponding assumptions about U given Z. Furthermore, a stratified sampling method can be used when drawing these random pairs: the random selection will occur within each stratum that represents mild, medium, and high sensitivity (i.e., the difference in values of the conditional probabilities of U given Z is small, medium, and large).

Fourth, multiple pre-determined regression coefficients associated with U should be chosen to examine the impact of altering the magnitude of the relationship between U and the outcome on the estimated treatment effects.

Finally, there are six assumptions when using VAMs to make causal argument. Only one of them (i.e., strong ignorability assumption) was tested through sensitivity analysis. Future studies are needed to examine all the applicable assumptions to justify the causal argument that attending City Connects invention leads students to prosper academically in a long run.

5.3 Final Remarks

The current study addressed the question of whether the City Connects intervention, both the elementary and the middle school versions, was effective in raising students' achievement in middle school. Although it is certainly important to answer this question, how City Connects managed to support students so well to face the academic challenges in middle school is a more critical question. To answer this question, future within-City Connects studies should be conducted.

First, the current treatment indicators, a binary Dose (ever City Connects) and a continuous Dosage (years spent with City Connects), did not fully specify whether a student stayed in a particular City Connects school for a really short period of time or for his or her entire elementary grades (Dosage cumulated years in City Connects but did not differentiate years with different City Connects schools) and the effect of entering City Connects at the specific grade (e.g., was City Connects more effective in latter elementary school grades than in earlier grades?). Although including years spent with City Connects at level 1 alleviated the problem, Dosage alone is far from enough.

Second, results of RQ1a showed that the maximum treatment effects occurred when a student received maximum years of dosage (i.e., six years) since a majority of the dosage dummies were associated with negative regression coefficients and these were in ascending order. Thus, although it may be reasonable to assume that it takes several years for City Connects to truly make an impact on students' academic achievement, the exact number of years for City Connects to cross the threshold of significance testing and to finally achieve such goal is still unclear. The specific question may deserve further study because it will help City Connects develop a more productive timeline for evaluation activities. City Connects is expanding rapidly to numerous districts and cities and it is important to have some evidence on roughly the number of years of exposure by which one should expect significant or substantively meaningful treatment effects to occur.

Third, the full nature of the City Connects intervention cannot be completely represented by the two treatment indicators. The Student Support Information System (SSIS) collects rich data on individual student plans, service referrals, and information on service providers. These data record key information on how the School Site Coordinator collaborates with parents, teachers, and other educators to evaluate each student's strengths and weaknesses, to develop a tailored remedial/enrichment plan, and then to link students to various services, and on how the implementation team tracks the fidelity of the implementation and provide consistent follow-ups. To truly understand why the City Connects treatment works as manifested in this study, the SSIS data deserve a separate well-constructed study.

REFERENCES

Adelman, H., & Taylor, L. (2005). *The School Leader's Guide to Student Learning Supports: New Directions for Addressing Barriers to Learning*. Thousand Oaks, CA: Corwin Press.

Allison, P.D. (2009). *Fixed Effects Regression Models.* Thousand Oaks, CA: Sage Publications.

Amatea, E. S., & Clark, M. A. (2005). Changing schools, changing counselors: A qualitative study of school administrators' conceptions of the school counselor role. *Professional School Counseling, 9*(1), 16-27.

American School Counselor Association. (2003). *The ASCA national model: A framework for school counseling programs.* Alexandria, VA: Author.

An, C., Lee-St. John, T., Raczek, S., Walsh, M. E., & Madaus, G. (2014). *Estimating the impact of optimized student support in Elementary School on Boston Exam School attendance.* Poster session presented at the annual meeting of the American Educational Research Association, Philadelphia, Pennsylvania.

An, C. & Wong, Y. (2012). *Appendix E of The impact of City Connects: Annual report 2012*. Chestnut Hill, MA: Boston College.

Anderson-Butcher, D., Lawson, H. A., Bean, J., Flaspohler, P., Boone, B., & Kwiatkowski, A. (2008). Community collaboration to improve schools: Introducing a new model from Ohio. *Children & Schools, 30*(3), 161-172.

Aronson, J. Z. (1996). How schools can recruit hard-to-reach parents. *Educational Leadership, 53*(7), 58-60.

Ayoub, C. C., Fischer, K. W. (2006). Developmental pathways and intersections among domains of development. In K. McCartney & D. Phillips (Eds.), *Handbook of early child development* (pp. 62-82). Oxford: Blackwell.

Ballou, D. (2008). Test scaling and value-added measurement. *Education Finance and Policy, 4*, 351-383.

Becker, B. E., & Luthar, S. S. (2002). Social-emotional factors affecting achievement outcomes among disadvantaged students: Closing the achievement gap. *Educational Psychologist, 37*(4), 197-214.

Berk, R. A. (2004). *Regression analysis: A constructive critique*. Thousand Oaks, CA: Sage Publications, Inc.

References

Boston Public Schools. (2013). *Student Assignment Policy*. Retrieved September 10, 2013, from http://www.bostonpublicschools.org/assignment.

Barton, P. E. (2004). Why does the gap persist? *Educational Leadership, 62*(3), 8-13.

Berliner, D. C. (2009). *Poverty and Potential: Out-of-School Factors and School Success*. Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved January 23, 2014, from http://epicpolicy.org/publication/poverty-and-potential.

Braun, H. I. (2005) Using student progress to evaluate teachers: A primer on Value-Added *Models.* Princeton, New Jersey: ETS Policy Information Center.

Briggs, D. C. (2008). *The Goals and Uses of Value-Added Models.* Paper prepared for the workshop of the Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability, National Research Council, Washington, DC, November 13-14. Retrieved September 1, 2013, from: http://www7.nationalacademies.org/bota/VAM_Workshop_Agenda.html.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

City Connects. (2011). *The impact of City Connects: Annual report 2010*. Chestnut Hill, MA: Boston College.

City Connects. (2013). *City Connects 2013 Practice Manual*. Chestnut Hill, MA: Boston College.

Clarke, P., Crawford, C., Steele, F., & Vigoles, A. (2010). *The choice between fixed and random effects models: Some considerations for educational research.* The Institute for the Study of Labor. Bon, Germany: Deutche Post Foundation.

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

Cook, T.D., & Shadish, W.R. (2012). *QE Workshop Day 5* [PowerPoint slides]. Retrieved from http://www.ipr.northwestern.edu/workshops/past-workshops/quasi-experimental-design-and-analysis-in-education/2012/docs/Day-5-2012.pdf

Columbo, G. (1995). Parental involvement: A key to successful schools. *NASSP Bulletin, 79*(567), 71-75.

References

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., & Wynder, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute, 22*, 173–203.

de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory.* New York, NY: The Guilford Press.

Diaconu, D. V. (2012). *Modeling science achievement differences between single-sex and coeducational schools: Analyses from Hong Kong, SAR and New Zealand from TIMSS 1995, 1999, and 2003*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. [UMI3521765])

Epstein, J. L. (1995, May). School/family/community partnerships: Caring for the children we share. *Phi Delta Kappan, 76*(9), 701-712.

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review, 13*(1), 1-22.

Fisher, R. A. (1935). *The design of experiments*. Oxford, England: Olive & Boyd.

Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School Class Size: Research and Policy.* Beverly Hills, California: Sage Publications.

Guo S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Los Angeles: Sage Publications, Inc.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? Journal of *Applied Econometrics, 18*(5), 527-544.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory.* Newbury Park, CA: Sage Press.

Harder, V. S., Stuart, E. A., Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychol Methods, 15*(3), 234-249.

Harrington, M. (1962). *The Other America: Poverty in the United States*. New York, NY: Simon & Schuster Inc.

Heageland, T. and L. Kirkeboen. (2008). *School performance and value-added indicators - What is the importance of controlling for socioeconomic background?: A simple empirical illustration using Norwegian data.* Retrieved July 23, 2013, from http://www.ssb.no/a/english/publikasjoner/pdf/doc_200808_en/doc_200808_en.pdf.

References

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

Heller, R., Calderon, S., & Medrich, E. (2003). *Academic achievement in the middle grades: What does research tell us?* Atlanta, GA: Southern Regional Education Board.

Hong, G., & Raudenbush, S.W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*(3), 205-224.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika, 87*(3), 706-710.

Johnson, K. A. (2000). *The peer effect on academic achievement among public elementary school students* (Report No. CDA00-06). Washington, DC: The Heritage Foundation. (ERIC Document Reproduction Series No. ED442916)

Jeynes, W. H. (2007, January). The relationship between parental involvement and urban secondary school student academic achievement: A meta-analysis. *Urban Education, 42*(1), 82-110.

Kennedy, P. (2003). A guide to econometrics (5th ed.). Cambridge: MIT Press.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Lee-St. John, T. (2013). Doctoral dissertation, Boston College, Lynch School of Education, 2013.

Lockwood, J. R., & McCaffrey, D. F. (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics, 1*, 223-252.

Loucks, H. (1992, April). Increasing parent/family involvement: Ten ideas that work. *NASSP Bulletin, 76*(543), 19-23.

Massachusetts Department of Elementary and Secondary Education. (2007). *MCAS 2007 technical report.* Retrieved October 10, 2013, from http://iservices.measuredprogress.org/documents/MA/Technical%20Report/2007/04-09-08%202007%20MCAS%20Tech%20Rpt%20Final%20PDF.pdf

Massachusetts Department of Elementary and Secondary Education. (2012). *Spring 2012 MCAS tests: Summary of state results.* Massachusetts: Chester, M.D.

