The Pennsylvania State University

The Graduate School

EFFECTS OF SUPPLEMENTAL SERVICES ON STUDENTS' MOTIVATION, ENGAGEMENT AND ACADEMIC ACHIEVEMENT

A Thesis in

Educational Psychology

by

Ming Yang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2022

The thesis of Ming Yang was reviewed and approved by the following:

Matthew McCrudden
Professor of Educational Psychology
Thesis Advisor
Professor in Charge of the Program of Educational Psychology

James DiPerna
Professor of Education

P. Karen Murphy
Professor of Education
Head of Department of Educational Psychology, Counseling and Special Education

ABSTRACT

Students who enter elementary school with low academic achievement are provided supplemental services to help address skill deficits. However, the effect of supplemental services has not been shown to be robust according to previous research using state-wide standardized tests to measure students' academic achievement. The probable explanations for these results are ineffectiveness of supplemental services, and interferences of potentially confounding variables demographic variables associated with the receipt of supplemental services. As such, the current research employed propensity score matching to statistically "control" for multiple demographic variables by across "control" (no supplemental services) and "treatment" (received serviced) groups. After matching first grade students' gender, race/ethnicity, primary language, home language and learning disability (reading and mathematics), I used repeated measure ANOVA to compare student's math and reading scores, motivation and engagement across groups (i.e., students receiving supplemental services and students not receiving services) at the three time points during the school year. The analysis showed that supplemental service did not significantly improve or reduce students' academic achievement, motivation and engagement compared to the "control" group. Future research regarding the evaluation of supplemental services would benefit from using the propensity score matching method, including more potential confounding variables that were unavailable in the current data set (e.g., SES), and including a larger sample size of students and schools.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my academic advisor, Dr. Pui-wa Lei, for her patience and support in my first two years of master program and in the process of developing my thesis. I would thank Dr. James DiPerna and the SSIS-CIP project team for collecting and sharing the data set. I would also thank Dr. Matthew McCrudden for serving as my thesis advisor and editing the thesis.

I am also grateful to Kyle Husmann from Human Development and Family Studies program and Tianying Sun from Educational Psychology program. Both have provided support and suggestion during my data analysis process.

Chapter 1

Introduction

Children enter primary school with various prior experiences and knowledge; some score lower on math or reading than others (National Center for Education Statistics, 2016a, 2016b). Though first-grade children don't have much experience with formal math and reading instruction, they have learned by observing and interacting with surrounding environments and develop individual differences at an early age (Schunk, 2012). Statistics have revealed individual differences in math and reading for first-grade children and identify family income as risk factor for students' academic performance (National Center for Education Statistics, 2009, 2012). By categorizing students into low- or high-income groups, National Center for Education Statistics (2009, 2012) reported that primary school students from low-income families tend to have lower academic performance in math and reading.

Students' prior academic achievement, motivation and engagement are related to later achievement, especially reading achievement (Alexander et al., 1997; DiPerna et al., 2002). According to goal theory, students are more motivated when they expect their efforts to contribute to goal achievement, whereas they are less motivated when they expect failure (Wigfield & Eccles, 2002). If a student finds they cannot achieve the goal even if they spend more time and effort than their peers, they are less likely to invest time and effort in that goal; students with low prior performance tend to lose motivation in classes (Elliot, 2005; Schunk, 1987). With low motivation, students will not actively engage into learning tasks and are not as enthusiastic about learning, which leads to unsatisfying academic outcomes. After years of low performance, students developed low learning competence, including low motivation and engagement. In this case, at-risk students may benefit from early additional assistance.

The U.S. Department of Education (2009) provides resources for supplemental educational services to facilitate the academic performance of students who grow up in high poverty areas and do not make Adequate Yearly Progress (AYP). Under the Every Student Succeeds Act (ESSA), students from low-income families with low academic performance have priority to receive supplemental service when the Title 1 funding is limited (U.S. Department of Education, 2009). Given the vulnerability of students living in poverty, additional services that improve students' academic performance may be needed to facilitate learning (Ascher, 2006).

Supplemental education service (abbreviated to supplemental service for the remainder of this paper), was designed to compensate for risk factors - low socioeconomic status and low prior reading and math score - and promote positive learning outcomes for at-risk children. Under No Child Left Behind, and more recently ESSA, when a child is eligible for reduced- or free-price lunch and attends a school that needs improvement, he or she will be eligible to receive supplemental service (U.S. Department of Education, 2009). To promote students' academic progress, supplemental educational services typically focus on academic skills, such as math and reading, to promote students' progress toward state academic standards (U.S. Department of Education, 2009).

In 2011, the U.S. Department of Education provided more flexible choices for supplemental services, including school readiness courses, to close the student opportunity gap. Since then, local schools have generated various supplemental services for at-risk students outside of school days (U.S. Department of Education, 2012). Researchers have suggested several characteristics of successful supplemental service, such as parents' involvement, interaction with local teachers, and one-to-one tutors (Barley & Wegner, 2010; Muñoz et al., 2012). Even if the content of supplemental service changes over time, these characteristics could facilitate students' attendance of supplemental service and enhance students' opportunities for achievement.

Evaluation of Supplemental Services

Researchers have tried to demonstrate the effectiveness of supplemental service but instead have reported inconsistent results about the benefits of supplemental service on students' academic achievement (Deke et al., 2014; Heinrich et al., 2010; Muñoz et al., 2008, 2009, 2012; Ross et al., 2009; Springer et al., 2014). Researchers who measured students' academic progress on statewide standard tests found moderate or negligible effects of supplemental service on students' academic performance (Deke et al., 2014; Heinrich et al., 2010; Muñoz et al., 2008, 2009; Springer et al., 2014). One research study even found a negative effect of supplemental service on students' math (Ross et al., 2009).

Students in these evaluations of supplemental service completed statewide standardized tests in Grade 3 and above. Such standardized tests are applied yearly to measure the school quality and determine students' learning proficiency. Heinrich, Meyer and Whitten (2010) employed The Wisconsin Knowledge and Concepts Examination (WKCE) in their research which is applicable for students in Grade 4, 8 and 10. Muñoz, Ross and Neergaard (2008) measured students' academic performance with Kentucky Core Content Tests (KCCT), which was applicable to students in Grade 3 through Grade 8. The Tennessee Comprehensive Assessment Program (TCAP) was commonly used for students on Grade 3 through 8. Muñoz, Potter and Ross (2008) argued that these high-stakes standardized tests, purposed for school evaluation and students' academic proficiency, are not sensitive to students' individual progress. Hence, these test scores cannot precisely represent students' academic progress.

Researchers have applied various methods to separate the effect of supplemental service from demographic variables that have been shown to be related to student's academic performance, such as race/ ethnicity, gender, family income, disability, and education variables, such as school and classroom (Deke et al., 2014; Heinrich et al., 2010; Muñoz et al., 2008, 2009, 2012; Ross et al., 2009; Springer et al., 2014). To isolate the effect of supplemental services on academic performance, researchers have controlled for the effect of these variables (Ascher, 2006). Muñoz and colleagues (2008, 2012) included prior reading, prior math, gender, race, family structure, eligibility for special education and English proficiency as covariates into analysis and found a slight positive effect of supplemental service. Deke, Dragoset, Bogen and Gill (2012) matched students' prior academic scores by selecting samples near cutoff scores for supplemental service and included race, primary language, disability status and school as covariates, and did not find evidence to demonstrate the effect of supplemental service. Ross and colleagues (2009) predicted students' prior academic achievement with students' scores in the last two years. They regressed student's scores on the predicted score, grade, service provider and teacher to reveal the effect of different service providers (Ross et al., 2009).

Statement of the Problem

Supplemental service has not been shown to enhance students' learning outcomes in previous research. Three possible explanations include: (a) separating the effect of supplemental services and controlling other confounding variables requires the use of sophisticated research designs (Ascher, 2006); (b) differences in supplemental services among education service providers, local education administrators, etc. (Barley & Wegner, 2010; Ross et al., 2009); and (c) high-stake statewide standard tests (such as KCCT, TCAP) may not be sufficiently sensitive to individual student's progress (Muñoz et al., 2012). The current study examined the effect of supplemental service by analyzing its effect on – students' academic skills as well as their motivation and engagement.

Purpose of the Study

The current study employed a quantitative method to examine the effect of supplemental service on Grade 1 students. The current research used a matching process that accounts for covariates related to students' academic performance. Compared to previous research, this study aimed to reveal students' academic progress by comparing students who receive supplemental services with those who do not. The achievement data are students' math and reading scores on a standardized test. Student motivation and engagement were also measured as indicators of student learning processes (DiPerna et al., 2002; DiPerna & Elliott, 2002).

Research Questions

Given the challenges in evaluating supplemental services for elementary school students, I plan to assess the effect of supplemental service on at-risk students by addressing the following research question:

Does supplemental service promote Grade 1 student academic achievement, motivation, and engagement when accounting for demographics and prior achievement?

After controlling for covariates (e.g., such as, gender, race, home language, prior math, prior reading and prior social skills, disability, etc.), I predict that students who receive supplemental service will have higher academic achievement than their peers who do not receive supplemental service. Due to the correlation between academic performance and students' motivation and engagement, I expect an increase in motivation and engagement for students who receive supplemental service, after controlling for demographic variables and prior academic achievement.

Chapter 2

Literature Review

This review of literature focuses on the operation of the supplemental service and research related to the effectiveness of supplemental services. As supplemental service initially focused on promoting academic achievement for at-risk students, the review will examine measurement of academic achievement and how researchers have isolated the effect of supplemental service from other variables, which are related to provision of supplemental service and academic achievement at primary school (Ascher, 2006).

Supplemental Service

Supplemental service, under No Child Left Behind, is an additional free service for students to compensate for these students' shortage in educational resources (U.S. Department of Education, 2009). The original target for supplemental service is to improve students' academic performance because students from low-income families and low-performance schools have fewer educational resources than their peers.

Evaluations for supplemental service.

To evaluate the effectiveness of supplemental service, researchers focus on stakeholders' academic progress. Some researchers measure students' academic performance quantitatively by statewide standardized tests, while some measure students' academic performance qualitatively by parents' and teachers' reports and feedback, and others apply both methods to evaluate the effectiveness (Deke et al., 2014; Muñoz et al., 2009, 2012; etc.). Parents tend to give positive feedback on supplemental services. Teacher feedback on supplemental service relates to the depth and frequency service providers communicate with them. Teachers tend to view supplemental service positively when the service providers reach out to them. Parents' and teachers' reports about supplemental service are to support students' progress with the insight of knowing that these students have received extra service. Standardized tests may reveal students' academic progress but may overlook students' learning process (such as motivation and engagement) and might include the effects of several factors related to students' academic performance. Unlike teachers' and parents' reports, most research with standardized tests have indicated nonsignificant effects of supplemental service (Ross et al., 2008, 2009).

Measurement of Academic Performance

Researchers have used standardized tests to measure students' academic proficiency and can be use determine whether students achieve a state-set standard at each grade level (Harding et al., 2012; Muñoz et al., 2012; Ross et al., 2009). Students who fail to meet a district-set cutoff score will be viewed as at-risk and eligible to receive supplemental service (Deke et al., 2012, 2014). If a school has a low average score on the test, it will be identified as a school that needs improvement (U.S. Department of Education, 2009). These standardized tests have high stakes in evaluating students and schools, such as KCCT, TCAP, MSA, etc. (Harding et al., 2012; Muñoz et al., 2012; Ross et al., 2009). Deke, Dragoset, Bogen and Gill (2012) did not mention the specific academic achievement tests in their research but noted that students with scores under a specific cutoff were eligible for supplemental service.