References

McCaffrey, D. F., & Lockwood, J. R. (2008). *Value-added Models: Analytic issues*. Paper prepared for the workshop of the Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability, National Research Council, Washington, DC, November 13-14. Retrieved September 1, 2013, from: http://www7.nationalacademies.org/bota/VAM_Workshop_Agenda.html.

Michalopoulos, C., Bloom, H. S., & Hill, C. J. (2004). Can propensity score methods match the findings form a random assignment evaluation of mandatory welfare-to-work programs? *Review of Economics and Statistics, 86*, 156-179.

Montgomery, M. R., Richards, T., & Braun, H. I. (1986). Child Health, Breast-Feeding, and Survival in Malaysia: A Random-Effects Logit Approach. *Journal of the American Statistical Association, 81*(394), 297-309.

Mooney, C. Z. (1997). *Monte Carlo Simulation*. Thousand Oaks, CA: Sage Publications.

National Middle School Association. (2003). *This We Believe: Successful Schools for Young Adolescents: A Position Paper of the National Middle School Association*. National Middle School Association.

National Research Council. (2010). Getting value out of value-added: Report of a workshop. Braun, H., Chudowsky, N. & Koenig, J. (Eds.). Washington, D.C.: The National Academies Press.

No Child Left Behind Act of 2001. (2002). Pub. L. No. 107-110, 115 Stat. 1425, 20 U.S.C. § 6301 et seq.

Noguera, P. A. (2011). A broader and bolder approach uses education to break the cycle of poverty. *Phi Delta Kappan, 93*(3), 9-14.

OECD. (2008). *Measuring improvements in learning outcomes: Best practices to assess the value-added of schools.* Paris, France: OECD.

OECD. (2013). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know*. Paris, France: OECD.

Ohio Mental Health Network for School Success. (2004). *OMNSS Information Brief: Non-Academic Barriers to Learning*. Retrieved February 13, 2014, from: http://www.units.muohio.edu/csbmhp/network/barriers.pdf.

Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and Science achievement of U.S. fourth- and eighth-grade students in an international context*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

References

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics, 29*(1), 121-129.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20*(4), 307-335.

Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy, 4*(4), 492-519.

Rosenbaum, P. R. (1986). Dropping out of high school in the United States: An observational study. *Journal of Educational Statistics, 11*(3), 207-224.

Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation tests in matched observational studies. *Biometrika, 74*(1), 13-26.

Rosenbaum, P. R. (1988). Sensitivity analysis for matching with multiple controls. *Biometrika, 75*(3), 577-581.

Rosenbaum, P. R. (1989). Sensitivity analysis for matched observational studies with many ordered treatments. *Scandinavian Journal of Statistics, 16*(3), 227-236.

Rosenbaum, P. R. (1991a). Discussing hidden bias in observational studies. *Annals of Internal Medicine, 115*(11), 901- 905.

Rosenbaum, P. R. (1991b). Sensitivity analysis for matched case-control studies. *Biometrics, 47*(1), 87-100.

Rosenbaum, P. R., & Krieger, A. M. (1990). Sensitivity of two-sample permutation inferences in observational studies. *Journal of the American Statistical Association, 85*(410), 493-498.

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

Rosenbaum, P. R., & Rubin, D. B. (1983b). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological), 45*(2), 212-218.

Rothstein, J. M. (2007). Do value-added models add value? Tracking, fixed effects, and causal inference. *CEPS Working Paper*, No. 159.

References

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127*, 757-763.

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

Rubin, D.B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied statistics, 2*(3), 808-840.

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2003). A potential outcome view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29*(1), 103-116.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee Value-Added Assessment System: A quantitative outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp. 137-162). Thousand Oaks, California: Corwin Press, INC.

Schanzenbach, D. W. (2014). *Does Class Size Matter?* Boulder, CO: National Education Policy Center. Retrieved October 20, 2014, from http://nepc.colorado.edu/files/pb_-_class_size.pdf.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Belmont, CA: Wadsworth.

Shah, B.R., Laupacis, A., Hux, J. E., & Austin, P.C. (2004) Propensity score methods gave similar results to traditional regression modeling in observational studies: A systematic review. *Journal of Clinical Epidemiology, 58,* 550-559.

Slavtcheva, D. (2010). *Propensity score analysis Appendix of The impact of City Connects: Annual report 2010.* Chestnut Hill, MA: Boston College.

StataCorp. (2013a). *Stata 13 Base Reference Manual.* College Station, TX: Stata Press.

StataCorp. (2013b). *Stata Statistical Software: Release 13 [computer software]*. College Station, TX: StataCorp LP.

StataCorp. (2013c). *Stata User's Guide: Release 13*. College Station, TX: Stata Press.

Stuart, E.A., & Rubin, D. B. (2007). *Best practices in quasi-experimental designs: Matching methods for causal inference*. Retrieved January 14, 2012, from http://www.gsbc.uni-jena.de/fileadmin/soec/media/GSBC/News/BBS_Steiner.pdf.

References

Tekwe, C.D., Carter, R. L., Ma, C., Algina, J., Lucas, M.E., Roth, J., Ariet, M., Fisher, T., & Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29*(1), 11-36.

The Mathworks Inc. (2014). *MATLAB (Version 8.4) [computer software]*. Natick, MA: The MathWorks Inc.

Todd, P. E., & Walpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal, 113*(485), F3-33.

Urdan, T., & Klein, S. (1998). *Early adolescence: A review of the literature*. A paper prepared for the U. S. Department of Educational Research and Improvement, Berkeley, CA, May 7-8. Retrieved February 20, 2014, from: http://www.rti.org/pubs/early_adolescence.pdf.

Walsh, M. E., & Brabeck, M.M. (2006). Resilience and risk in learning: Complex interactions and comprehensive interventions. In R.J. Sternberg & R.F. Subotnik (Eds.), *Optimizing student success in school with the other three Rs: Reasoning, resilience, and responsibility*. (pp. 113-142). Greenwich, CT: Information Age Publishing.

Walsh, M. E., & Depaul, J. (2008). The essential role of school-community partnerships in school counseling. In H. L. K. Coleman & C. Yeh, *Handbook of School Counseling* (pp. 765-783). New York: Taylor & Francis Group.

Walsh, M. E., Kenny, M. E., Wieneke, K. M., & Harrington, K. R. (2008). The Boston Connects program: Promoting learning and healthy development. *Professional School Counseling, 12*(2), 166-169.

Walsh, M. E., Madaus, G. F., Raczek, A. E., Dearing, E., Foley, C., An, C., Lee-St. John, T. J., & Beaton, A. (2014). A new model for student support in high-poverty urban elementary schools: effects on elementary and middle school academic outcomes. *American Educational Research Journal, 51*(4), 704-737.

Walsh, M. E., Murphy, J. A. (2003). *Children, Health, and Learning*. Westport, CT: Greenwood Publishing Group.

Webster, W. J., & Mendro, R. L. (1997). The Dallas Value-Added Accountability System. In J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp. 81-99). Thousand Oaks, California: Corwin Press, INC.

What Works Clearinghouse. (2011). *What works clearinghouse: Procedures and standards handbook*. Retrieved February 23, 2012, from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effect on student achievement: Implication for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57-67.

Xiang, Y., & Wang, S. (2013). An application of propensity score stratification using multilevel models: Do charter schools make a difference in student achievement and growth? Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA, April 28-30. Retrieved February 24, 2015, from: https://www.nwea.org/content/uploads/2013/08/An-Application-of-Propensity-Score-Stratification-Using-Multilevel-Models.pdf.

Yen, W. M. (1986). The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement, 23*(4), 299-325.

Zimmer, R. W., & Toma, E. F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management, 19*(1), 75-92.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *The Review of Economics and Statistics, 85*(1), 9-23.

Appendix A

APPENDIX A. RESULTS OF RQ1A MIDDLE SCHOOL ADJUSTEMENT MODELS

AND THE ASSOCIATED STATISTICAL EQUATIONS

Table A.1

Results of RQ1A Middle School Adjustment Models: Hierarchical Linear Models Predicting MCAS ELA Standardized Scores in Middle School

Fixed Effects	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	0.23	0.02	0.000	0.07	0.03	0.018	0.10	0.03	0.001
Male	-0.12	0.01	0.000	-0.20	0.02	0.000	-0.16	0.02	0.000
is_Black	-0.06	0.02	0.009	0.02	0.03	0.591	-0.01	0.06	0.886
is_Asian	0.06	0.02	0.006	0.05	0.05	0.314	0.09	0.09	0.296
is_Hispanic	-0.03	0.02	0.132	0.03	0.04	0.345	0.04	0.05	0.388
is_Other	-0.03	0.06	0.672	-0.02	0.10	0.860	0.13	0.15	0.412
Bilingual	-0.02	0.01	0.130	-0.02	0.03	0.583	0.02	0.03	0.415
Special Needs 2	-0.07	0.02	0.000	-0.09	0.03	0.001	-0.11	0.04	0.002
Special Needs 3	-0.20	0.02	0.000	-0.25	0.04	0.000	-0.27	0.04	0.000
Reduced Lunch	-0.04	0.05	0.351	0.00	0.09	0.961	0.14	0.11	0.203
Free Lunch	-0.15	0.03	0.000	-0.07	0.07	0.314	-0.01	0.08	0.874
Foreign Born	0.06	0.02	0.005	0.08	0.03	0.013	0.06	0.03	0.049
MCAS ELA_gr5	0.57	0.01	0.000	0.46	0.01	0.000	0.46	0.02	0.000
MCAS Math_gr5	0.18	0.01	0.000	0.14	0.01	0.000	0.18	0.02	0.000
# School Moves	-0.01	0.00	0.027	0.00	0.01	0.684	0.00	0.01	0.772
Age	-0.05	0.02	0.001	-0.06	0.02	0.002	-0.09	0.02	0.000
Distance from School	0.00	0.00	0.290	0.00	0.00	0.404	0.00	0.00	0.349

Appendix A

Table A.2

Results of RQ1A Middle School Adjustment Models: Hierarchical Linear Models Predicting MCAS Mathematics Standardized Scores in Middle School

Fixed Effects	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	0.20	0.03	0.000	-0.01	0.04	0.782	-0.02	0.04	0.609
Male	-0.01	0.01	0.197	0.01	0.02	0.778	-0.02	0.02	0.348
is_Black	-0.10	0.02	0.000	-0.09	0.04	0.036	-0.10	0.05	0.051
is_Asian	0.14	0.03	0.000	0.23	0.06	0.001	0.27	0.06	0.000
is_Hispanic	-0.04	0.02	0.056	-0.03	0.04	0.512	-0.01	0.05	0.780
is_Other	-0.13	0.06	0.023	-0.23	0.07	0.001	-0.41	0.09	0.000
Bilingual	0.02	0.02	0.343	0.05	0.03	0.066	0.04	0.02	0.060
Special Needs 2	-0.04	0.03	0.131	-0.04	0.03	0.201	-0.05	0.03	0.162
Special Needs 3	-0.16	0.04	0.000	-0.17	0.04	0.000	-0.15	0.06	0.024
Reduced Lunch	-0.05	0.04	0.184	-0.10	0.07	0.145	-0.06	0.11	0.553
Free Lunch	-0.09	0.03	0.001	-0.06	0.05	0.184	-0.06	0.05	0.189
Foreign Born	0.03	0.02	0.186	-0.01	0.03	0.597	0.06	0.03	0.059
MCAS ELA_gr5	0.14	0.01	0.000	0.10	0.01	0.000	0.08	0.01	0.000
MCAS Math_gr5	0.64	0.01	0.000	0.57	0.01	0.000	0.58	0.02	0.000
# School Moves	-0.01	0.01	0.385	-0.01	0.01	0.621	-0.02	0.01	0.131
Age	-0.08	0.02	0.000	-0.12	0.02	0.000	-0.17	0.03	0.000
Distance from School	0.00	0.00	0.220	0.00	0.00	0.864	0.00	0.00	0.749

Appendix A

Table A.3

Results of RQ1A Middle School Adjustment Models: Hierarchical Linear Models Predicting Weighted GPA in Middle School

Fixed Effects	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	2.69	0.03	0.000	2.46	0.04	0.000	2.54	0.04	0.000
Male	-0.35	0.02	0.000	-0.37	0.03	0.000	-0.40	0.04	0.000
is_Black	-0.11	0.05	0.026	-0.06	0.07	0.367	-0.01	0.06	0.861
is_Asian	0.27	0.05	0.000	0.39	0.09	0.000	0.48	0.11	0.000
is_Hispanic	-0.06	0.05	0.168	-0.04	0.06	0.524	0.01	0.06	0.925
is_Other	-0.14	0.07	0.039	-0.11	0.13	0.393	-0.36	0.14	0.010
Bilingual	0.08	0.02	0.000	0.09	0.03	0.002	0.06	0.03	0.075
Special Needs 2	0.03	0.03	0.326	0.07	0.04	0.098	0.05	0.05	0.242
Special Needs 3	-0.01	0.03	0.783	0.03	0.04	0.528	0.05	0.04	0.215
Reduced Lunch	-0.11	0.05	0.015	-0.12	0.07	0.076	-0.20	0.11	0.075
Free Lunch	-0.25	0.04	0.000	-0.23	0.05	0.000	-0.31	0.08	0.000
Foreign Born	0.10	0.03	0.001	0.11	0.05	0.018	0.11	0.05	0.030
MCAS ELA_gr5	0.18	0.01	0.000	0.12	0.02	0.000	0.08	0.02	0.000
MCAS Math_gr5	0.32	0.02	0.000	0.23	0.02	0.000	0.24	0.02	0.000
# School Moves	-0.06	0.01	0.000	-0.06	0.01	0.000	-0.06	0.02	0.001
Age	-0.15	0.02	0.000	-0.16	0.03	0.000	-0.12	0.04	0.003
Distance from School	-0.01	0.01	0.157	0.00	0.01	0.836	-0.01	0.01	0.077

Appendix A

Statistical Equations for RQ1a Middle School Adjustment Models Predicting

MCAS ELA and Mathematics Standardized Scores and Weighted GPA in Middle School

Level 1 (Student Level):

$Y_{ij} = \beta_{0j} + \beta_{1j}$ *(Male) + β_{2j} *(is_Black) + β_{3j} *(is_Asian) + β_{4j} *(is_Hispanic) + β_{5j} *(is_Other) + β_{6j} *(Bilingual) + β_{7j} *(Special Needs 2) + β_{8j} *(Special Needs 3) + β_{9j} *(Reduced Lunch) + β_{10j} *(Free Lunch) + β_{11j} *(Foreign Born) + β_{12j} *(MCAS_ELA_gr5) + β_{13j} *(MCAS_Math_gr5) + β_{14j} *(# School Moves) + β_{15j} *(Age) + β_{16j} *(Distance from School) + r_{ij}

Level 2 (School Level):

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

….

$$\beta_{16j} = \gamma_{160}$$

Appendix B

APPENDIX B. RESULTS OF RQ1A AND THE ASSOCIATED STATISTICAL MODELS

Table B.1

Results of RQ1a: Hierarchical Linear Models Predicting MCAS ELA Standardized Scores in Middle School

Fixed Effects	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	-0.08	0.05	0.165	-2.43	0.04	0.000	-2.44	0.04	0.000
EDose (Ever City Connects)	0.43	0.17	**0.016**	0.25	0.08	**0.004**	0.38	0.09	**0.000**
1-Year vs. 6-Year Dosage	-0.33	0.17	0.048	-0.19	0.06	0.002	-0.29	0.10	0.004
2-Year vs. 6-Year Dosage	-0.15	0.18	0.388	-0.25	0.07	0.001	-0.46	0.12	0.000
3-Year vs. 6-Year Dosage	-0.08	0.20	0.675	-0.06	0.13	0.609	-0.10	0.15	0.537
4-Year vs. 6-Year Dosage	-0.13	0.14	0.335	-0.04	0.08	0.635	-0.28	0.13	0.032
5-Year vs. 6-Year Dosage	0.15	0.08	0.060	0.07	0.14	0.604	-0.11	0.21	0.605
Male	-0.16	0.02	0.000	-0.23	0.02	0.000	-0.20	0.03	0.000
is_Black	-0.20	0.04	0.000	-0.07	0.06	0.251	-0.20	0.11	0.080
is_Asian	0.15	0.04	0.000	0.10	0.07	0.159	-0.01	0.13	0.957
is_Hispanic	-0.13	0.04	0.000	-0.03	0.06	0.641	-0.08	0.09	0.372
is_Other	-0.05	0.13	0.690	-0.04	0.13	0.781	0.00	0.24	0.991
Bilingual	-0.15	0.03	0.000	-0.15	0.07	0.029	-0.06	0.07	0.430
Special Needs 2	-0.31	0.07	0.000	-0.17	0.07	0.014	-0.12	0.07	0.090
Special Needs 3	-0.63	0.06	0.000	-0.54	0.06	0.000	-0.46	0.07	0.000
Reduced Lunch	-0.02	0.06	0.687	0.23	0.14	0.098	0.28	0.19	0.142
Free Lunch	-0.28	0.06	0.000	0.05	0.09	0.609	0.02	0.11	0.829
Foreign Born	0.26	0.03	0.000	0.21	0.06	0.000	0.16	0.04	0.001
RC_Reading_gr1	0.12	0.03	0.001	0.05	0.04	0.215	0.00	0.06	0.949
RC_Math_gr1	0.03	0.02	0.104	0.07	0.02	0.006	0.08	0.03	0.015
RC_Writing_gr1	0.03	0.02	0.159	-0.04	0.02	0.027	-0.06	0.04	0.168
RC_WorkHabit_gr1	0.07	0.02	0.005	0.04	0.02	0.037	0.14	0.04	0.000
RC_Behavior_gr1	-0.03	0.03	0.397	0.00	0.03	0.995	-0.04	0.03	0.230
RC_Effort_gr1	0.03	0.04	0.490	0.04	0.04	0.267	0.03	0.03	0.283
Age_gr1	0.04	0.04	0.364	0.02	0.04	0.590	-0.03	0.06	0.556
Distance from School_gr1	-0.01	0.01	0.234	0.01	0.01	0.408	0.01	0.01	0.393
# School Moves_gr1	0.03	0.03	0.371	0.06	0.03	0.079	0.10	0.06	0.096

Note. Bolded values are statistically significant school-level treatment effects at the .05 level.