Matching Technique

To evaluate the effects of supplemental service, researchers have used quasi-experiments to compare students who received supplemental service to those who did not (Ascher, 2006; Austin et al., 2018). Regarding parents' determination and students' needs, the researchers cannot randomly assign supplemental services to students but can analyze students' performance afterward. Therefore, researchers have used statistical tests to examine how supplemental service contributes to students' academic progress (on math and reading scores) and have tried to control for the effects of confounding variables that are related to both the attendance of supplemental service and the students' academic performance (Deke et al., 2012; Springer et al., 2014).

Researchers have used ANCOVA to include covariates in statistical analysis. Researchers compare students' prior math and reading scores, gender, race, and eligibility for free- or reduced-price lunch across the 'control' group and 'treatment' group (Muñoz et al., 2008, 2009, 2012). They control for variables distributed differently between groups as covariates, such as preliminary reading and race. For example, when testing the effect of reading services, Muñoz, Potter and Ross (2008) identified prior reading score and race as covariates and analyzed the impact of supplemental service on reading. When addressing ANCOVA analysis, researchers assume that the relationships between covariates and outcomes are known. Baker et al. (2019) stated that explaining the significant contribution of covariates is difficult as different sets of covariates can show different results.

Some researchers have controlled confounding variables by selecting students with similar levels on covariates, such as academic performance and family income (Deke et al., 2012; Muñoz et al., 2009). Muñoz, Ross and Neergaard (2008) filtered their sample to only include students eligible for free- or reduced-price lunch to examine the effect of supplemental service on students from low-income families. Deke, Drag, Bogen and Gill (2012) conducted research in school districts where the funding is not adequate to support all students' supplemental service; they included only eligible students with scores near the cutoff. With this research design, Deke et al. (2012) selected comparable groups of students with similar prior academic performance as well as a similar level of family income. As such, their results for students near cutoff scores may not be generalizable to students far below the cutoff (Deke et al., 2012). When selecting students based on one or two covariates, the researchers reduced the sample size and limited the generalizability of their findings.

In order to reconcile the difficulty in dealing with confounding variables, Rosenbaum and Rubin introduced propensity score matching to exclude effects of covariates in quasi-experiment designs (1983, 1985). When researchers cannot randomly assign students into treatment and control groups, the confounding variables influence both independent and dependent variables. As a result, the researcher cannot determine the causality between the independent and dependent variables. For example, in supplemental service research, students who come from low-income families are more likely to receive supplemental service, and their low-income status is related to their academic achievement (National Center for Education Statistics, 2009, 2012, 2019; U.S. Department of Education, 2009). The propensity scores could estimate the probability that students receive the supplemental service, then create comparable groups. After propensity score matching, we assume that students in the 'control' group and the 'treatment' group have equal chances to enter each group.

For example, a student from a low-income family who attends an underperforming school is more likely to participate in supplemental service, and the propensity score matching finds another student who has similar probabilities of participating in supplemental service but didn't receive the service. When two groups of students have paired propensity scores, we can assume that students in the control group are comparable to students in the treatment group; students in each group have similar levels of covariates. Researchers can assume that students in different groups are influenced by covariates similarly if students in both groups have similar covariates, and they can exclude the effect of covariates by comparing the control group and treatment group.

Several researchers have employed propensity score matching in educational research to examine the students' improvement while controlling for unmanipulated variables, such as socio-economic status, afterschool programs, etc. (Baker et al., 2019; Belfi et al., 2016). Belfi, Haelermans and De Fraine (2016) matched students' who attended low-, medium-, high- and mixed socio-economic status (SES) schools and compared students' achievement growth across the schools. The researchers could not assign students to the schools with a particular SES type, but they estimated each student's probability of entering a specific type of school with covariates (e.g., demographics, social background indicators, ethnic background, home, prior achievement, etc.). Their results showed that students had more positive achievement growth in high-SES compared to low-SES and mixed-SES schools. Springer et al. (2014) estimated the effect of supplemental service by matching several covariates: students' reading scores, math scores, race, gender, eligibility for free lunch, English proficiency, special education status, student attendance rate at school, and grade level of the student. They found that supplemental service promoted students' math performance but did not promote students' reading scores.

Confounding Variables to Control

In evaluations of supplemental service, researchers have included some covariates that relate to students' attendance of supplemental service and academic performance. Students from at-risk groups tend to score lower on math and reading and tend to apply for supplemental services. However, even students who apply for supplemental service do not always attend.

Researchers aim to examine the pure effect of supplemental service when including as many covariates as possible in their research.

Family income

Family income (the average financial income of each family member in the students' families) is a predictor of students' academic performance in both kindergarten and elementary school (National Center for Education Statistics, 2016a; Sirin, 2005). The National Center for Education Statistics (2019) distinguished poverty (a per-person income under the poverty threshold) as a statistically risky factor for students' reading scores from kindergarten to 5th grade. The NCES (2012) also reported that higher socio-economic status (a variable that combines family income and parents' education) is positively related to students' math scores in elementary school. A high proportion of students from low-income families enter elementary school with low literacy as measured by standardized tests and show lower academic performance throughout the elementary grades and beyond (Alexander et al., 1997; Sonnenschein et al., 2010).

Under ESSA (and its predecessor No Child Left Behind), all students who attend supplemental service are eligible for free- or reduced-price lunch, another assistant service for students from low-income families (Pennsylvania Department of Education, 2022; U.S. Department of Education, 2009). Though not all eligible students would attend supplemental service due to parents' choices, students who receive and participate in supplemental service are eligible under these criteria. This can be shown by several research studies that examined the effectiveness of supplemental service. All students who participated in one study were eligible for free or reduced-price lunch and came from low-income families (Deke et al., 2012; Muñoz et al., 2009). However, Springer, Pepper and Ghosh-Dastidar (2014) included eligibility for free- or reduced- lunch as a covariate when they matched and created comparable groups of students. Both studies identified eligibility for free- or reduced-price lunch as important factors for examining their participation in supplemental service, though their methods for identifying the influence varied.

Race and Ethnicity

Students from some racial and ethnic groups are more likely to receive supplemental service. Most research on supplemental service has used race or ethnic background as covariates; students from some ethnic groups have lower academic performance and thus tend to attend supplemental service to make academic progress (Muñoz et al., 2012; Springer et al., 2014). Muñoz, Chang and Ross (2012) include race as a covariate, though the ANCOVA analysis failed to indicate race as a significant covariate. Springer, Pepper and Gosh-Dastidar (2014) noticed that a more substantial percentage of black students enrolled and attended supplemental service while fewer white students were eligible for and attended supplemental service. Hispanic students enrolled in supplemental service were slightly less likely to participate in supplemental service sessions. Though the researchers did not check the disproportionality of these ethnic groups based on their attendance with statistical analysis, they included race and ethnicity groups as covariates when evaluating the effects of supplemental service (Springer et al., 2014).

When examining the effects of supplemental services, some researchers have included gender as a covariate (Harding et al., 2012; Muñoz et al., 2009, 2012). Muñoz, Chang, and Ross (2012) included gender as a covariate and found that students who received reading supplemental service had a more significant proportion of boys than students who didn't attend the service. Researchers didn't detect a gender difference on math supplemental service (Muñoz et al., 2012). In addition, researchers did not detect significant gender differences in students' academic achievement in Grade 3 (Harding et al., 2012). Though the correlations between gender and attendance of supplemental service or academic performance are not clear, in the current study, I checked for gender balance and included it as a covariate when matching students who received supplemental service to students who did not.

Prior Academic Achievement

Students' previous test scores were highly related to students' academic performance after receiving supplemental service. Some researchers included students' prior academic scores as covariates. Ross, Neergaard, Harrison, Ford and Paek (2009) viewed students' prior achievement scores as representative of students' learning ability. This, in turn contributed to students' learning progress and regressed students' scores on prior learning abilities and supplemental service providers, therefore finding a weak effect of supplemental service. Muñoz, Chang, and Ross (2012) assumed that students who attended supplemental service had lower academic scores and included prior academic performance as covariates in their ANCOVA. They assumed that students who participated in extra math service had lower math scores, students who attended additional reading service had lower reading scores, and students who attended both services had low math and reading scores. Their analysis found that supplemental service had an overall weak effect.

Some researchers viewed prior academic achievements as the result of factors in students' early learning experiences. For students who attended supplemental service, their low academic performance embodied several known or unknown drawbacks of previous learning. Springer, Pepper and Ghosh-Dastidar (2014) stated that students' academic test scores before attending supplemental service might capture some unobserved covariates of student achievement. They included pre-test scores as covariates in their propensity matching, estimated the fixed effect of supplemental service, and found that students who attended supplemental service made significant progress in math.

Primary Language/Home Language

Students' primary language and home language impact their academic performance. Students' early language experiences are highly related to their English literacy in Grade 1, such that these students show lower achievement in English literacy (Goldenberg, 2008; Haager & Windmueller, 2016). Muñoz, Chang, and Ross (2012) found that students who attended math services had a higher proportion of limited English proficiency. Deke, Dragoset, Bogen and Gill (2012) applied intent to treat (ITT) impact estimates, and their analysis didn't indicate English learners' intention to attend supplemental service. As primary and home language is related to attendance and the impact of supplemental service, the current research includes students' language experience as a covariate.

Disability Status

Some researchers included disability status as a covariate when they analyzed the impact of supplemental service on students' academic learning progress. Students with disabilities receive special education related to their disability type and attended supplemental service if eligible. Examples of disabilities include intellectual disability, hearing impairment, speech or language impairment, visual impairment, autism, traumatic brain injury, a specific learning disability, etc. (National Center for Education Statistics, 2022). Some students with disabilities are eligible for supplemental service due to their low family income. Muñoz, Chang and Ross (2012) checked the disability status in both student groups (received vs. not received supplemental service) and found that students who attend math supplemental service had a higher proportion of disability. Deke, Dragoset, Bogen, and Gill (2012) included disability status as a covariate and found that students with disabilities benefited less from math supplemental service. The current research examines whether the groups of students are comparable in their disability status.

Motivation and Engagement

Academic learning proficiency is not the only contributor to students' academic performance. When students have developed adequate learning proficiency, they need to fully use their in-school learning resources to gain equal growth as their peers. Ross, Potter, Paek et al. (2008) found that students' in-school environment contributed to students' learning progress significantly. Their finding was consistent with the academic competence learning model, developed by DiPerna (2002), that students' in-school behaviors and attitudes, such as motivation and engagement, contributed to students' academic outcomes. To make adequate learning progress, students benefit from prior knowledge and positive interactions with their teachers, peers, and schoolwork.

Motivation

Motivation underlies students' in-class learning processes. To reach higher academic achievement, students make achievement-related choices by their expectation of success and the value they place on the learning task (Wigfield & Eccles, 2002). Students' previous achievement-related experiences and their interpretations of these experiences contribute to their goal setting and affective responses. Subsequently, the children's goal and affective responses contribute to their expectations for success and perceptions of task value and may be more likely to use their in-school learning resources.

The current study examined students' motivation as a key outcome of supplemental service. I assume that when students attend supplemental service, which aims to promote their academic performance, it will increase students' motivation and help students make full use of their in-class learning resources as well. The Motivation subscale from Academic Competence Evaluation Scale (ACES) assesses student persistence in academic tasks and includes items pertaining to student responsibility and goal-directed behaviors.

Engagement

Students made academic progress by engaging with academic works (DiPerna et al., 2002). Researchers recorded students' academic related behaviors (e.g., persist in class, respond to teacher, attend discussion) to indicate students' academic engagement (Birch & Ladd, 1997; Finn et al., 1995; Skinner & Belmont, 1993). When students perform more academic related behaviors, they are more involved in the classroom, and tend to gain higher academic achievement in the class (Lee, 2014). Previous research has examined the effect of supplemental service on the individual student's learning engagement. The current study includes an engagement subscale that assesses a student's participation level in the classroom based on the students' frequency in asking and answering questions and engaging in leadership roles with peers.