Appendix B

Final Statistical Models for RQ1a Predicting

MCAS ELA Standardized Scores in Middle School

Level 1 (Student Level):

$$Y_{ik}^{j} = \beta_{ok} + \beta_{1k} *(Dosage\_1yr) + \beta_{2k} *(Dosage\_2yrs) + \beta_{3k} *(Dosage\_3yrs) + \beta_{4k} *(Dosage\_4yrs) + \beta_{5k} *(Dosage\_5yrs) + \beta_{6k} *(Male) + \beta_{7k} *(is\_Black) + \beta_{8k} *(is\_Asian) + \beta_{9k} *(is\_Hispanic) + \beta_{10k} *(is\_Other) + \beta_{11k} *(Bilingual) + \beta_{12k} *(Special\ Needs\ 2) + \beta_{13k} *(Special\ Needs\ 3) + \beta_{14k} *(Reduced\ Lunch) + \beta_{15k} *(Free\ Lunch) + \beta_{16k} *(Foreign\ Born) + \beta_{17k} *(RC\_Reading\_gr1) + \beta_{18k} *(RC\_Math\_gr1) + \beta_{19k} *(RC\_Writing\_gr1) + \beta_{20k} *(RC\_WorkHabit\_gr1) + \beta_{21k} *(RC\_Behavior\_gr1) + \beta_{22k} *(RC\_Effort\_gr1) + \beta_{23k} *(Age\_gr1) + \beta_{24k} *(Distance\ from\ School\_gr1) + \beta_{25k} *(\#\ School\ Moves\_gr1) + r_{ik}^{j}$$

Level 2 (School Level):

	Grade 6	Grade 7	Grade 8
$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + u_{0k}$			
$\beta_{1k} = \gamma_{10}$			
$\beta_{2k} = \gamma_{20}$			
$\beta_{3k} = \gamma_{30}$			
$\beta_{4k} = \gamma_{40}$			
$\beta_{5k} = \gamma_{50}$			
$\beta_{6k} = \gamma_{60}$			
$\beta_{7k} = \gamma_{70}$	$+ u_{7k}$		
$\beta_{8k} = \gamma_{80}$			

Appendix B

$\beta_{9k} = \gamma_{90}$			
$\beta_{10k} = \gamma_{100}$			
$\beta_{11k} = \gamma_{110}$		$+ u_{11k}$	
$\beta_{12k} = \gamma_{120}$	$+ u_{12k}$	$+ u_{12k}$	
$\beta_{13k} = \gamma_{130}$	$+ u_{13k}$		
$\beta_{14k} = \gamma_{140}$			
$\beta_{15k} = \gamma_{150}$			
$\beta_{16k} = \gamma_{160}$			
$\beta_{17k} = \gamma_{170}$	$+ u_{17k}$	$+ u_{17k}$	$+ u_{17k}$
$\beta_{18k} = \gamma_{180}$			
$\beta_{19k} = \gamma_{190}$			
$\beta_{20k} = \gamma_{200}$			
$\beta_{21k} = \gamma_{210}$	$+ u_{21k}$	$+ u_{21k}$	$+ u_{21k}$
$\beta_{22k} = \gamma_{220}$	$+ u_{22k}$	$+ u_{22k}$	
$\beta_{23k} = \gamma_{230}$	$+ u_{23k}$		
$\beta_{24k} = \gamma_{240}$	$+ u_{24k}$		
$\beta_{25k} = \gamma_{250}$			

Appendix B

Table B.2

Results of RQ1a: Hierarchical Linear Models Predicting MCAS Math Standardized Scores in Middle School

Fixed Effects	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	-0.09	0.06	0.139	0.13	0.05	0.008	0.09	0.05	0.073
EDose (Ever City Connects)	0.67	0.25	**0.009**	0.38	0.16	**0.022**	0.63	0.16	**0.000**
1-Year vs. 6-Year Dosage	-0.54	0.29	0.065	-0.29	0.21	0.163	-0.47	0.18	0.011
2-Year vs. 6-Year Dosage	-0.36	0.24	0.135	-0.30	0.19	0.111	-0.55	0.16	0.001
3-Year vs. 6-Year Dosage	-0.26	0.23	0.245	-0.25	0.14	0.066	-0.26	0.11	0.019
4-Year vs. 6-Year Dosage	-0.33	0.16	0.043	-0.18	0.11	0.105	-0.24	0.12	0.034
5-Year vs. 6-Year Dosage	-0.12	0.15	0.421	-0.12	0.16	0.448	-0.30	0.18	0.102
Male	0.06	0.03	0.063	0.05	0.03	0.067	0.12	0.03	0.001
is_Black	-0.20	0.05	0.000	-0.22	0.05	0.000	-0.27	0.07	0.000
is_Asian	0.33	0.04	0.000	0.37	0.06	0.000	0.27	0.09	0.002
is_Hispanic	-0.07	0.05	0.195	-0.09	0.04	0.038	-0.13	0.06	0.034
is_Other	-0.04	0.09	0.641	-0.17	0.11	0.112	-0.09	0.18	0.615
Bilingual	-0.06	0.04	0.176	-0.01	0.06	0.836	-0.16	0.05	0.003
Special Needs 2	-0.27	0.07	0.000	-0.14	0.06	0.011	-0.11	0.08	0.146
Special Needs 3	-0.63	0.07	0.000	-0.54	0.07	0.000	-0.42	0.08	0.000
Reduced Lunch	-0.07	0.08	0.364	-0.16	0.18	0.397	0.17	0.26	0.509
Free Lunch	-0.27	0.07	0.000	-0.24	0.09	0.008	0.03	0.11	0.787
Foreign Born	0.22	0.04	0.000	0.22	0.05	0.000	0.33	0.07	0.000
RC_Reading_gr1	0.04	0.03	0.154	0.03	0.05	0.579	-0.04	0.04	0.374
RC_Math_gr1	0.13	0.03	0.000	0.11	0.04	0.014	0.10	0.04	0.013
RC_Writing_gr1	0.05	0.02	0.033	-0.02	0.03	0.544	0.00	0.05	0.963
RC_WorkHabit_gr1	0.10	0.03	0.000	0.07	0.03	0.006	0.13	0.05	0.011
RC_Behavior_gr1	-0.02	0.03	0.410	-0.01	0.03	0.691	-0.03	0.06	0.544
RC_Effort_gr1	0.01	0.02	0.542	0.01	0.04	0.790	0.03	0.03	0.387
Age_gr1	-0.11	0.05	0.038	-0.14	0.04	0.000	-0.10	0.04	0.012
Distance from School_gr1	-0.02	0.01	0.027	-0.02	0.01	0.145	0.00	0.01	0.955
# School Moves_gr1	0.00	0.03	0.863	0.03	0.04	0.513	0.00	0.06	0.999

Note. Bolded values are statistically significant school-level treatment effects at the .05 level.

Appendix B

Final Statistical Models for RQ1a Predicting

MCAS Mathematics Standardized Scores in Middle School

Level 1 (Student Level):

$$Y_{ik}^{j} = \beta_{ok} + \beta_{1k} *(Dosage\_1yr) + \beta_{2k} *(Dosage\_2yrs) + \beta_{3k} *(Dosage\_3yrs) + \beta_{4k} *(Dosage\_4yrs) + \beta_{5k} *(Dosage\_5yrs) + \beta_{6k} *(Male) + \beta_{7k} *(is\_Black) + \beta_{8k} *(is\_Asian) + \beta_{9k} *(is\_Hispanic) + \beta_{10k} *(is\_Other) + \beta_{11k} *(Bilingual) + \beta_{12k} *(Special\ Needs\ 2) + \beta_{13k} *(Special\ Needs\ 3) + \beta_{14k} *(Reduced\ Lunch) + \beta_{15k} *(Free\ Lunch) + \beta_{16k} *(Foreign\ Born) + \beta_{17k} *(RC\_Reading\_gr1) + \beta_{18k} *(RC\_Math\_gr1) + \beta_{19k} *(RC\_Writing\_gr1) + \beta_{20k} *(RC\_WorkHabit\_gr1) + \beta_{21k} *(RC\_Behavior\_gr1) + \beta_{22k} *(RC\_Effort\_gr1) + \beta_{23k} *(Age\_gr1) + \beta_{24k} *(Distance\ from\ School\_gr1) + \beta_{25k} *(\#\ School\ Moves\_gr1) + r_{ik}^{j}$$

Level 2 (School Level):

	Grade 6	Grade 7	Grade 8
$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + u_{0k}$			
$\beta_{1k} = \gamma_{10}$			
$\beta_{2k} = \gamma_{20}$			
$\beta_{3k} = \gamma_{30}$			
$\beta_{4k} = \gamma_{40}$			
$\beta_{5k} = \gamma_{50}$			
$\beta_{6k} = \gamma_{60}$			
$\beta_{7k} = \gamma_{70}$	$+ u_{7k}$		
$\beta_{8k} = \gamma_{80}$			

Appendix B

$\beta_{9k} = \gamma_{90}$	$+ u_{9k}$		
$\beta_{10k} = \gamma_{100}$			
$\beta_{11k} = \gamma_{110}$			
$\beta_{12k} = \gamma_{120}$	$+ u_{12k}$		
$\beta_{13k} = \gamma_{130}$	$+ u_{13k}$		
$\beta_{14k} = \gamma_{140}$			
$\beta_{15k} = \gamma_{150}$			
$\beta_{16k} = \gamma_{160}$	$+ u_{16k}$		
$\beta_{17k} = \gamma_{170}$	$+ u_{17k}$	$+ u_{17k}$	$+ u_{17k}$
$\beta_{18k} = \gamma_{180}$			
$\beta_{19k} = \gamma_{190}$			
$\beta_{20k} = \gamma_{200}$			
$\beta_{21k} = \gamma_{210}$	$+ u_{21k}$	$+ u_{21k}$	$+ u_{21k}$
$\beta_{22k} = \gamma_{220}$			
$\beta_{23k} = \gamma_{230}$			
$\beta_{24k} = \gamma_{240}$	$+ u_{24k}$	$+ u_{24k}$	
$\beta_{25k} = \gamma_{250}$			$+ u_{25k}$

Appendix B

Table B.3

Results of RQ1a: Hierarchical Linear Models Predicting Weighted GPA in Middle School

Fixed Effects	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	-0.14	0.04	0.002	-0.03	0.03	0.442	-0.02	0.04	0.621
EDose (Ever City Connects)	0.64	0.07	**0.000**	0.38	0.06	**0.000**	0.34	0.06	**0.000**
1-Year vs. 6-Year Dosage	-0.52	0.11	0.000	-0.18	0.10	0.069	-0.27	0.06	0.000
2-Year vs. 6-Year Dosage	-0.32	0.07	0.000	-0.31	0.09	0.001	-0.40	0.06	0.000
3-Year vs. 6-Year Dosage	-0.33	0.09	0.000	-0.10	0.08	0.200	-0.06	0.09	0.523
4-Year vs. 6-Year Dosage	-0.16	0.07	0.024	0.02	0.10	0.854	-0.11	0.05	0.013
5-Year vs. 6-Year Dosage	-0.13	0.09	0.115	-0.05	0.14	0.727	-0.11	0.16	0.476
Male	-0.29	0.03	0.000	-0.33	0.03	0.000	-0.35	0.04	0.000
is_Black	-0.20	0.04	0.000	-0.02	0.06	0.721	0.08	0.10	0.419
is_Asian	0.38	0.04	0.000	0.63	0.09	0.000	0.83	0.16	0.000
is_Hispanic	-0.11	0.05	0.016	0.05	0.07	0.438	0.12	0.11	0.281
is_Other	-0.08	0.05	0.106	0.16	0.13	0.221	0.11	0.17	0.544
Bilingual	0.03	0.03	0.422	0.02	0.04	0.553	0.04	0.05	0.365
Special Needs 2	-0.06	0.05	0.230	0.00	0.07	0.943	0.09	0.06	0.142
Special Needs 3	-0.31	0.05	0.000	-0.13	0.05	0.005	-0.04	0.05	0.376
Reduced Lunch	-0.09	0.07	0.244	-0.13	0.11	0.260	0.03	0.14	0.845
Free Lunch	-0.25	0.06	0.000	-0.28	0.07	0.000	-0.20	0.07	0.004
Foreign Born	0.18	0.03	0.000	0.23	0.04	0.000	0.24	0.06	0.000
RC_Reading_gr1	-0.01	0.02	0.824	-0.04	0.03	0.277	0.00	0.03	0.999
RC_Math_gr1	0.03	0.02	0.105	0.01	0.04	0.720	-0.06	0.04	0.167
RC_Writing_gr1	0.05	0.01	0.000	0.00	0.02	0.800	-0.03	0.02	0.157
RC_WorkHabit_gr1	0.11	0.02	0.000	0.07	0.04	0.039	0.12	0.04	0.004
RC_Behavior_gr1	0.09	0.02	0.000	0.09	0.02	0.000	0.08	0.03	0.017
RC_Effort_gr1	0.03	0.02	0.202	0.05	0.03	0.070	0.07	0.03	0.026
Age_gr1	-0.01	0.04	0.750	-0.04	0.05	0.378	0.00	0.05	0.989
Distance from School_gr1	-0.02	0.01	0.006	-0.01	0.01	0.347	0.00	0.01	0.679
# School Moves_gr1	-0.03	0.05	0.452	-0.07	0.03	0.022	-0.07	0.04	0.078

Note. Bolded values are statistically significant school-level treatment effects at the .05 level.

Appendix B

Final Statistical Models for RQ1a Predicting Weighted GPA in Middle School

Level 1 (Student Level):

Y_{ik}^{j} = β_{ok} + β_{1k} *(Dosage_1yr) + β_{2k} *(Dosage_2yrs) + β_{3k} *(Dosage_3yrs) + β_{4k} *(Dosage_4yrs) + β_{5k} *(Dosage_5yrs) + β_{6k} *(Male) + β_{7k} *(is_Black) + β_{8k} *(is_Asian) + β_{9k} *(is_Hispanic) + β_{10k} *(is_Other) + β_{11k} *(Bilingual) + β_{12k} *(Special Needs 2) + β_{13k} *(Special Needs 3) + β_{14k} *(Reduced Lunch) + β_{15k} *(Free Lunch) + β_{16k} *(Foreign Born) + β_{17k} *(RC_Reading_gr1) + β_{18k} *(RC_Math_gr1) + β_{19k} *(RC_Writing_gr1) + β_{20k} *(RC_WorkHabit_gr1) + β_{21k} *(RC_Behavior_gr1) + β_{22k} *(RC_Effort_gr1) + β_{23k} *(Age_gr1) + β_{24k} *(Distance from School_gr1) + β_{25k} *(# School Moves_gr1) + r_{ik}^{j}

Level 2 (School Level):

	Grade 6	Grade 7	Grade 8
$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + u_{0k}$			
$\beta_{1k} = \gamma_{10}$			
$\beta_{2k} = \gamma_{20}$			
$\beta_{3k} = \gamma_{30}$			
$\beta_{4k} = \gamma_{40}$			
$\beta_{5k} = \gamma_{50}$			
$\beta_{6k} = \gamma_{60}$			
$\beta_{7k} = \gamma_{70}$			
$\beta_{8k} = \gamma_{80}$			
$\beta_{9k} = \gamma_{90}$	$+ u_{9k}$		

Appendix B

$\beta_{10k} = \gamma_{100}$			
$\beta_{11k} = \gamma_{110}$	$+ u_{11k}$		
$\beta_{12k} = \gamma_{120}$	$+ u_{12k}$		
$\beta_{13k} = \gamma_{130}$	$+ u_{13k}$		
$\beta_{14k} = \gamma_{140}$			
$\beta_{15k} = \gamma_{150}$	$+ u_{15k}$		
$\beta_{16k} = \gamma_{160}$			
$\beta_{17k} = \gamma_{170}$	$+ u_{17k}$	$+ u_{17k}$	
$\beta_{18k} = \gamma_{180}$		$+ u_{18k}$	
$\beta_{19k} = \gamma_{190}$			
$\beta_{20k} = \gamma_{200}$		$+ u_{20k}$	$+ u_{20k}$
$\beta_{21k} = \gamma_{210}$	$+ u_{21k}$		
$\beta_{22k} = \gamma_{220}$	$+ u_{22k}$		
$\beta_{23k} = \gamma_{230}$	$+ u_{23k}$	$+ u_{23k}$	
$\beta_{24k} = \gamma_{240}$	$+ u_{24k}$	$+ u_{24k}$	
$\beta_{25k} = \gamma_{250}$	$+ u_{25k}$		

Appendix C

APPENDIX C. RESULTS OF RQ1C AND THE ASSOCIATED STATISTICAL MODELS

Table C.1

Results of RQ1c: Hierarchical Linear Models Predicting MCAS ELA Standardized Scores in Middle School

Fixed Effects	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
School Level									
Intercept	-0.07	0.06	0.201	-2.43	0.04	0.000	-2.44	0.04	0.000
% Low Income	-0.01	0.00	0.011						
% Students with Disabilities	-0.03	0.01	0.005						
Student/Teacher Ratio	-0.08	0.03	0.017						
Edose (Ever City Connects)	0.42	0.16	**0.011**	0.25	0.08	**0.004**	0.38	0.09	**0.000**
Student Level									
1-Year vs. 6-Year Dosage	-0.33	0.16	0.045	-0.19	0.06	0.002	-0.29	0.10	0.004
2-Year vs. 6-Year Dosage	-0.16	0.18	0.379	-0.25	0.07	0.001	-0.46	0.12	0.000
3-Year vs. 6-Year Dosage	-0.08	0.20	0.668	-0.06	0.13	0.609	-0.10	0.15	0.537
4-Year vs. 6-Year Dosage	-0.13	0.13	0.328	-0.04	0.08	0.635	-0.28	0.13	0.032
5-Year vs. 6-Year Dosage	0.15	0.08	0.060	0.07	0.14	0.604	-0.11	0.21	0.605
Male	-0.16	0.02	0.000	-0.23	0.02	0.000	-0.20	0.03	0.000
is_Black	-0.20	0.04	0.000	-0.07	0.06	0.251	-0.20	0.11	0.080
is_Asian	0.15	0.04	0.000	0.10	0.07	0.159	-0.01	0.13	0.957
is_Hispanic	-0.13	0.04	0.000	-0.03	0.06	0.641	-0.08	0.09	0.372
is_Other	-0.05	0.13	0.695	-0.04	0.13	0.781	0.00	0.24	0.991
Bilingual	-0.16	0.03	0.000	-0.15	0.07	0.029	-0.06	0.07	0.430
Special Needs 2	-0.31	0.07	0.000	-0.17	0.07	0.014	-0.12	0.07	0.090
Special Needs 3	-0.63	0.06	0.000	-0.54	0.06	0.000	-0.46	0.07	0.000
Reduced Lunch	-0.02	0.06	0.685	0.23	0.14	0.098	0.28	0.19	0.142
Free Lunch	-0.28	0.06	0.000	0.05	0.09	0.609	0.02	0.11	0.829
Foreign Born	0.26	0.03	0.000	0.21	0.06	0.000	0.16	0.04	0.001
RC_Reading_gr1	0.12	0.03	0.001	0.05	0.04	0.215	0.00	0.06	0.949
RC_Math_gr1	0.03	0.02	0.108	0.07	0.02	0.006	0.08	0.03	0.015
RC_Writing_gr1	0.03	0.02	0.161	-0.04	0.02	0.027	-0.06	0.04	0.168
RC_WorkHabit_gr1	0.07	0.02	0.005	0.04	0.02	0.037	0.14	0.04	0.000
RC_Behavior_gr1	-0.02	0.03	0.424	0.00	0.03	0.995	-0.04	0.03	0.230
RC_Effort_gr1	0.02	0.04	0.517	0.04	0.04	0.267	0.03	0.03	0.283
Age_gr1	0.04	0.04	0.356	0.02	0.04	0.590	-0.03	0.06	0.556
Distance from School_gr1	-0.01	0.01	0.220	0.01	0.01	0.408	0.01	0.01	0.393
# School Moves_gr1	0.03	0.03	0.378	0.06	0.03	0.079	0.10	0.06	0.096

Note. Bolded values are statistically significant school-level treatment effects at the .05 level.