Summary

Grade 1 students' prior academic achievement is highly correlated with demographics. Some children are more at-risk and could benefit from supplemental services. However, previous research has not consistently demonstrated the effect of supplemental service on students' academic outcomes. Possible explanations for these results may be due to the difficulty in controlling for covariates and the over-reliance on high-stake standardized tests. Previous research has only examined the academic achievement, but not motivation and engagement. The current research will control for demographic variables and prior academic achievement to evaluate whether supplemental service facilitates primary school students' academic achievement, motivation and engagement

Chapter 3

Method

This chapter describes the method of the study, including the introduction of data collection, a review of the instruments, and data screening. First, I describe the participants, measures, and acquiring procedures of the current data set. Second, I introduce two different data analytic approaches: 1) propensity score matching and its method, and 2) repeated measure ANOVA for comparing students from two conditions.

Participants

This study derived data collected during 2012 - 2013. The current data analysis only includes Grade 1 students from the original research sample. These 244 Grade 1 students came from five schools in the U.S. First grade classrooms within these schools were randomly assigned to receive SSIS-CIP intervention, a class-wide course that aimed to improve students' social skills and reduce problem behaviors.

Measures

Supplemental Service

The dataset includes students' receipt of supplemental services and the categories of the services. The three dominant supplemental service methods are Title 1, Instructional Support, and

Response to Intervention. Further, some students also attended other types of supplemental services, such as gifted, ESL, read to succeed, etc.

Demographic Variables

The demographic variables were collected from school records, including race/ethnicity, gender, primary language, home language, and disability status. Students' race and ethnicity were recorded with a multiple-choice question including alternatives - White, Black, Hawaiian and Pacific islander, Asian, Hispanic, other, and unknown – allowing participants to select more than one options.

Students' primary and home language were recorded as English, Spanish or other (participants described specific language when they selected 'other'). In the current study, as the learning outcome is indicated by English reading, I re-coded primary language and home language into English vs. non-English. Students whose primary language was not English were regarded as English learners.

The dataset had several questions about disability status and special education, surveying students' current and past disability status as well as specific types of disability. There was a question asking about the past-year enrollment in special education and another question asking about the current year enrollment in special education. After asking about special education enrollment, the teachers indicated the type of disability. Usually, a student diagnosed with a disability would receive special education service designed for this disability. The types of disability included speech/language impairment, learning disability, emotional behavioral disorder/serious emotional disturbance, autism, traumatic brain injury, deafness, deaf-blindness, hearing impairment, visual impairment (including blindness), and orthopedic impairment. As for students with learning disabilities, their disability status was indicated with another multiple-choice question showing their learning disability on specific learning topics: language (expressive or receptive), reading, writing, and mathematics. As for students with ‘other health impairment’, their disability was recorded with a follow-up question indicating the specific health impairment. In the current analysis, due to my interest in learning, I include learning disabilities that pertained to difficulty in reading and mathematics.

Academic Performance

Trained research assistants administered the STAR reading and math tests. STAR tests are computer-adaptive tests to measure students’ reading and mathematics performance. STAR tests accurately measure preschool to K-12 students’ reading and math achievement (Renaissance Learning, 2009, 2010) and can be administered up to five times in a year to track the growth of students’ academic performance longitudinally. In addition, with developed norms, test administrators could align students’ reading and math achievement to national norms and indicate whether the students have reached grade-level reading and math standards.

STAR reading assesses students’ reading comprehension and overall reading achievement through vocabulary-in-context test items (Renaissance Learning, 2010). In a test session, a student reflects on several vocabulary-in-context items on a computer, which requires background information, vocabulary knowledge, and reading strategies. The computer program determines the difficulty of the items based on students’ reflections on previous items. For example, if a student makes a mistake on an earlier item, they will next receive an easier item. After a 10-minute test session, the testing system generates a STAR Reading scaled score representing the student’s current reading skills. The STAR reading test has high reliability in assessing Grade 1 students’ reading comprehension, with split-half reliability at 0.88 and test-retest reliability at 0.89. The STAR test has high validity; both scaled scores are highly related to other reading comprehension tests (e.g., Stanford Achievement Test, $r = 0.73$). The STAR test score could represent students' academic performance in reading (Renaissance Learning, 2010).

STAR math test is a computer-adaptive test that helps teachers assess students' mathematical ability in 15 minutes (Renaissance Learning, 2009). The STAR math test is organized into eight strands – numeration concepts, computation process, estimation, geometry, measurement, data analysis and statistics, word problems, and algebra – which are frequently taught in schools and are matched to national standards (e.g., Principles and Standards for School Mathematics of the National Council of Teachers of Mathematics, the National Assessment of Educational Progress, Trends in International Mathematics and Science Study). The split-half reliability of the test was 0.82, and its alternate-form reliability is 0.73 on grade 1 students. The STAR math test is highly related to other math tests (e.g., Iowa Test of Basic Skills, $r = 0.56$), Metropolitan Achievement Test, $r = 0.55$).

Motivation and Engagement

Motivation and engagement were measured by subscales from the Academic Competence Evaluation Scale Teacher Record Form (ACES; DiPerna & Elliott, 2000). To assess the elementary school students' motivation and engagement, the teachers rated several items on a 5-point Likert-type scale, ranging from 1 (*never*) to 5 (*almost always*) (DiPerna & Elliott, 2000).

Student motivation was measured by the mean of 11 items (ranging from 1 to 5) which assessed a student's participation level in the classroom based on the students' frequency in asking and answering questions and engaging in leadership roles with peers. The reliability of motivation subscale was high, $\alpha = 0.94$, for elementary school students (DiPerna & Elliott, 1999, 2000). The test-retest reliability in a 2- to 3- week interval was 0.96.

Student engagement was represented by the mean of eight items which measures students' persistence in academic tasks and students' responsibility and goal-directed behaviors. The Engagement subscale had high internal consistency, $\alpha = 0.97$, for kindergarten through second-grade students in the standardization sample. In addition, the test-retest reliability in a 2- to 3- week interval was .92 (DiPerna & Elliott, 2000).

Procedures

Before data collection, researchers asked for consent from parents, teachers, and children. In October 2012 (Wave 1), researcher assistants administered the STAR Reading and Math tests to participating students with parental and child consent, and classroom teachers rated these students' motivation and engagement on ACES scales. Then, in the next four months, some students received a class-wide social-emotional learning curriculum while others did not. The curriculum was randomly assigned at the class level. In March 2013 (Wave 2) and May 2013 (Wave 3), all measures were completed again. The students' demographic variables and receipt of supplemental services throughout the year were collected in May 2013 as well.

For the current analysis I first checked for missing variables in the data set and reported descriptive data. Then, I used propensity score matching to balance the covariates over the students who attended and who did not attend supplemental service. Finally, after creating balanced and comparable student groups, I conducted a repeated measure ANOVA to compare students' learning outcomes, motivation, and engagement levels after attending supplemental service to students who did not receive supplemental service.

Data Analysis

Missing Data

I analyzed the missing pattern and noticed that the data contained a high proportion of missing values. In the 2012-2013 data, there were 244 students in Grade 1, and 139 cases had missing values on at least one variable. The math score in Oct. 2012 (Wave 1) had 27.45% missing data, and the reading score in Oct. 2012 (Wave 1) had 33.19% missing data. The missing pattern was not completely at random, Little's χ^2 = 466.635, df = 332, p < 0.001. Figure **3-1** shows the missing pattern.

Descriptive Data on Covariates

I conducted descriptive statistics on all covariates and examined the characteristics of the distribution of these variables. The number and percentage of participants on each variable level are reported for demographic variables, including gender, race/ethnicity, primary language, and home language (Table 3-1). The race/ethnicity indicators show two unique characteristics. In the original questionnaire, race/ethnicity was recorded by a multi-select question which allows more than one alternative each time. In the current data, students' races did not reach 100% because one student didn't report the race. In addition, there were no Grade 1 students with 'other' or 'unknown' race in the current sample. In this case, I would remove Race_other and Race_unknown from the covariates in matching. Hence, the demographic variables in current research would be gender (male vs. female), race/ethnicity (White, Black, Hawaiian and Pacific islander, Asian, Hispanic), primary language (English vs. non-English), home language (English vs. non-English), and learning disability (reading, mathematic, vs. none).

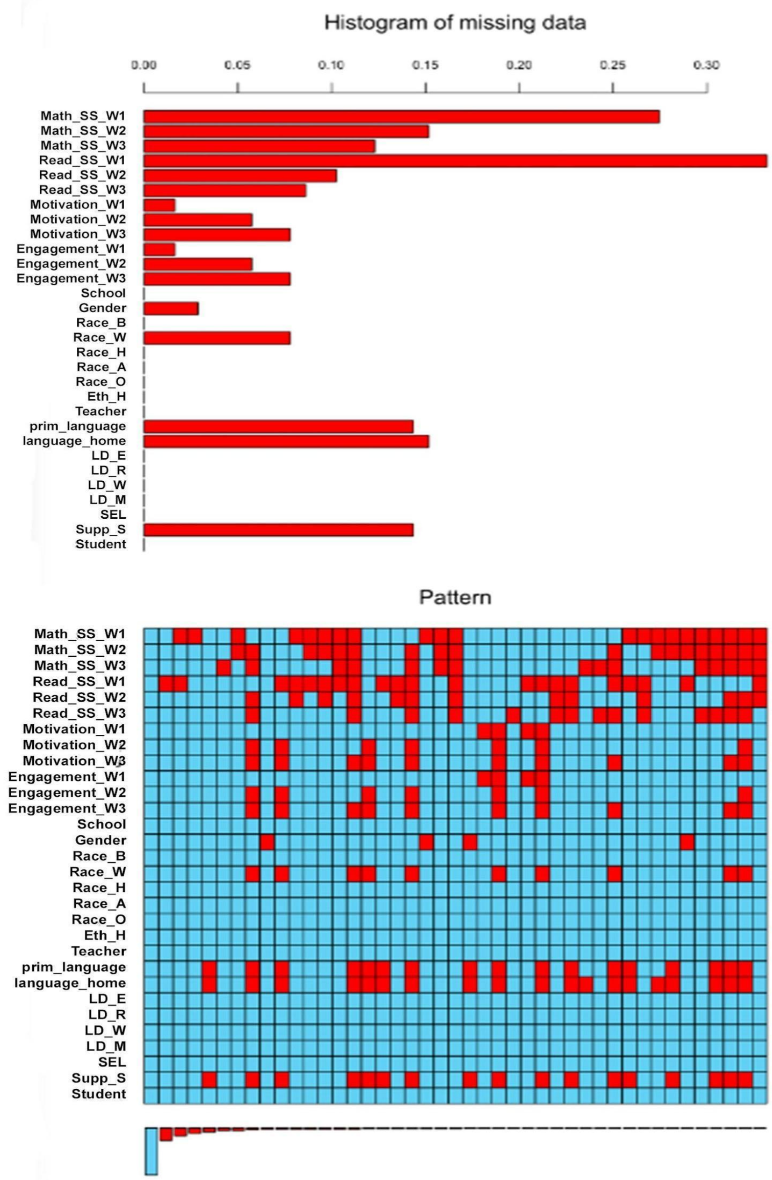


Figure **3-1**: Missing Values for Grade 1 Students.

Note. Total number of Grade 1 students N = 244. Math_SS_W1, Math_SS_W2, Math_SS_W3 refer to students' STAR math scaled score in October 2012 (Wave 1), March 2013 (Wave 2), May 2013 (Wave 3). Read_SS_W1, Read_SS_W2, Read_SS_W3refer to students' STAR read scaled score in October 2012 (Wave 1), March 2013 (Wave 2), May 2013 (Wave 3). Motivation_W1, Motivation_W2, Motivation_W3 refer to student's score on motivation subscale of ACES-T test in October 2012 (Wave 1), March 2013 (Wave 2) and May 2013 (Wave 3). Engagement_W1, Engagement _W2, Engagement _W3 refer to student's score on engagement subscale of ACES-T test in October 2012 (wave 1), March 2013 (Wave 2) and May 2013 (Wave 3). Race_B, Race_W, Race_H, Race_A, and Race_O, indicated whether a student is Black, White, Hawaiian or other Pacific islanders, Asian, and other. Eth_H indicates whether a student is Hispanic or not. Prim_language recorded students' primary language. Language_home indicates students' home language. LD_E, LD_R, LD_W, and LD_M indicated students' status of learning disability, including language (expressive or receptive), reading, writing, and mathematics. SEL is the social-emotional learning course, the intervention in the original study. Supp_S indicates students' current level of supplemental service, 0 = not attend, 1 = attended. Student refers to student ID in data set.