Appendix C

Final Statistical Models for RQ1c Predicting

MCAS ELA Standardized Scores in Middle School

Level 1 (Student Level):

$Y_{ik}^{j} = \beta_{ok} + \beta_{1k}$ *(Dosage_1yr) + β_{2k} *(Dosage_2yrs) + β_{3k} *(Dosage_3yrs) + β_{4k} *(Dosage_4yrs) + β_{5k} *(Dosage_5yrs) + β_{6k} *(Male) + β_{7k} *(is_Black) + β_{8k} *(is_Asian) + β_{9k} *(is_Hispanic) + β_{10k} *(is_Other) + β_{11k} *(Bilingual) + β_{12k} *(Special Needs 2) + β_{13k} *(Special Needs 3) + β_{14k} *(Reduced Lunch) + β_{15k} *(Free Lunch) + β_{16k} *(Foreign Born) + β_{17k} *(RC_Reading_gr1) + β_{18k} *(RC_Math_gr1) + β_{19k} *(RC_Writing_gr1) + β_{20k} *(RC_WorkHabit_gr1) + β_{21k} *(RC_Behavior_gr1) + β_{22k} *(RC_Effort_gr1) + β_{23k} *(Age_gr1) + β_{24k} *(Distance from School_gr1) + β_{25k} *(# School Moves_gr1) + r_{ik}^{j}

Level 2 (School Level):

Grade 6:

$$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + \gamma_{02} * (\%\ Low\_Income_k) + \gamma_{03} * (\%\ Student\_with\_Disabilities_k) + \gamma_{04} * (Student\_Teacher\_Ratio_k) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

[13]....

Grade 7:

$$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

[13]....

[13] The random effect components of the model are the same as the ones on pages 159-160.

Appendix C

Grade 8:

$$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

[13]….

Appendix C

Table C.2

Results of RQ1c: Hierarchical Linear Models Predicting MCAS Math Standardized Scores in Middle School

Fixed Effects	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
School Level									
Intercept	-0.10	0.06	0.087	0.12	0.05	0.009	0.10	0.05	0.042
Average Class Size	-0.07	0.02	0.001	-0.04	0.02	0.029	-0.05	0.02	0.011
% Low Income	-0.01	0.00	0.066						
% Free Lunch				-0.01	0.00	0.020			
% Reduced Lunch				-0.04	0.02	0.021			
Edose (Ever City Connects)	0.73	0.24	**0.003**	0.43	0.15	**0.005**	0.59	0.15	**0.000**
Student Level									
1-Year vs. 6-Year Dosage	-0.54	0.29	0.066	-0.29	0.21	0.158	-0.47	0.18	0.010
2-Year vs. 6-Year Dosage	-0.36	0.24	0.134	-0.30	0.19	0.107	-0.55	0.15	0.001
3-Year vs. 6-Year Dosage	-0.26	0.22	0.242	-0.25	0.13	0.061	-0.26	0.11	0.017
4-Year vs. 6-Year Dosage	-0.33	0.16	0.042	-0.18	0.11	0.105	-0.25	0.11	0.033
5-Year vs. 6-Year Dosage	-0.13	0.15	0.409	-0.12	0.16	0.456	-0.30	0.18	0.101
Male	0.06	0.03	0.063	0.05	0.03	0.073	0.12	0.03	0.001
is_Black	-0.20	0.05	0.000	-0.22	0.05	0.000	-0.28	0.07	0.000
is_Asian	0.33	0.04	0.000	0.37	0.06	0.000	0.27	0.09	0.002
is_Hispanic	-0.07	0.05	0.195	-0.09	0.04	0.044	-0.13	0.06	0.030
is_Other	-0.04	0.09	0.653	-0.16	0.11	0.123	-0.09	0.18	0.605
Bilingual	-0.05	0.04	0.193	-0.01	0.06	0.864	-0.15	0.05	0.003
Special Needs 2	-0.27	0.07	0.000	-0.14	0.06	0.010	-0.11	0.08	0.147
Special Needs 3	-0.64	0.07	0.000	-0.54	0.07	0.000	-0.42	0.08	0.000
Reduced Lunch	-0.07	0.08	0.356	-0.15	0.19	0.428	0.17	0.26	0.516
Free Lunch	-0.27	0.07	0.000	-0.23	0.09	0.014	0.03	0.11	0.796
Foreign Born	0.22	0.04	0.000	0.22	0.05	0.000	0.33	0.07	0.000
RC_Reading_gr1	0.04	0.03	0.155	0.03	0.05	0.588	-0.04	0.04	0.387
RC_Math_gr1	0.13	0.03	0.000	0.11	0.04	0.014	0.10	0.04	0.014
RC_Writing_gr1	0.05	0.02	0.035	-0.02	0.03	0.533	0.00	0.05	0.952
RC_WorkHabit_gr1	0.10	0.03	0.000	0.07	0.03	0.006	0.13	0.05	0.011
RC_Behavior_gr1	-0.02	0.03	0.402	-0.01	0.03	0.689	-0.03	0.06	0.550
RC_Effort_gr1	0.01	0.02	0.557	0.01	0.04	0.797	0.03	0.03	0.403
Age_gr1	-0.11	0.05	0.039	-0.14	0.04	0.000	-0.10	0.04	0.013
Distance from School_gr1	-0.02	0.01	0.024	-0.02	0.01	0.139	0.00	0.01	0.942
# School Moves_gr1	-0.01	0.03	0.840	0.03	0.04	0.530	0.00	0.06	0.975

Note. Bolded values are statistically significant school-level treatment effects at the .05 level.

Appendix C

Final Statistical Models for RQ1c Predicting

MCAS Mathematics Standardized Scores in Middle School

Level 1 (Student Level):

$Y_{ik}^{j} = \beta_{ok} + \beta_{1k}$ *(Dosage_1yr) + β_{2k} *(Dosage_2yrs) + β_{3k} *(Dosage_3yrs) + β_{4k} *(Dosage_4yrs) + β_{5k} *(Dosage_5yrs) + β_{6k} *(Male) + β_{7k} *(is_Black) + β_{8k} *(is_Asian) + β_{9k} *(is_Hispanic) + β_{10k} *(is_Other) + β_{11k} *(Bilingual) + β_{12k} *(Special Needs 2) + β_{13k} *(Special Needs 3) + β_{14k} *(Reduced Lunch) + β_{15k} *(Free Lunch) + β_{16k} *(Foreign Born) + β_{17k} *(RC_Reading_gr1) + β_{18k} *(RC_Math_gr1) + β_{19k} *(RC_Writing_gr1) + β_{20k} *(RC_WorkHabit_gr1) + β_{21k} *(RC_Behavior_gr1) + β_{22k} *(RC_Effort_gr1) + β_{23k} *(Age_gr1) + β_{24k} *(Distance from School_gr1) + β_{25k} *(# School Moves_gr1) + r_{ik}^{j}

Level 2 (School Level):

Grade 6:

$$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + \gamma_{02} * (Average\_Class\_Size_k) + \gamma_{03} * (\%\ Low\_Income_k) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

[14]….

Grade 7:

$$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + \gamma_{02} * (Average\_Class\_Size_k) + \gamma_{03} * (\%\ Free\_Lunch_k) + \gamma_{04} * (\%\ Reduced\_Lunch_k) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

[14] The random effect components of the model are the same as the ones on pages 162-163.

Appendix C

[14]….

Grade 8:

$$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + \gamma_{02} * (Average\_Class\_Size_k) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

[14]….