Table **2-1**: Demographic data for the Sample before Matching.

Covariate	N	Proportion
Gender		
Male	59	56.2%
Female	46	43.8%
Race/Ethnicity*		
White	84	80%
Black	15	14.3%
Hawaiian and Pacific Islander	1	1%
Asian	4	3.8%
Other	0	0%
Unknown	0	0%
Hispanic	3	2.9%
School		
School 1	48	45.7%
School 2	25	23.8%
School 3	8	7.6%
School 4	15	14.3%
School 5	9	8.6%
Primary language		
English	103	98.1%
Non-English	2	1.9%
Home language		
English	99	94.3%
Non-English	6	5.8%
Learning Disability		
Reading	2	1.9%
Mathematics	2	1.9%

Note. All race categories did not make up 100%.

For continuous variables in my covariates set (math scaled score and reading scaled score at Wave 1), I examined their distribution grouped by supplemental service (Figure 3-2). The scaled math score at Wave 1 is normally distributed, $M = 335.35$, $SD = 99.784$. The scaled reading score at Wave 1 is negatively skewed, $M = 117.32$, $SD = 74.341$.

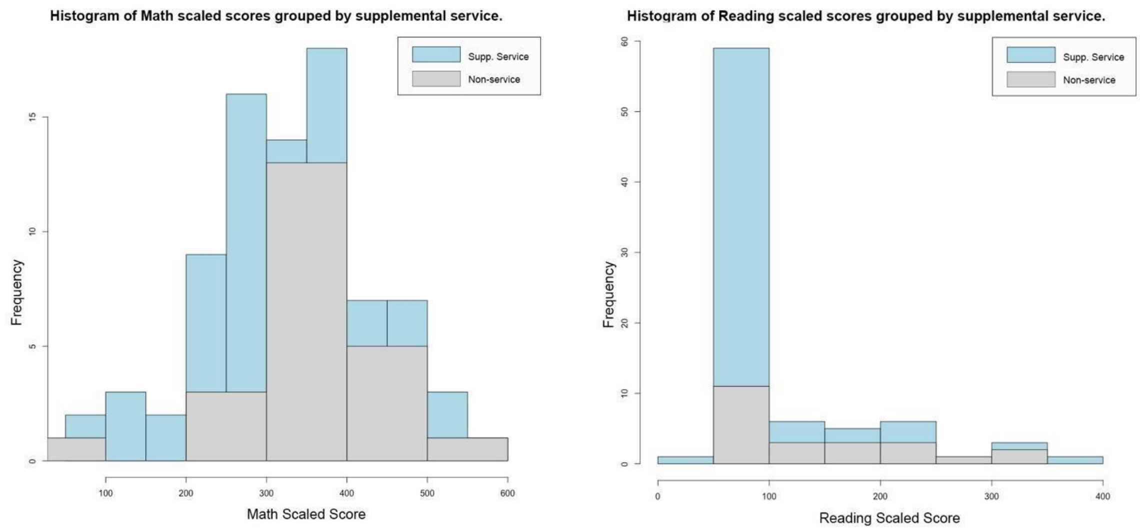


Figure **3-2**: Histograms for Covariates (*N* = 105).

Propensity Score Matching Models

The propensity matching was introduced by Rosenbaum (1969, 2020) to create a comparable 'control' group and 'treatment' group from observational data. In an experimental design, the students are randomly assigned to the control and treatment groups, and each student has equal probability of being assigned to each comparable group, $\pi_{control} = \pi_{treatment}$. If the students are already assigned to comparable groups (control group and treatment group), for example, there would be several covariates influencing the assignment, thus affecting the probability that student i is assigned to each group, $\pi_{control} \neq \pi_{treatment}$. However, for each student k from the control group with the probability of π_k, we could find a comparable individual l in the treatment group with the probability of π_l, and individual k and l satisfied the equation $\pi_k = \pi_l$. Then, we can assume that this pair of cases are comparable, and the intervention is randomly assigned to the groups.

The propensity score matching in this research is operated with MatchIt package 4.4.0 in R(Ho et al., 2011). With propensity matching, I am trying to find the comparable ‘pairs’ of the students.

First, I calculated the propensity score with several covariates related to students’ assignments. The covariates included in the analysis were gender (male vs. female), race/ethnicity (White, Black, Asian, Hispanic, etc.), primary language (English vs. non-English), home language (English vs. non-English), learning disability (abled vs. reading disability vs. mathematic disability), prior math score, and prior reading score. For variables that had more than one level, for example, learning disability, it was dummy coded into multiple dummy variables. The estimate of propensity score could be addressed with a generalized linear model that predicts probability based on several continuous and discrete variables. The formula for estimating the probability is

$$\hat{p} = \frac{1}{1+e^{B_0+B_1X_1+B_2X_2+\cdots\ldots+B_kX_k}},$$

where $X_1, X_2, \ldots\ldots, X_k$ refer to covariates in propensity score matching, and $B_0, B_1, B_2, \ldots\ldots, B_k$ are generated parameters in prediction.

Second, I used the MatchIt package in R 4.4.0 to match individuals from different groups. According to my research purposes that focus on educationally deprived students, the estimation focuses on the average treatment effect in the treated group (ATT) in matching. Therefore, all students who received supplement service would be kept in the ‘treatment’ group, then students who didn’t receive supplemental service would be matched and assigned to the ‘control’ group(Ho et al., 2007; Stuart, 2010). Three methods are compared and selected in matching: nearest neighbor matching, optimal pair matching, and full matching (which is described in the following paragraph). After creating a comparable group, I reevaluated the balance between the two groups with the effect size of the differences expecting that the mean differences across the groups are approximately 0.

The current study applied three matching methods and selected the method that created the most balanced groups. Nearest neighbor matching runs through all individuals in the treatment group and finds the case with the closest propensity score (Stuart, 2010), which is the most common method in matching (Thoemmes & Kim, 2011; Zakrison et al., 2018). Optimal pair matching also runs through all individuals in the treatment group but attempts to eliminate the overall paired differences across both groups. Optimal full matching creates inside subclasses which consist of two cases with approximately the same propensity scores and drop no case from unmatched groups. The balance, indicated by the mean difference between both groups, reveals the most ideal matching when the mean is similar to 0. In the educational field, the acceptable balance is less than 3% or 5%.

Evaluating the Effect of Supplemental Services

To evaluate the students' change in math and reading scores, motivation, and engagement over the three time points, I used repeated measure ANOVA with SPSS to compare the difference between students who received supplemental to students who did not. Each outcome - math performance, reading performance, motivation, and engagement – was examined independently. For each analysis, the performance of a student Y_{ijk} consisted of mean level of prior development, growth in the time period, effect of supplemental service and interaction between supplemental service and time,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

(μ refers to the grand mean; α_i refers to the effect of supplemental service enrollment, i = 0 or 1; β_j refers to the impact of time, e_{ijk} refers to the random error in the model). The time/wave is a within-subject variable which has three levels, wave 1, wave 2 and wave 3. The enrollment of supplemental service is a between-subject variable which has two levels for attendance at supplemental service: yes or no.

To evaluate students' growth across time points, I used a regression model through SPSS. In the regression model, I predicted students' post academic performance, motivation, or engagement with their enrollment of supplemental service and their prior academic performance or learning approach,

$$\widehat{Y_{ij}} = B_0 + B_1 Supp + B_2 Prior.$$

If supplemental service facilitates the growth of supplemental service, the slope of supplemental service B_1 would be significant.

Chapter 4

Results

Descriptive Analysis

Supplemental Service

After removing cases with missing data, there were 105 Grade 1 cases left for analysis, including 23 Grade 1 students who attended supplemental service in the 2012-2013 academic year. Of these 23 Grade 1 students, 17 attended Title 1 service, two received instructional service, three received response to intervention, nine received tutoring, and six received other supplemental services, including enrichment, ESL, and speech. In addition, eight students received two categories of supplemental service, and two students received four categories of supplemental service.

Balance on Covariates

Before the group matching, I checked the balance on all the covariates across the two groups. All means and standard deviations are shown in Table 4-1. Comparing the mean of students across groups with independent t-test (two-tailed), the statistical results indicate that only race and prior reading score are significantly different across the groups.

Table **4-1**: Balance for Covariates on Grade 1 Unmatched Sample.

	Non-supp. service (N = 82)		Supp. service (N = 23)			
Covariates	*M*	*SD*	*M*	*SD*	*t*-test	*p*
Gender (Male)	0.585	0.496	0.478	0.511	0.910	0.183
Race (Black)	0.171	0.379	0.435	0.209	2.110	0.019**
Race (White)	0.780	0.416	0.870	0.344	-1.045	0.151
Race (Hawaiian and islander)	0.012	0.110	0.000	0.000	0.528	0.299
Race (Asian)	0.037	0.189	0.043	0.209	-0.151	0.440
Ethnicity (Hispanic)	0.024	0.155	0.043	0.209	-0.481	0.316
Primary lang. (Eng.)	0.988	0.110	0.957	0.209	0.965	0.168
Language home (Eng.)	0.951	0.217	0.913	0.288	0.692	0.245
Learning Disability (reading)	0.012	0.110	0.043	0.209	-0.965	0.168
Learning Disability (Math)	0.488	0.503	0.696	0.470	-0.965	0.168
Prior Math	329.878	103.272	354.870	85.393	-1.062	0.145
Prior Reading	109.439	71.214	145.435	79.978	-2.085	0.002**
Motivation (wave 1)	3.543	1.073	3.728	1.083	-0.722	0.475
Engagement (wave 1)	3.668	1.045	3.739	1.027	-0.294	0.771

Note. The covariate is significantly different across the two groups, $p < 0.05$. Math and reading score used are STAR scaled scores at wave 1.

Matching

I calculated the probability that each individual would enter the treatment group, propensity score, with a logistic regression model and covariates. The formula is:

$$\hat{p} = \frac{1}{1 + e^{B_0 + B_1 Male + B_2 Race_B + B_3 Race_W + B_4 Race_H + B_5 Race_A + B_6 Eth_H + B_7 PrimLang + B_8 LangHome + B_9 LD_M + B_{10} LD_R + B_{11} Math + B_{12} Read}}.$$

The propensity score ($\hat{p}$) for all participants in the whole sample has a mean of 0.22 and a standard deviation of 0.13.

Next, I conducted nearest neighbor matching, optimal paired matching and optimal full matching methods based on propensity scores. Under each propensity scores matching method, I visualized the outcome of the balance.

After nearest neighbor matching, 59 students who didn't receive supplemental service were removed because their propensity scores were not matched with students who had received supplemental service (as Figure 4-1 indicated), and 23 matched students entered both groups. The standardized mean difference showed that reading score, gender, learning disability, primary language, home language, and race/ethnicity (particularly Black and Hispanic) were matched and balanced across the groups. The balance on prior math score was problematic, with the absolute standardized mean difference = -0.0642. The balance across some racial groups (Asian and White) was not achieved, with absolute standardized mean difference = 0.213, -0.129.

After optimal pair matching, 59 students who didn't receive supplemental service were removed because their propensity scores did not match with those who had received supplemental service (as Figure 4-2 indicated), with 23 matched students left in both groups. The balance of prior reading and math was successful, with the absolute standardized mean difference = 0.013, -0.0117. Several covariates, including race/ethnicity, learning disability and gender, matched perfectly with an absolute standardized mean near 0. The balances of primary language and home language were problematic, with absolute standardized mean difference = -0.213, -0.154.

After optimal full matching, no students were removed from the sample (see Figure 4-3). The balance was improved by weighting each case, creating new "control" and "treatment" groups with weighted cases. The balance of multiple covariates was problematic, including learning disability (reading and math), primary language, home language, and Asian racial/ethnic identity, with effect sizes of mean difference more than 0.1. The math score and reading score were not perfectly balanced either, with standardized mean difference = 0.067, 0.038.

Comparing the balance after three propensity matching methods, the best method for current data was optimal pair matching. Though in this case, optimal pair matching failed to match participants on all covariates, it successfully created groups with comparable prior math and reading scores, which are the main variables in supplemental service enrollment. The optimal pair matching dropped 59 cases from the “control” group; therefore, creating two comparable groups which included 23 cases in each group.

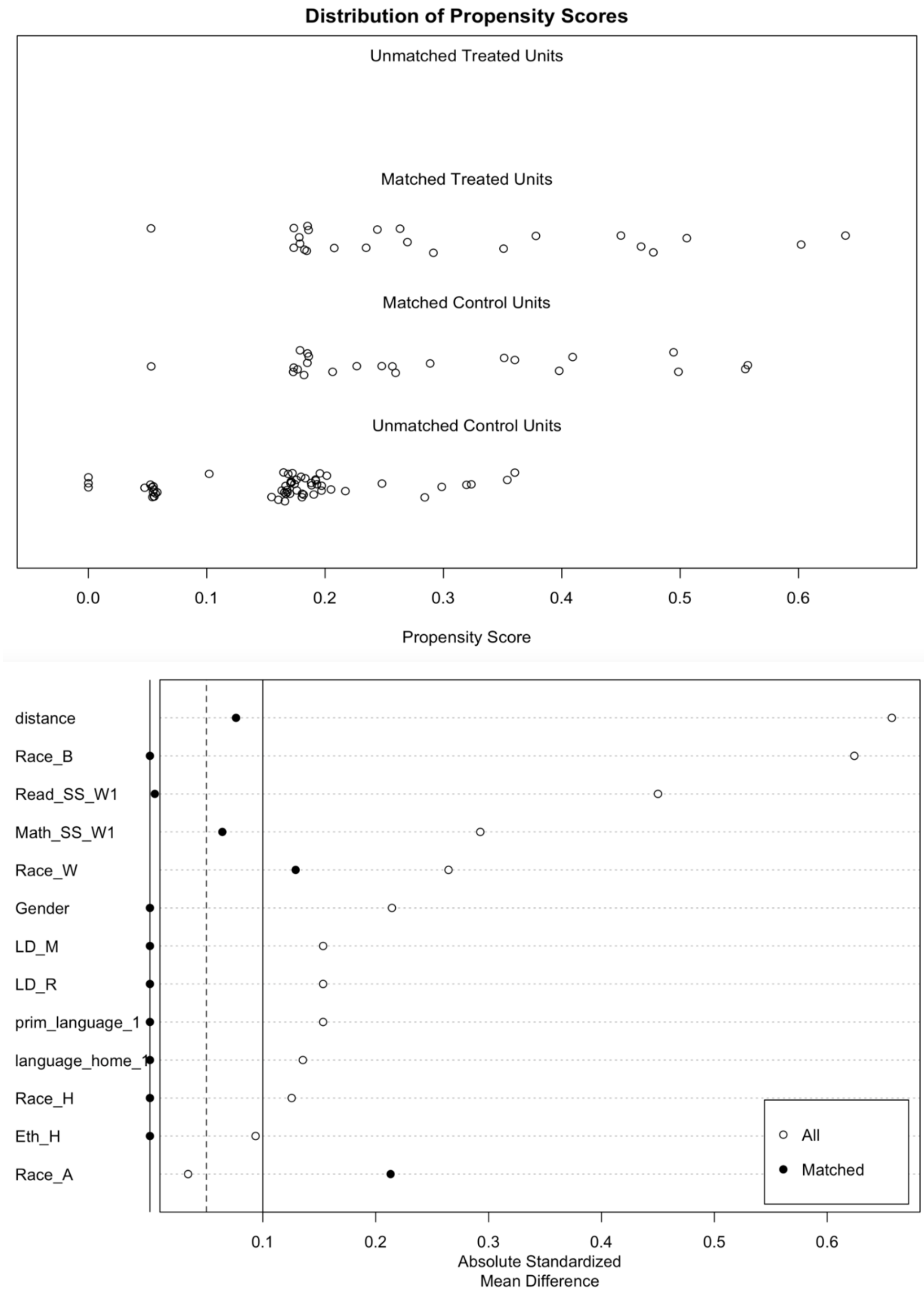


Figure **4-1**: Visualizations for Comparing Balance Before and After Nearest Neighbor Matching.

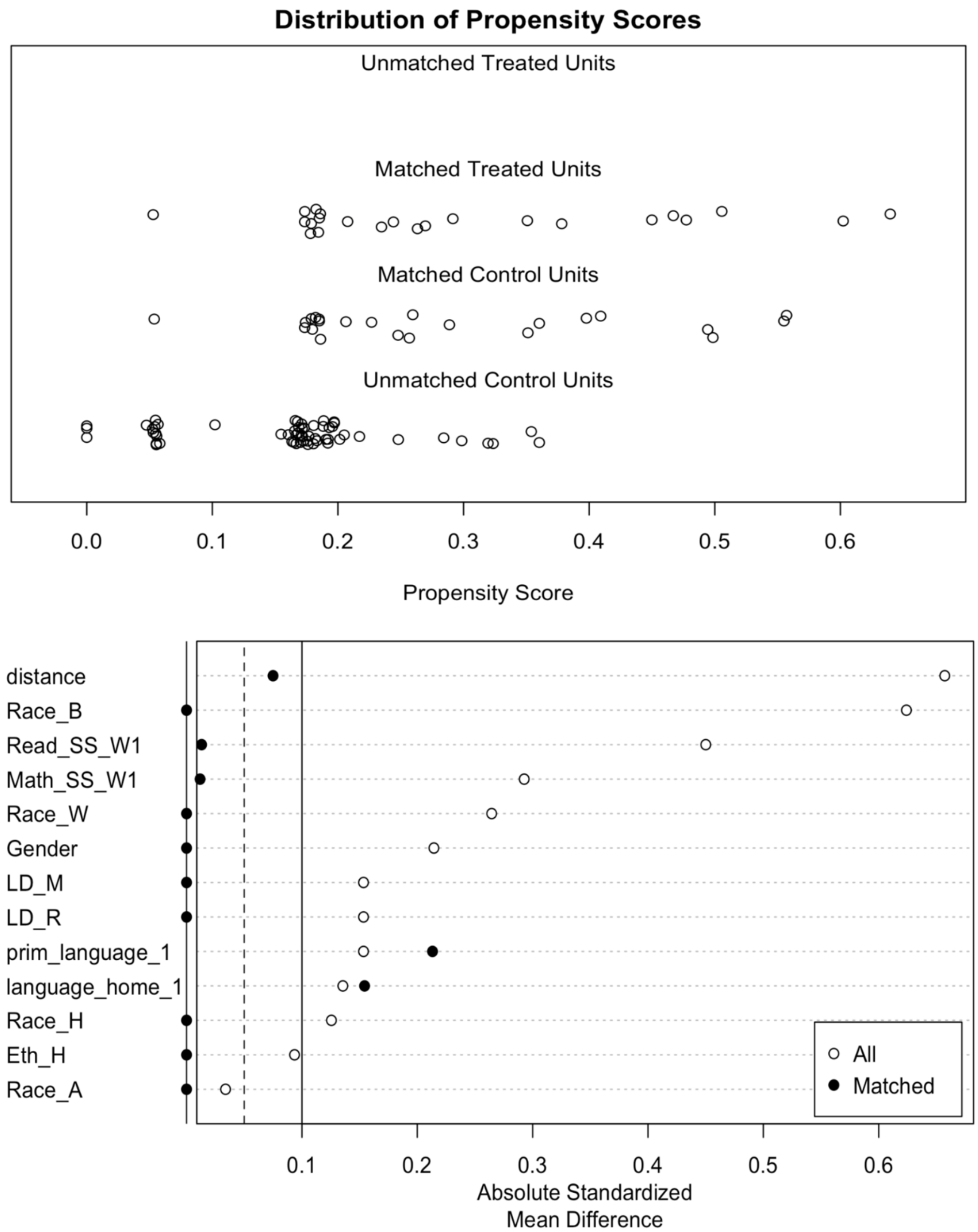
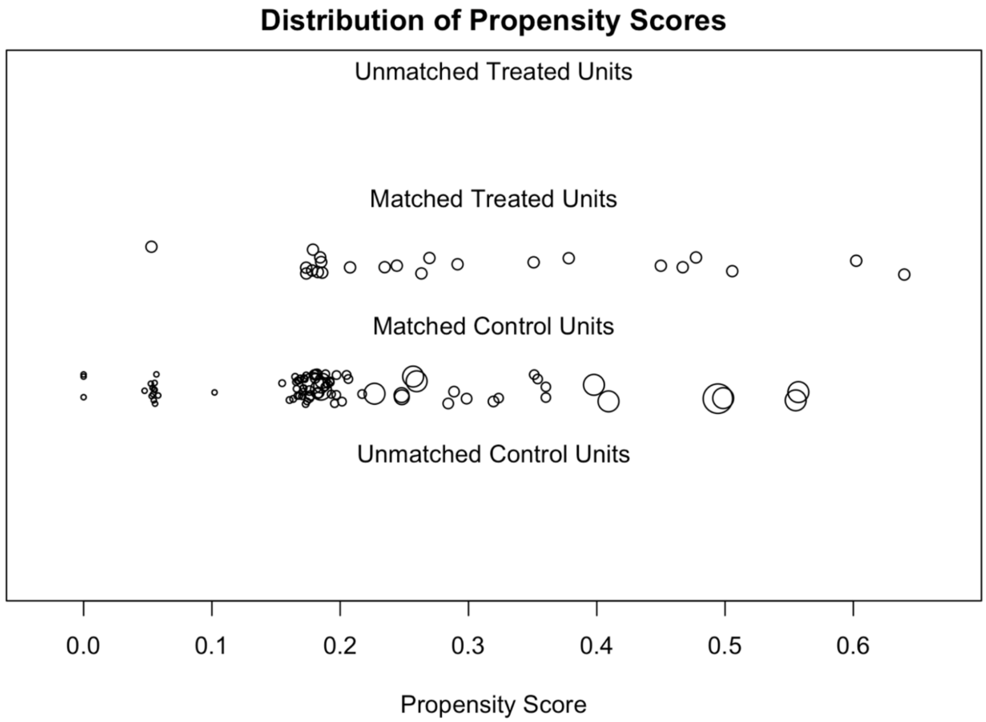


Figure **4-2**: Visualizations for Comparing Balance Before and After Optimal Pair Matching.