Appendix C

Table C.3

Results of RQ1c: Hierarchical Linear Models Predicting Weighted GPA in Middle School

Fixed Effects	Grade 6			Grade 7			Grade 8		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
School Level									
Intercept	-0.15	0.04	0.001	-0.03	0.04	0.421	-0.02	0.04	0.620
Average Class Size							-0.03	0.02	0.095
% Free Lunch	0.00	0.00	0.091	-0.01	0.00	0.116			
% Reduced Lunch	-0.05	0.02	0.001	-0.03	0.01	0.030			
Edose (Ever City Connects)	0.72	0.07	**0.000**	0.42	0.06	**0.000**	0.33	0.06	**0.000**
Student Level									
1-Year vs. 6-Year Dosage	-0.52	0.11	0.000	-0.18	0.10	0.074	-0.27	0.06	0.000
2-Year vs. 6-Year Dosage	-0.32	0.07	0.000	-0.31	0.09	0.001	-0.39	0.06	0.000
3-Year vs. 6-Year Dosage	-0.32	0.09	0.000	-0.10	0.08	0.210	-0.05	0.09	0.594
4-Year vs. 6-Year Dosage	-0.16	0.07	0.025	0.02	0.10	0.850	-0.11	0.04	0.011
5-Year vs. 6-Year Dosage	-0.13	0.09	0.116	-0.04	0.13	0.740	-0.11	0.16	0.469
Male	-0.29	0.03	0.000	-0.33	0.03	0.000	-0.35	0.04	0.000
is_Black	-0.20	0.04	0.000	-0.02	0.06	0.766	0.08	0.10	0.432
is_Asian	0.38	0.04	0.000	0.64	0.09	0.000	0.83	0.16	0.000
is_Hispanic	-0.11	0.05	0.015	0.06	0.07	0.407	0.12	0.11	0.288
is_Other	-0.08	0.05	0.096	0.16	0.13	0.220	0.10	0.17	0.559
Bilingual	0.03	0.03	0.424	0.03	0.04	0.523	0.04	0.05	0.356
Special Needs 2	-0.06	0.05	0.236	0.00	0.07	0.944	0.09	0.06	0.148
Special Needs 3	-0.30	0.05	0.000	-0.13	0.05	0.004	-0.04	0.05	0.374
Reduced Lunch	-0.08	0.07	0.250	-0.12	0.11	0.298	0.02	0.14	0.863
Free Lunch	-0.25	0.06	0.000	-0.27	0.08	0.001	-0.20	0.07	0.005
Foreign Born	0.18	0.03	0.000	0.23	0.04	0.000	0.25	0.06	0.000
RC_Reading_gr1	-0.01	0.02	0.807	-0.04	0.03	0.290	0.00	0.03	0.982
RC_Math_gr1	0.03	0.02	0.121	0.01	0.04	0.731	-0.06	0.04	0.169
RC_Writing_gr1	0.05	0.01	0.000	0.00	0.02	0.880	-0.04	0.02	0.154
RC_WorkHabit_gr1	0.11	0.02	0.000	0.07	0.04	0.038	0.12	0.04	0.004
RC_Behavior_gr1	0.09	0.02	0.000	0.09	0.02	0.000	0.08	0.03	0.016
RC_Effort_gr1	0.03	0.02	0.171	0.05	0.03	0.060	0.07	0.03	0.026
Age_gr1	-0.01	0.04	0.751	-0.04	0.05	0.386	0.00	0.05	0.989
Distance from School_gr1	-0.02	0.01	0.007	-0.01	0.01	0.323	0.00	0.01	0.664
# School Moves_gr1	-0.03	0.05	0.456	-0.07	0.03	0.022	-0.08	0.04	0.074

Note. Bolded values are statistically significant school-level treatment effects at the .05 level.

Appendix C

Final Statistical Models for RQ1 Predicting Weighted GPA in Middle School

Level 1 (Student Level):

Y_{ik}^{j} = β_{ok} + β_{1k} *(Dosage_1yr) + β_{2k} *(Dosage_2yrs) + β_{3k} *(Dosage_3yrs) + β_{4k} *(Dosage_4yrs) + β_{5k} *(Dosage_5yrs) + β_{6k} *(Male) + β_{7k} *(is_Black) + β_{8k} *(is_Asian) + β_{9k} *(is_Hispanic) + β_{10k} *(is_Other) + β_{11k} *(Bilingual) + β_{12k} *(Special Needs 2) + β_{13k} *(Special Needs 3) + β_{14k} *(Reduced Lunch) + β_{15k} *(Free Lunch) + β_{16k} *(Foreign Born) + β_{17k} *(RC_Reading_gr1) + β_{18k} *(RC_Math_gr1) + β_{19k} *(RC_Writing_gr1) + β_{20k} *(RC_WorkHabit_gr1) + β_{21k} *(RC_Behavior_gr1) + β_{22k} *(RC_Effort_gr1) + β_{23k} *(Age_gr1) + β_{24k} *(Distance from School_gr1) + β_{25k} *(# School Moves_gr1) + r_{ik}^{j}

Level 2 (School Level):

Grade 6:

$$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + \gamma_{02} * (\%\, Free\_Lunch_k) + \gamma_{03} * (\%\, Reduced\_Lunch_k) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

[15]....

Grade 7:

$$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + \gamma_{02} * (\%\, Free\_Lunch_k) + \gamma_{03} * (\%\, Reduced\_Lunch_k) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

[15]....

[15] The random effect components of the model are the same as the ones on pages 165-166.

Appendix C

Grade 8:

$$\beta_{0k} = \gamma_{00} + \gamma_{01} * (EDose_k) + \gamma_{02} * (Average\_Class\_Size_k) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

[15]….

Appendix D

APPENDIX D. RESULTS OF RQ2A AND THE ASSOCIATED STATISTICAL MODELS

Table D.1

Results of RQ2a: Hierarchical Linear Models Predicting Middle School Outcomes in Grade 6

Fixed Effects	MCAS ELA			MCAS Mathematics			Weighted GPA		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	0.34	0.06	0.000	0.28	0.05	0.000	2.80	0.06	0.000
MDose (Ever City Connects)	-0.06	0.03	0.104	0.08	0.05	0.086	0.06	0.06	0.342
Male	-0.13	0.01	0.000	-0.02	0.01	0.067	-0.34	0.03	0.000
is_Black	-0.07	0.04	0.092	-0.07	0.03	0.008	-0.19	0.03	0.000
is_Asian	0.04	0.02	0.035	0.16	0.04	0.001	0.18	0.02	0.000
is_Hispanic	-0.09	0.03	0.005	-0.04	0.02	0.157	-0.16	0.05	0.004
is_Other	-0.12	0.08	0.103	-0.13	0.03	0.000	-0.27	0.05	0.000
Bilingual	-0.01	0.02	0.666	0.04	0.02	0.034	0.14	0.04	0.000
Special Needs 2	-0.10	0.05	0.051	-0.10	0.02	0.000	0.04	0.04	0.303
Special Needs 3	-0.21	0.08	0.015	-0.12	0.04	0.014	-0.01	0.05	0.910
Reduced Lunch	-0.10	0.06	0.078	-0.12	0.06	0.043	-0.15	0.10	0.156
Free Lunch	-0.19	0.03	0.000	-0.10	0.03	0.007	-0.27	0.02	0.000
Foreign Born	0.13	0.07	0.079	0.05	0.02	0.030	0.14	0.02	0.000
MCAS ELA_gr5	0.56	0.01	0.000	0.14	0.01	0.000	0.16	0.01	0.000
MCAS Math_gr5	0.20	0.02	0.000	0.66	0.01	0.000	0.34	0.02	0.000
Age_gr5	-0.05	0.04	0.229	-0.06	0.03	0.106	-0.14	0.04	0.002
Distance from School_gr5	0.00	0.01	0.791	0.00	0.00	0.375	0.00	0.00	0.538
# School Moves_gr5	0.02	0.01	0.111	-0.01	0.01	0.325	-0.03	0.01	0.003

Note. Bolded values are statistically significant school-level treatment effects at the .05 level.

Appendix D

Final Statistical Models for RQ2a Predicting Middle School Outcomes in Grade 6

Level 1 (Student Level):

$$Y_{ij} = \beta_{0j} + \beta_{1j} *(Male) + \beta_{2j} *(is\_Black) + \beta_{3j} *(is\_Asian) + \beta_{4j} *(is\_Hispanic) + \beta_{5j} *(is\_Other) +$$

$$\beta_{6j} *(Bilingual) + \beta_{7j} *(Special\ Needs\ 2) + \beta_{8j} *(Special\ Needs\ 3) + \beta_{9j} *(Reduced\ Lunch) +$$

$$\beta_{10j} *(Free\ Lunch) + \beta_{11j} *(Foreign\ Born) + \beta_{12j} *(MCAS\ ELA\_gr5) + \beta_{13j} *(MCAS\ Math\_gr5) +$$

$$\beta_{14j} *(Age\_gr5) + \beta_{15j} *(Distance\ from\ School\_gr5) + \beta_{16j} *(\#\ School\ Moves\_gr5) + r_{ij}$$

Level 2 (School Level):

	MCAS ELA	MCAS Mathematics	Weighted GPA
$\beta_{0j} = \gamma_{00} + \gamma_{01} * (MDose_j) + u_{0j}$			
$\beta_{1j} = \gamma_{10}$			$+ u_{1j}$
$\beta_{2j} = \gamma_{20}$	$+ u_{2j}$		
$\beta_{3j} = \gamma_{30}$			
$\beta_{4j} = \gamma_{40}$			$+ u_{4j}$
$\beta_{5j} = \gamma_{50}$			
$\beta_{6j} = \gamma_{60}$			
$\beta_{7j} = \gamma_{70}$	$+ u_{7j}$		$+ u_{7j}$
$\beta_{8j} = \gamma_{80}$	$+ u_{8j}$	$+ u_{8j}$	
$\beta_{9j} = \gamma_{90}$			$+ u_{9j}$

Appendix D

$\beta_{10j} = \gamma_{100}$	$+ u_{10j}$	$+ u_{10j}$	
$\beta_{11j} = \gamma_{110}$	$+ u_{11j}$		
$\beta_{12j} = \gamma_{120}$			
$\beta_{13j} = \gamma_{130}$	$+ u_{13j}$	$+ u_{13j}$	$+ u_{13j}$
$\beta_{14j} = \gamma_{140}$	$+ u_{14j}$	$+ u_{14j}$	$+ u_{14j}$
$\beta_{15j} = \gamma_{150}$			
$\beta_{16j} = \gamma_{160}$			