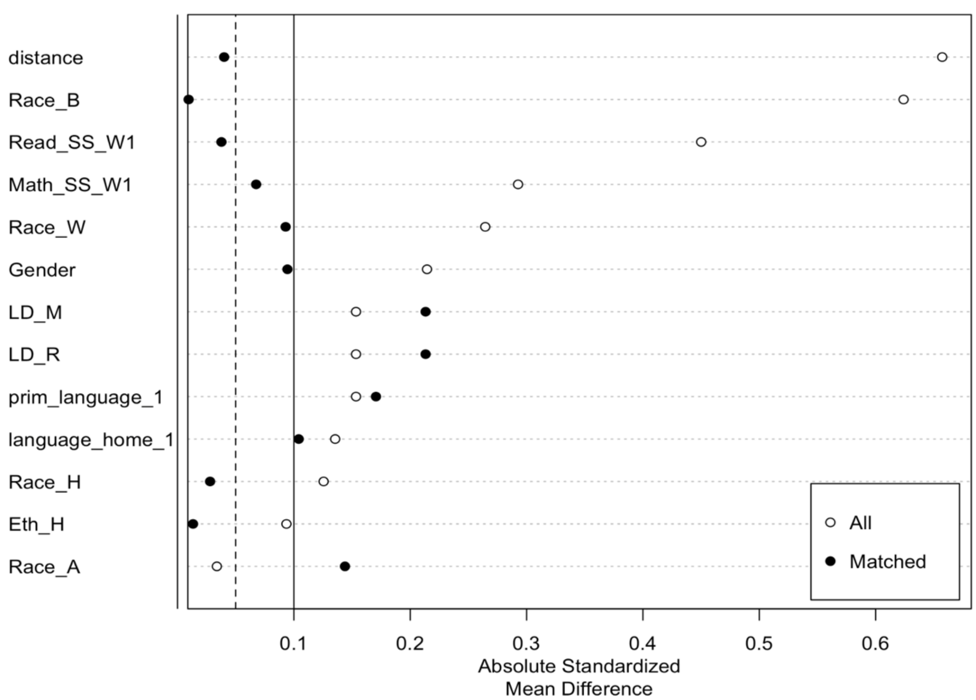


Figure **4-3**: Visualizations for Comparing Balance Before and After Optimal Full Matching.

Effects of Supplemental Service

After students from the two groups were matched on the covariates, I used repeated measure ANOVA (Cohen, 2013) with SPSS to compare students' performance on four outcomes (math achievement, reading achievement, motivation and engagement).

Math Achievement

I checked assumptions of repeated measure ANOVA. The math scores were not normally distributed with the Shapiro-Wilk test < 0.05. The correlation between residual and observed score was linear.

Repeated measures ANOVA was used to examine students' math scaled scores between the two groups. The interaction between wave and supplemental service was not statistically significant, F (2, 88) = 0.576, p = 0.564. The main effect of the wave, however, was significant, F (2, 88) = 37.840, p < 0.001. The main effect of supplemental service, on the other hand, is not significant, F (1, 44) = 0.359, p = 0.552. A Supp*wave line chart is also included in Figure 4-4.

At wave 1, when all demographic variables and academic performance were matched, the two groups of students had similar math scores, p = 0.970. At Wave 2, students who didn't receive supplemental service (M = 355.87, SD = 96.233) had higher score than students who received (M = 354.87, SD = 85.393), though the difference was not significant, p = 0.464, Cohen's d = 0.011. At Wave 3, students who didn't receive supplemental service (M = 461.78, SD = 100.93) also had higher score than students who received (M = 437.35, SD = 115.50), though again, the difference was not significant, p = 0.449, Cohen's d = 0.225.

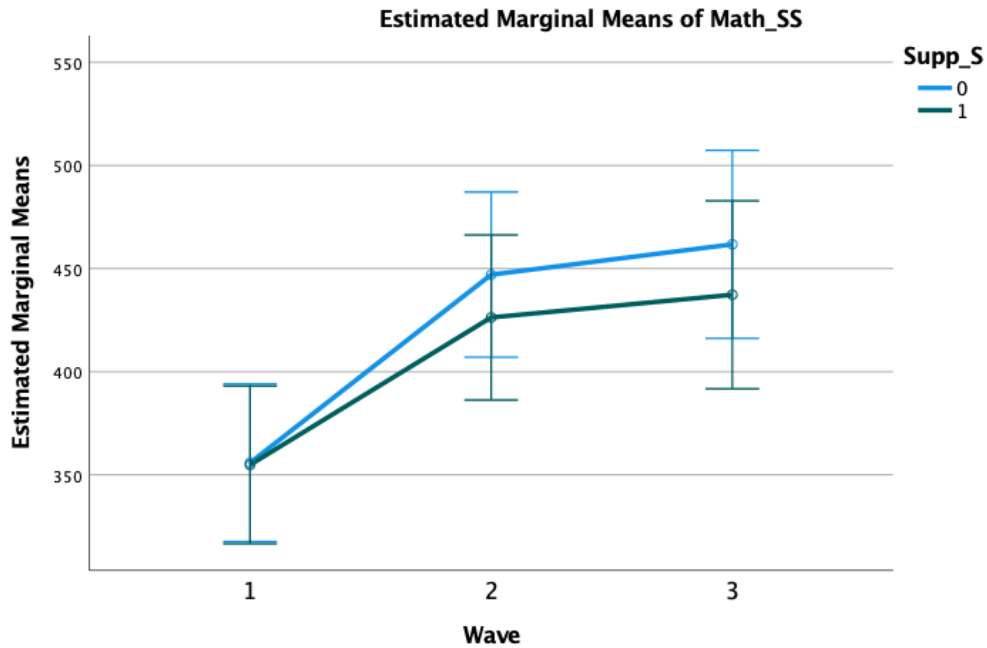


Figure **4-4**: STAR Math Score at Three Timepoints.

Note. Supp_S = 0 refers to students who did not attend supplemental service. Supp_S = 1 refers to students who attended supplemental service.

Error bars refer to 95% confidence interval in comparing means of the groups at each time point.

The significant main effect of wave indicated improvements in students' math achievement over time, and the difference between waves could be analyzed by pairwise comparison. Students who received supplemental service made significant progress on reading from wave 1 (M = 354.870, SD = 85.393) to Wave 2 (M = 426.348, SD = 99.518), $p < 0.001$. Students who received supplemental service did not make significant progress on reading from wave 2 (M = 426.348, SD = 99.518) to Wave 3 (M = 437.348, SD = 115.500), $p = 0.505$. Students who didn't receive supplemental service made significant progress on math from wave 1 (M = 355.870, SD = 96.233) to Wave 2 (M = 447.087, SD = 90.784), $p < 0.001$. However,

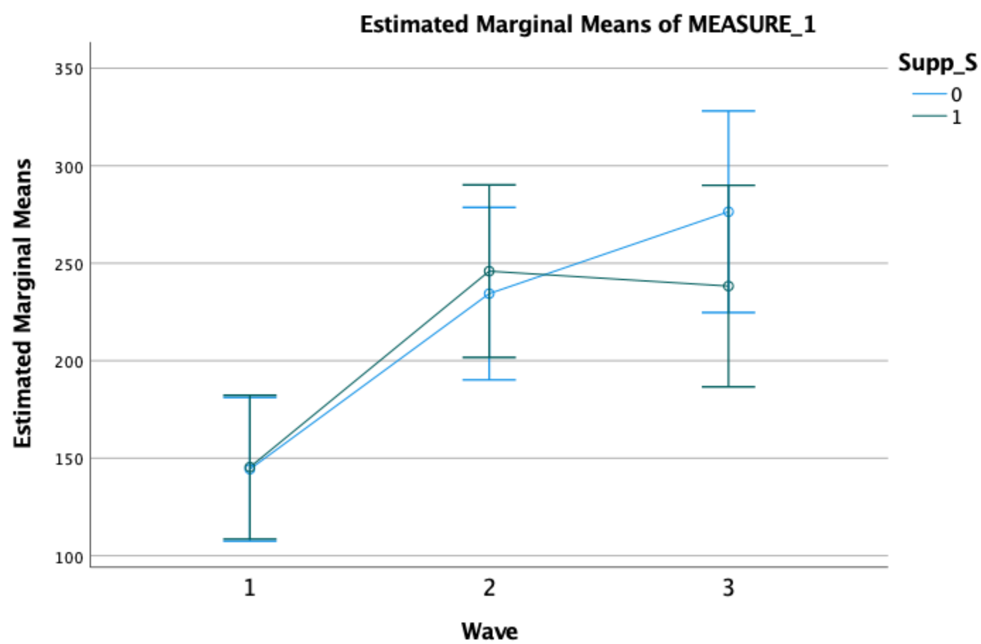


Figure **4-5**: STAR Reading Score at Three Timepoints.

Note. Supp_S = 0 refers to students who did not attend supplemental service. Supp_S = 1 refers to students who attended supplemental service.

Error bars refer to 95% confidence interval in comparing means of group at each time point.

students who didn't receive supplemental service did not make significant progress on math from Wave 2 (M = 447.087, SD = 90.784) to wave 3 (M = 461.783, SD = 110.928), p = 0.374.

Reading Achievement

First, I checked assumptions of repeated measure ANOVA. The math scores of two groups were not normally distributed at wave 1, Shapiro-Wilk test < 0.05. The correlation between the residual and observed score was linear. Then, I conduct a line chart of the means of the student groups at all three time points (Figure 4-5).

Repeated measure ANOVA examines students' math scaled scores over two groups. The interaction between wave and supplemental service is statistically significant, F (2, 88) = 3.095, p = 0.05. The main effect of the wave is significantly different over time, F (2, 88) = 66.347, p < 0.001. The main effect of supplemental service, on the other hand, is not significant F (1, 44) = 0.087, p = 0.769. The Supp_S*Wave graph shows that students' supplemental service and wave have an intersection from Wave 2 to Wave 3.

The significant main effect of wave indicated improvements in students' reading achievement over time, and the difference between waves could be analyzed by pairwise comparison. Students who received supplemental service made significant progress on reading from Wave 1 (M = 145.435, SD = 18.283) to Wave 2 (M = 245.957, SD = 21.949), p < 0.001. However, students who received supplemental service did not make significant progress on reading from Wave 2 (M = 234.435, SD = 21.949) to Wave 3 (M = 238.304, SD = 25.646), p = 0.505. Students who didn't receive supplemental service made significant progress on reading from Wave 1 (M = 144.391, SD = 18.283) to Wave 2 (M = 234.435, SD = 21.949), p < 0.001. Students who didn't receive supplemental service did not make significant progress on reading from Wave 2 (M = 234.435, SD = 21.949) to Wave 3 (M = 276.435, SD = 25.646), p = 0.003 < 0.05.

At Wave 1, when all demographic variables and academic performance were matched, the two groups had similar reading scores, p = 0.966. At Wave 2, students who didn't receive supplemental service (M = 234.43, SD = 103.423) had lower scores than students who received supplemental service (M = 245.96, SD = 107.069); however, the difference was not significant, p = 0.712, Cohen's d = 0.110. At Wave 3, students who didn't receive supplemental service (M = 276.43, SD = 117.39) had higher score than students who received supplemental service (M = 238.30, SD = 128.352), though the difference again was not significant, p = 0.299, Cohen's d = 0.310.

I used a regression model to test the contribution of supplemental service on students' academic performance at Wave 3. The regression model for predicting the students' score at Wave 3 was:

$$\widehat{Y_{ij}} = 44.033 + 0.991\,Math\_W2 - 49.552\,Supp.S$$

(Y_{ij} refers to students' reading scaled score at wave 3).

The model explained 72.7% variance of the Wave 3 reading score, and the supplemental service explained 4.1% variance uniquely. Wave 2 reading score predicted student Wave 3 reading score, t (43) = 10.510, p < 0.001. Supplemental service predicted student Wave 3 reading score as well, t (43) = - 0.203, p = 0.015 < 0.05.

The pairwise comparison were used to analyze the difference between students across each pair of timepoints. Students who received supplemental service made significant progress on reading from Wave 1 (M = 143.435, SD = 79.978) to Wave 2 (M = 245.967, SD = 107.069), p < 0.001. Students who received supplemental service didn't make significant progress on reading from Wave 2 (M = 245.967, SD = 107.069) to Wave 3 (M = 238.304, SD = 128.352), p = 0.576. Students' who didn't receive supplemental service made significant progress on reading from Wave 1 (M = 144.391, SD = 94.762) to Wave 2 (M = 234.435, SD = 103.423), p < 0.001. Students who didn't receive supplemental service made significant progress on reading from Wave 2 (M = 234.435, SD = 103.423) to Wave 3 (M = 276.435, SD = 117.388), p < 0.001.

Motivation

First, I checked the assumptions of repeated measure ANOVA. The motivation scores of two groups were not normally distributed across Wave 1, Wave 2, and Wave 3, Shapiro-Wilk test < 0.05. The correlation between the residual and observed score was linear. Then, I conduct a line chart consisting of the means of the student group at each time point (Figure 4-6).

I use repeated measures ANOVA to examine the effect of supplemental service on students' motivation. The interaction term was not significant, F (2, 88) = 0.045, p = 0.956. The main effect of supplemental service was not significant either, F (1, 44) = 0.005, p = 0.943. At Wave 1, when all demographic variables and prior scores are matched, the students who didn't attend supplemental service (M =3.735, SD = 1.106) had slightly higher motivation than students who attended supplemental service (M =3.727, SD = 1.083). The difference of motivation across the groups was not significant, p = 0.981. The Supp_S*Wave graph shows that students' supplemental service and wave intersected from Wave 1 to Wave 2 on motivation. At Wave 2, the students who didn't attend supplemental service (M = 3.874, SD = 0.992) had slightly lower motivation than students who attended supplemental service (M = 3.913, SD = 1.028) though this difference was not statistically significant, p = 0.896, Cohen's d = 0.038. At Wave 3, the students who didn't attend supplemental service (M = 4.004, SD = 1.015) had slightly lower motivation than students who attended supplemental service (M = 4.036, SD = 1.064), though again not significant, p = 0.918, Cohen's d = 0.034.

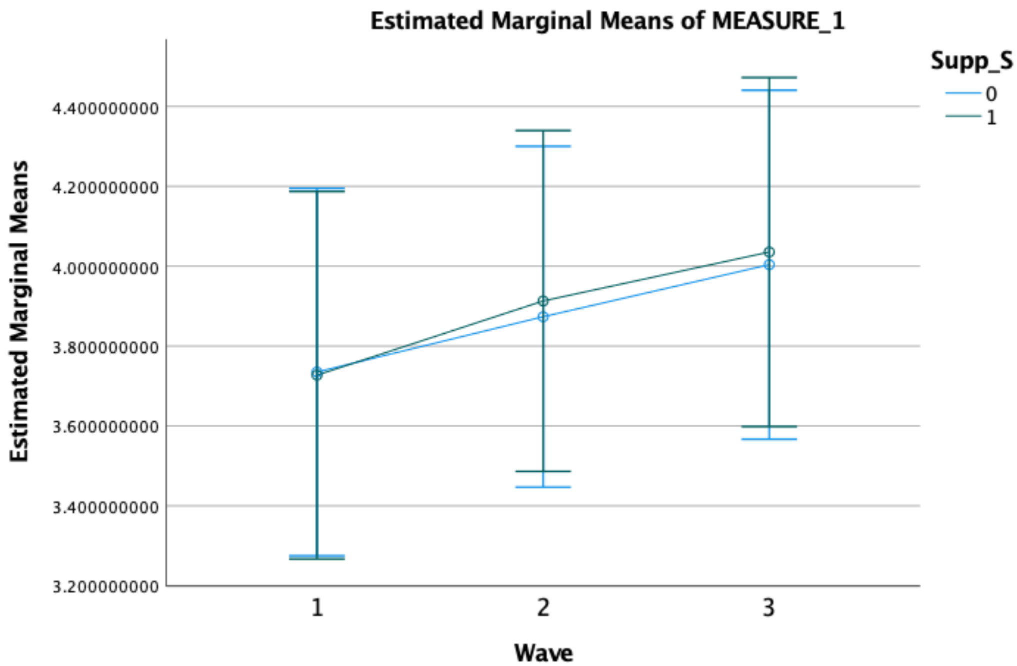


Figure **4-6**: Student Motivation at Three Timepoints.

Note. Supp_S = 0 refers to students who did not attend supplemental service. Supp_S = 1 refers to students who attended supplemental service.

Error bars refer to a 95% confidence interval in comparing the means of group at each time point.

Students' motivation improved across time and pairwise comparisons indicated the difference between time points. For students who received supplemental service, their mean motivation at Wave 2 was slightly greater than their mean motivation at Wave 1, $p = 0.119$, and their motivation at Wave 3 was slightly greater than motivation at wave 2, $p = 0.294$. These students' motivations at Wave 3 were moderately and significantly greater than their motivation at Wave 1, $p = 0.019$, Cohen's $d = 0.584$. For students who didn't receive supplemental service, their mean motivation at Wave 2 was slightly greater than their mean motivation at Wave 1, $p = 0.243$, and their motivation at Wave 3 was slightly greater than motivation at Wave 2, $p = 0.265$.

These students' motivation at Wave 3 was significantly greater than their motivation at Wave 1, $p = 0.040$, Cohen's $d = 0.634$.

Engagement

First, I checked the assumptions of repeated measures ANOVA. Students' engagement from the 'control' group was normally distributed at Wave 1, Shapiro-Wilk test = 0.123, while students' engagement scores of groups were not normally distributed at the other two time points, Shapiro-Wilk test = 0.123. The correlation between residual and observed score was linear. Then, I created a line chart with the means of the student groups at each time point (Figure 4-7).

I used repeated measure ANOVA to examine the effect of supplemental service on students' motivation. The interaction term was not significant, F (2, 88) = 0.323, $p = 0.725$. The main effect of supplemental service was not significant either, F (1, 44) = 0.001, $p = 0.972$. The main effect of wave was significant, F (1, 44) = 10.316, $p = 0.002$.

At Wave 1, when all demographic variables and prior scores were matched, the students who didn't attend supplemental service ($M = 3.755$, $SD = 0.991$) had slightly higher engagement than students who attend supplemental service ($M = 3.739$, $SD = 1.027$). The difference of engagement across the groups was not significant at Wave 1, $p = 0.957$. At Wave 2, the students who didn't attend supplemental service ($M = 4.065$, $SD = 0.922$) had slightly higher engagement than students who attended supplemental service ($M = 4.027$, $SD = 0.872$), $p = 0.886$. At Wave 3, the students who didn't attend supplemental service ($M = 4.125$, $SD = 0.858$) had slightly lower (but not statistically significant) motivation than students who attended supplemental service ($M = 4.207$, $SD = 0.781$), $p = 0.738$.

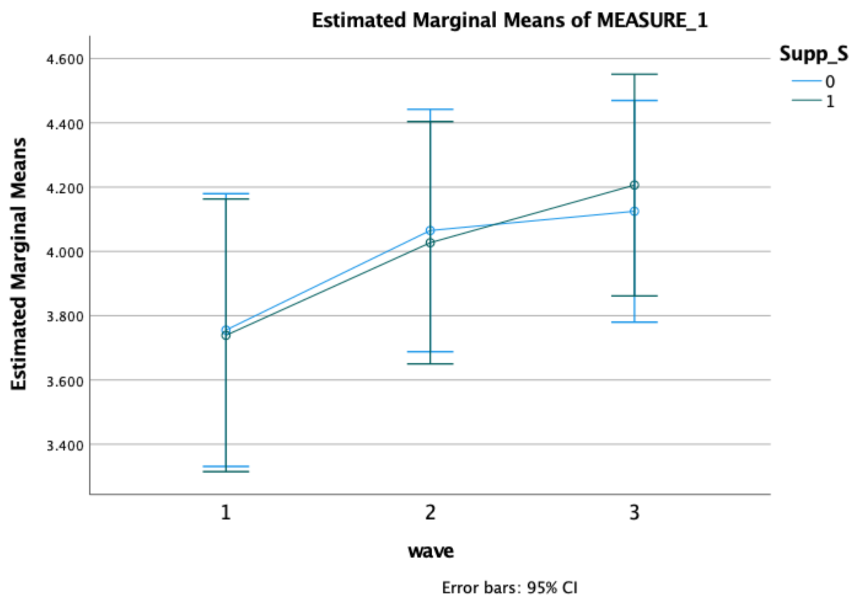


Figure **4-7**: Student Engagement at Three Timepoints.

Note. Supp_S = 0 refers to students who did not attend supplemental service. Supp_S = 1 refers to students who attended supplemental service.

Error bars refer to 95% confidence interval in comparing means of group at each time point.

Students' engagement improved across the time, and the pairwise comparisons were used to compare the difference between time points. For students who received supplemental service, their mean engagement at Wave 2 was significantly greater than their mean engagement at Wave 1, $p = 0.016 < 0.05$, Cohen's $d = 0.507$, their engagement at Wave 3 was greater than engagement at Wave 2, $p = 0.061$, Cohen's $d = 0.351$. These students' engagement at Wave 3 was significantly greater than their motivation at Wave 1, $p < 0.001$, Cohen's $d = 0.637$. For students who didn't receive supplemental service, their mean engagement at Wave 2 was significantly greater than their mean engagement at Wave 1, $p = 0.010$, Cohen's $d = 0.539$, and their engagement at Wave 3 was slightly greater than engagement at Wave 2, $p = 0.093$. These students’ engagement at Wave 3 was significantly greater than their engagement at Wave 1, p = 0.005, Cohen’s d = 0.562.

The Supp_S*Wave graph shows that students’ supplemental service and waves had an interaction from Wave 2 to Wave 3 on engagement. I regressed Wave 3 engagement on Wave 2 engagement and supplemental service and compared the regression model with regressing Wave 3 engagement on Wave 2 engagement. The supplemental service uniquely explained 5% variance of student engagement at Wave 3, but did not significantly predict student engagement at Wave 3, F (1, 43) = 0.362.

Chapter 5

Discussion

The purpose of the current study was to examine the effects of supplemental service on students' academic outcomes. Specifically, I analyzed students' academic performance across three timepoints in Grade 1 after matching students' academic score and demographic variables at the beginning of the year. With propensity scores matching, students' prior academic performance and some demographic variables were matched successfully. The results indicated that students who attended supplemental service made slightly less progress than their peers on math and reading scores but achieved similar levels of motivation and engagement to their peers.

Math and Reading Performance

For math and reading performance, when students' scores at Wave 1 are matched, the students who attended supplemental service neither made more progress nor demonstrated significantly higher scores than their peers who didn't participate in supplemental service. This result is consistent with previous research that failed to demonstrate the effectiveness of supplemental service on academic performance (e.g., Muñoz et al., 2008, 2012; Ross et al., 2009). To answer my research question, supplemental service did not improve Grade 1 students' academic achievement in math or reading based on the data from the current study.

In the current study, supplemental service influences students' reading score differently in each time period. The statistical analysis showed that supplemental service did not influence students' reading progress from Wave 1 (October) to Wave 2 (March) but was associated with decreases in students' reading progress from Wave 2 (March) to Wave 3 (May). As previous studies examined the effect of service in a more extended period (more than one year), this short period change could be due to reduced programming and participation at the end of the school year. Unfortunately, the available data did not include how often students received services (and for how long) throughout the school year, so it is difficult to know if a reduction in services occurred.

In current analysis, the effect of supplemental service on reading and math are not the same. Students' math progress did not change when receiving supplemental service, but students' reading progress negatively changed at the end of the year. Unfortunately, another limitation of the current dataset is that the focus (reading, math, language) of supplemental services was not reported. As a result, it is unknown if reading or math was more commonly targeted, or not even targeted at all, for each student receiving services. Thus, it is difficult to know if the "intervention" of supplemental service was well-aligned with the academic outcomes featured in this study.

Motivation and Engagement

After controlling the students' demographic variables and prior academic achievement, the students who received supplemental service demonstrated a similar level of motivation and engagement to their peers who did not attend supplemental service. For students with similar level of prior academic achievements and demographic variables, they demonstrated similar levels of motivation and engagement when entering elementary school. The motivation and engagement of both groups increased across their first-grade year, and inconsistent with the academic performance findings, supplemental service did not improve students' motivation and engagement beyond gains made by students who did not receive services.