Appendix D

Table D.2

Results of RQ2a: Hierarchical Linear Models Predicting Middle School Outcomes in Grade 6 with the Elementary Dose Indicator

Fixed Effects	MCAS ELA			MCAS Mathematics			Weighted GPA		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
Intercept	0.34	0.06	0.000	0.28	0.05	0.000	2.80	0.06	0.000
MDose (Ever City Connects)	-0.05	0.03	0.175	0.04	0.05	0.419	0.00	0.05	0.970
Edose	-0.02	0.03	0.558	0.05	0.03	0.120	0.09	0.04	**0.035**
Male	-0.13	0.01	0.000	-0.02	0.01	0.085	-0.33	0.03	0.000
is_Black	-0.07	0.04	0.092	-0.07	0.03	0.009	-0.20	0.03	0.000
is_Asian	0.04	0.02	0.035	0.16	0.05	0.001	0.17	0.02	0.000
is_Hispanic	-0.09	0.03	0.004	-0.04	0.03	0.167	-0.16	0.05	0.004
is_Other	-0.12	0.08	0.103	-0.13	0.03	0.000	-0.28	0.05	0.000
Bilingual	-0.01	0.03	0.715	0.03	0.02	0.054	0.14	0.03	0.000
Special Needs 2	-0.10	0.05	0.051	-0.10	0.02	0.000	0.04	0.04	0.340
Special Needs 3	-0.21	0.08	0.015	-0.11	0.04	0.013	0.00	0.05	0.973
Reduced Lunch	-0.10	0.06	0.071	-0.12	0.06	0.048	-0.14	0.10	0.171
Free Lunch	-0.19	0.03	0.000	-0.10	0.03	0.008	-0.27	0.02	0.000
Foreign Born	0.13	0.07	0.078	0.05	0.02	0.027	0.14	0.02	0.000
MCAS ELA_gr5	0.56	0.01	0.000	0.14	0.01	0.000	0.16	0.01	0.000
MCAS Math_gr5	0.20	0.02	0.000	0.66	0.02	0.000	0.34	0.02	0.000
Age_gr5	-0.05	0.04	0.225	-0.06	0.03	0.113	-0.14	0.04	0.002
Distance from School_gr5	0.00	0.01	0.807	0.00	0.00	0.441	0.00	0.00	0.457
# School Moves_gr5	0.02	0.01	0.110	-0.01	0.01	0.330	-0.02	0.01	0.003

Note. Bolded values are statistically significant school-level treatment effects at the .05 level.

Appendix D

Final Statistical Models for RQ2a Predicting Middle School Outcomes in Grade 6 with the Elementary Dose Indictor

Level 1 (Student Level):

$Y_{ij} = \beta_{oj} + \beta_{1j}$ *(EDose) + β_{2j} *(Male) + β_{3j} *(is_Black) + β_{4j} *(is_Asian) + β_{5j} *(is_Hispanic) + β_{6j} *(is_Other) +

β_{7j} *(Bilingual) + β_{8j} *(Special Needs 2) + β_{9j} *(Special Needs 3) + β_{10j} *(Reduced Lunch) +

β_{11j} *(Free Lunch) + β_{12j} *(Foreign Born) + β_{13j} *(MCAS ELA_gr5) + β_{14j} *(MCAS Math_gr5) +

β_{15j} *(Age_gr5) + β_{16j} *(Distance from School_gr5) + β_{17j} *(# School Moves_gr5) + r_{ij}

Level 2 (School Level):

	MCAS ELA	MCAS Mathematics	Weighted GPA
$\beta_{0j} = \gamma_{00} + \gamma_{01} * (MDose_j) + u_{0j}$			
$\beta_{1j} = \gamma_{10}$			
$\beta_{2j} = \gamma_{20}$			$+ u_{2j}$
$\beta_{3j} = \gamma_{30}$	$+ u_{3j}$		
$\beta_{4j} = \gamma_{40}$			
$\beta_{5j} = \gamma_{50}$			$+ u_{5j}$
$\beta_{6j} = \gamma_{60}$			
$\beta_{7j} = \gamma_{70}$			

Appendix D

$\beta_{8j} = \gamma_{80}$	$+ u_{8j}$		$+ u_{8j}$
$\beta_{9j} = \gamma_{90}$	$+ u_{9j}$	$+ u_{9j}$	
$\beta_{10j} = \gamma_{100}$			$+ u_{10j}$
$\beta_{11j} = \gamma_{110}$	$+ u_{11j}$	$+ u_{11j}$	
$\beta_{12j} = \gamma_{120}$	$+ u_{12j}$		
$\beta_{13j} = \gamma_{130}$			
$\beta_{14j} = \gamma_{140}$	$+ u_{14j}$	$+ u_{14j}$	$+ u_{14j}$
$\beta_{15j} = \gamma_{150}$	$+ u_{15j}$	$+ u_{15j}$	$+ u_{15j}$
$\beta_{16j} = \gamma_{160}$			
$\beta_{17j} = \gamma_{170}$			

Appendix E

APPENDIX E. RESULTS OF RQ2C AND THE ASSOCIATED STATISTICAL MODELS

Table E

Results of RQ2c: Hierarchical Linear Models Predicting Middle School Outcomes in Grade 6 with School-level Covariates

Fixed Effects	MCAS ELA			MCAS Mathematics			Weighted GPA		
	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
School Level									
Intercept	0.50	0.12	0.001	0.27	0.03	0.000	2.74	0.09	0.000
% Foreign Language not English				-0.01	0.00	0.002	-0.01	0.00	0.001
Students per Computer	0.06	0.01	0.001	0.07	0.01	0.000	0.05	0.01	0.004
Student/Teacher Ratio	-0.03	0.01	0.011						
MDose (Ever City Connects)	-0.04	0.03	0.144	0.13	0.02	**0.000**	0.14	0.04	**0.003**
Student Level									
Male	-0.13	0.01	0.000	-0.02	0.01	0.067	-0.34	0.03	0.000
is_Black	-0.07	0.04	0.110	-0.08	0.03	0.007	-0.19	0.03	0.000
is_Asian	0.04	0.02	0.028	0.16	0.04	0.001	0.18	0.02	0.000
is_Hispanic	-0.09	0.03	0.007	-0.03	0.02	0.169	-0.15	0.05	0.005
is_Other	-0.12	0.08	0.118	-0.12	0.03	0.000	-0.27	0.05	0.000
Bilingual	-0.01	0.02	0.644	0.04	0.02	0.030	0.14	0.04	0.000
Special Needs 2	-0.10	0.05	0.048	-0.09	0.02	0.000	0.05	0.04	0.293
Special Needs 3	-0.21	0.08	0.013	-0.12	0.04	0.010	0.00	0.05	0.915
Reduced Lunch	-0.10	0.06	0.087	-0.12	0.06	0.046	-0.15	0.10	0.146
Free Lunch	-0.19	0.03	0.000	-0.09	0.03	0.015	-0.27	0.02	0.000
Foreign Born	0.13	0.07	0.068	0.05	0.02	0.033	0.14	0.02	0.000
MCAS ELA_gr5	0.56	0.01	0.000	0.14	0.01	0.000	0.16	0.01	0.000
MCAS Math_gr5	0.20	0.02	0.000	0.66	0.01	0.000	0.34	0.02	0.000
Age_gr5	-0.05	0.04	0.242	-0.06	0.03	0.112	-0.13	0.04	0.002
Distance from School_gr5	0.00	0.01	0.863	0.00	0.00	0.274	0.00	0.00	0.625
# School Moves_gr5	0.02	0.01	0.108	-0.01	0.01	0.328	-0.03	0.01	0.003

Note. Bolded values are statistically significant school-level treatment effects at the .05 level.

Appendix E

Final Statistical Models for RQ2c Predicting Middle School Outcomes in Grade 6

Level 1 (Student Level):

$Y_{ij} = \beta_{0j} + \beta_{1j}$ *(Male) + β_{2j} *(is_Black) + β_{3j} *(is_Asian) + β_{4j} *(is_Hispanic) + β_{5j} *(is_Other) + β_{6j} *(Bilingual) + β_{7j} *(Special Needs 2) + β_{8j} *(Special Needs 3) + β_{9j} *(Reduced Lunch) + β_{10j} *(Free Lunch) + β_{11j} *(Foreign Born) + β_{12j} *(MCAS ELA_gr5) + β_{13j} *(MCAS Math_gr5) + β_{14j} *(Age_gr5) + β_{15j} *(Distance from School_gr5) + β_{16j} *(# School Moves_gr5) + r_{ij}

Level 2 (School Level):

MCAS ELA:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (MDose_j) + \gamma_{02} * (Students\_per\_Computer_j) + \gamma_{03} * (Student\_Teacher\_Ratio_k) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

[16]....

MCAS Mathematics:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (MDose_j) + \gamma_{02} * (\%\, Foreign\_Language\_not\_English_j) + \gamma_{03} * (Students\_per\_Computer_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

[16]....

Weighted GPA:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * \left(MDose_j\right) + \gamma_{02} * (\%\, Foreign\_Language\_not\_English_j) + \gamma_{03} * (Students\_per\_Computer_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

[16]....

[16] The random effect components of the model are the same as the ones on pages 177-178.