Implications

The current study has two key implications for education research method and practice related to supplemental service. With regard to research, the current study demonstrated the use of propensity score matching to created comparable groups on several covariates, including demographic variables and academic achievement. With propensity score matching, researchers can control for the effect of a set of covariates without reducing the degrees of freedom. For example, the current study attempts to match students' academic performance to analyze the effect of an additional educational service correlated with serval pre-service variables (including demographic variables). Before matching, students who attended supplemental service were much fewer than those who did not, which conflicted with the assumption of repeated measure ANOVA. In addition, the supplemental service group had a higher proportion of Black students than the group without service. Propensity matching solved these conflicts by creating comparable groups and enabling direct comparison of the groups.

The current research indicated that supplemental service did not improve students' motivation and engagement, which suggests that attending supplemental service may not have benefits for these key learning-related variables that could help students be successful in other academic domains as well. Given the academic enablers, motivation and engagement, could improve students' academic achievement (DiPerna & Elliott, 2002), students' motivation and engagement could represent an additional consideration and focus for supplemental services

Limitations and Directions for Future Research

There are three main limitations to the study.

First, the current data did not allow for splitting the group of students who attended supplemental service by subject and category. In the present research sample, the supplemental services reported for each student had various contents (reading, math, unknown), various types of services (Title 1, instructional and response to intervention, tutoring), and various dosage (some students attend more than one supplemental service). The type of the services could influence students' performance differently, and services with worse effects may cover the impact of other services. In this case, more sophisticated examinations need be conducted with larger sample sizes and more specific data regarding service type, focus, and duration.

Second, the current research failed to include the effect of school and teacher. The school selection embodied a couple of demographic variables, and these school-level variables may correlate with students' enrollment in supplemental service. For example, students from lower-income households may tend to enter an elementary school in their neighborhood with a higher proportion of students with similar economic challenges. In addition, schools and teachers take responsibility for education, contributing to students learning progress. The propensity analysis could match students on several school-level demographic variables, including school as a matching covariate. When including teacher and school in the effect analysis phase, the effect of teachers and schools can be isolated from the impact of supplemental service (Wright et al., 1997).

Third, the sample size of the current study was small, which limited the analysis of matching procedure and the repeated measure ANOVA. Ina small sample size every unique student represents a combination of a set of covariates, and other combinations may not be represented. In this case, the matching procedure, based on current sample, might be biased. In addition, the small sample limited the power of the repeated measure ANOVA to detect effects of supplemental service. A larger sample size and collecting greater details on supplemental service received would enable future researchers to more definitively examine the effects of supplemental service on students' academic skills, motivation, and engagement.

References

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education*, *70*(2), 87–107. https://doi.org/10.2307/2673158

Ascher, C. (2006). NCLB's supplemental educational services: Is this what our students need? *Phi Delta Kappan*, *88*(2). https://doi.org/10.1177/003172170608800211

Austin, C. R., Wanzek, J., Scammacca, N. K., Vaughn, S., Gesel, S. A., Donegan, R. E., & Engelmann, M. L. (2018). The relationship between study quality and the effects of supplemental reading interventions: A meta-analysis. *Exceptional Children*, *85*(3), 347–366. https://doi.org/10.1177/0014402918796164

Baker, S. K., Kamata, A., Wright, A., Farmer, D., & Nippert, R. (2019). Using propensity score matching to estimate treatment effects of afterschool programs on third-grade reading outcomes. *Journal of Community Psychology*, *47*(1), 117–134. https://doi.org/10.1002/jcop.22104

Barley, Z. A., & Wegner, S. (2010). An examination of the provision of supplemental educational services in nine rural schools. *Journal of Research in Rural Education*, *25*(5), 1–13. https://jrre.psu.edu/sites/default/files/2019-08/25-5.pdf

Belfi, B., Haelermans, C., & de Fraine, B. (2016). The long-term differential achievement effects of school socioeconomic composition in primary education: A propensity score matching approach. *British Journal of Educational Psychology*, *86*(4), 501–525. https://doi.org/10.1111/bjep.12120

Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology*, *35*(1), 61–79. https://doi.org/10.1016/S0022-4405(96)00029-5

Deke, J., Dragoset, L., Bogen, K., & Gill, B. (2012). *Impacts of Title I supplemental educational services on student achievement*. https://files.eric.ed.gov/fulltext/ED532016.pdf

Deke, J., Gill, B., Dragoset, L., & Bogen, K. (2014). effectiveness of supplemental educational services. *Journal of Research on Educational Effectiveness*, *7*(2), 137–165. https://doi.org/10.1080/19345747.2013.809178

DiPerna, J. C., & Elliott, S. N. (1999). Development and validation of the academic competence evaluation scales. *School Effectiveness and School Improvement*, *17*(3), 207–225.

DiPerna, J. C., & Elliott, S. N. (2000). *Academic Competence Evaluation Scales*. The Psychological Corporation.

DiPerna, J. C., & Elliott, S. N. (2002). Promoting academic enablers to improve student achievement: An introduction to the mini-series. *School Psychology Review*, *31*(3), 293–297. https://doi.org/10.1080%2F02796015.2002.12086156

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2002). A model of academic enablers and elementary reading/language arts achievement. *School Psychology Review*, *31*(3), 298–312. https://doi.org/10.1080/02796015.2002.12086157

Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 52–72). Guilford Press.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *The Elementary School Journal*, *95*(5), 421–434. https://doi.org/10.1086/461853

Goldenberg, C. (2008). Teaching English language learners: What the research does - and does not - say. *ESED 5234 - Master List. 27.* https://digitalcommons.georgiasouthern.edu/esed5234-master/27

Haager, D., & Windmueller, M. P. (2016). Early reading intervention for English language learners at-risk for learning disabilities: Student and teacher outcomes in an urban school. *Learning Disability Quarterly*, *24*(4), 235–250. https://doi.org/10.2307/1511113

Harding, H. R., Harrison-Jones, L., & Rebach, H. M. (2012). A study of the effectiveness of supplemental educational services for Title I students in Baltimore City public schools. *Journal of Negro Education*, *81*(1), 52–66. https://doi.org/10.7709/jnegroeducation.81.1.0052

Heinrich, C. J., Meyer, R. H., & Whitten, G. (2010). Supplemental education services under no child left behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis*, *32*(2), 273–298. https://doi.org/10.3102/0162373710361640

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. https://doi.org/10.1093/PAN/MPL013

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1–28. https://doi.org/10.18637/JSS.V042.I08

Lee, J. S. (2014). The relationship between student engagement and academic performance: Is it a myth or reality? *The Journal of Educational Research*, *107*(3), 177–185. https://doi.org/10.1080/00220671.2013.807491

Muñoz, M. A., Chang, F., & Ross, S. M. (2012). No Child Left Behind and tutoring in reading and mathematics: Impact of supplemental educational services on large scale assessment.

Journal of Education for Students Placed at Risk, *17*(3), 186–200. https://doi.org/10.1080/10824669.2012.688170

Muñoz, M. A., Potter, A. P., & Ross, S. M. (2008). Supplemental educational services as a consequence of the NCLB legislation: Evaluating its impact on student achievement in a large urban district. *Journal of Education for Students Placed at Risk (JESPAR)*, *13*(1), 1–25. https://doi.org/10.1080/10824660701860342

Muñoz, M. A., Ross, S. M., & Neergaard, L. L. (2009). Supplemental educational services as a component of No Child Left Behind: A mixed-method analysis of its impact on student achievement. *Planning and Changing*, *40*(3 & 4), 135–159.

National Center for Education Statistics. (2009). *Mean reading scale scores and specific reading skills of fall 1998 first-time kindergartners, by time of assessment and selected characteristics: Selected years, fall 1998 through spring 2007*. https://nces.ed.gov/programs/digest/d09/tables/dt09_114.asp

National Center for Education Statistics. (2012). *Mean mathematics and science scale scores and specific mathematics skills of fall 1998 first-time kindergartners, by time of assessment and selected characteristics: Selected years, fall 1998 through spring 2007*. https://nces.ed.gov/programs/digest/d12/tables/dt12_139.asp

National Center for Education Statistics. (2016a). *Fall 2010 first-time kindergartners' approaches to learning, interpersonal skills, and self-control scale scores, by selected child, family, and school characteristics: Fall 2010, spring 2011, spring 2012, and spring 2013*. https://nces.ed.gov/programs/digest/d15/tables/dt15_220.45.asp

National Center for Education Statistics. (2016b). *Fast facts: Knowledge and skills at kindergarten*. https://nces.ed.gov/FastFacts/display.asp?id=680

National Center for Education Statistics. (2019). *Fall 2010 first-time kindergartners' reading scale scores and standard deviations through spring of fifth grade, by selected child, family,* *and school characteristics during the kindergarten year: Fall 2010 and spring 2011 through spring 2016*. https://nces.ed.gov/programs/digest/d18/tables/dt18_220.40.asp

National Center for Education Statistics. (2022, July 8). *Fast Facts: Students with disabilities (64)*. https://nces.ed.gov/fastfacts/display.asp?id=64

Pennsylvania Department of Education. (2022). *Title I*. https://www.education.pa.gov/Teachers%20-%20Administrators/Federal%20Programs/TitleI/Pages/default.aspx#:~:text=Program%20Description%3A%20Title%20I%20is,opportunities%20for%20educationally%20deprived%20children.

Renaissance Learning. (2009). *STAR Math technical manual*. Renaissance Learning.

Renaissance Learning. (2010). *STAR Reading technical manual*. Renaissance Learning.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/BIOMET/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, *39*(1), 33–38. https://doi.org/10.1080/00031305.1985.10479383

Ross, S. M., Neergaard, L. L., Harrison, L., Ford, J., & Paek, J. (2009). *Implementation and outcomes of supplemental educational services: The 2007-2008 Tennessee state-wide evaluation study*. https://ncspe.tc.columbia.edu/working-papers/files/OP176.pdf

Ross, S. M., Potter, A., Paek, J., McKay, D., Sanders, W., & Ashton, J. (2008). Implementation and outcomes of supplemental educational services: The Tennessee state-wide evaluation study. *Journal of Education for Students Placed at Risk (JESPAR)*, *13*(1), 26–58. https://doi.org/10.1080/10824660701860391

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, *57*(2), 149. https://doi.org/10.2307/1170234

Schunk, D. H. (2012). *Learning theories : An educational perspective*. Pearson.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*(4), 571–581. https://doi.org/10.1037/0022-0663.85.4.571

Sonnenschein, S., Stapleton, L. M., & Benson, A. (2010). The relation between the type and amount of instruction and growth in children's reading competencies. *American Educational Research Journal*, *47*(2), 358–389. https://doi.org/10.3102/0002831209349215

Springer, M. G., Pepper, M. J., & Ghosh-Dastidar, B. (2014). Supplemental educational services and student test score gains: Evidence from a large, urban school district. *Journal of Education Finance*, *39*(4), 370–403. https://muse.jhu.edu/article/546720

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. https://doi.org/10.1214/09-STS313

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, *46*(1), 90–118. https://doi.org/10.1080/00273171.2011.540475

U.S. Department of Education. (2009). *Supplemental educational services non-regulatory guidance*. https://www2.ed.gov/nclb/choice/help/ses/guidance.html

U.S. Department of Education. (2012, December 21). *Description of supplemental educational services*. https://www2.ed.gov/nclb/choice/help/ses/description.html

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 91–120). Academic Press.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*(1), 57–67. https://doi.org/10.1023/A:1007999204543

Zakrison, T. L., Austin, P. C., & McCredie, V. A. (2018). A systematic review of propensity score methods in the acute care surgery literature: avoiding the pitfalls and proposing a set of reporting guidelines. *European Journal of Trauma and Emergency Surgery*, *44*(3), 385–395. https://doi.org/10.1007/S00068-017-0786-6/TABLES/3