Preparing for High-Stakes Admissions Tests: A Moderation Mediation Analysis

Jed. I. Appelrouth[1], DeWayne Moore[2], Karen M. Zabrucky[1] & Janelle H. Cheung[2]

[1] Georgia State University, USA

[2] Clemson University, USA

Correspondence: Jed. I. Appelrouth, Georgia State University, USA.

Abstract

For decades researchers have examined the effects of SAT preparation, but only recently have they begun to explore the factors that inform successful test preparation (Appelrouth & Zabrucky, 2717). In their regression analysis of the factors of successful SAT preparation, Appelrouth, Moore, & Zabrucky (2015) found significant effects of homework completion, instructional hours, practice and official testing, distribution of study, and timing of test preparation. The current study builds upon that research in constructing a functional model of SAT preparatory factors. It was hypothesized that direct and indirect relationships would exist between preparatory factors, and that some of these relationships would be moderated by student characteristics such as gender and socioeconomic status. Archival data from 1,933 students were analyzed, and significant direct relations were reported between tutoring start time and the following variables: session distribution, individual tutoring hours, group tutoring hours, homework completion, number of official tests, number of practice tests and total SAT increase. Commencing test preparation earlier yielded positive direct and indirect effects, and session distribution, individual and group tutoring hours, and official SAT and practice SAT tests all mediated the relationship between start time and SAT score increase. School type and socioeconomic status moderated the relationship between start time and individual tutoring hours, and school type also moderated the relationship between homework completion and score increase. The results of this analysis have implications for the thousands of high schools and educational entities that offer SAT coaching programs. By encouraging earlier program start times, adequate instructional hours, distribution of sessions and practice effects, administrators can create more effective SAT preparation programs to serve their college-bound students.

Keywords: college admissions, SAT, SAT coaching, high-stakes tests, college entrance examinations, test preparation, mediation, moderation

1. Introduction

For more than half a century the SAT has been one of the most influential tests in the American educational landscape, capturing the interest of journalists, academic researchers, parents and students. Our focus on the SAT is a reflection of the significance afforded this test by college admissions offices, with over 88% of four-year colleges and universities placing moderate or considerable importance on SAT and ACT test scores (National Association for College Admission Counseling, 2015). As application numbers continue to surge at colleges and universities across the country (Hoover, 2016), admissions officers rely upon admission tests scores as tools to manage the applicant load. While relatively small changes in SAT scores, in the range of 20-30 points, can significantly improve the likelihood a student will be given an offer of admission (Briggs, 2009), larger gains, in the range of 80-100 points, can influence "merit money" and determine whether a student receives the highest possible merit-based scholarship or no scholarship at all (Arenson, 2006; Lilly & Montgomery, 2011).

With such weight placed on SAT outcomes, it is not surprising that families, schools and community organizations go to great lengths to ensure that their students are adequately prepared for the SAT. Many of the 1,637,589 students who took the SAT in the class of 2016 (College Board, 2016) engaged in formal test preparation offered in their high schools, through community organizations, or through commercial test preparation companies (Lilly & Montgomery, 2011). Test preparation programs vary greatly in format, instructional method and efficacy. Some have drawn

criticism for their lack of efficacy, their focus on "relentless drilling" and their lack of research-based methodology (Xie, 2013). There is a need to better understanding how to craft effective and efficient test-preparation programs.

Researchers interested in the SAT have historically focused their inquiries on questions of fairness, accuracy and "coachability." Numerous meta-analyses have confirmed the efficacy of SAT preparation (e.g., Briggs, 2005; Lilly & Montgomery, 2011), however, there is a substantial debate regarding the magnitude of coaching effects. Powers & Rock (1999) found that students who received formal SAT preparation were more likely than un-coached peers to attain large score increases. Lilly & Montgomery (2011) conducted a meta-analysis of 14 randomized studies of SAT preparation effectiveness, and found that students who received formal test preparation achieved SAT gains of 56.2 points over control groups. The question of *whether* SAT preparation is effective has been supplanted by the questions of *why* SAT preparation is effective and *how* we can make it more efficient and useful for students.

Lilly & Montgomery (2011) found that *duration* of coaching was related to the efficacy of SAT preparation programs, whereby math scores were significantly increased in programs with more than eight hours of instruction. Rubenstein (2003) emphasized the value of accurately calibrated and timed practice tests to allow students to reinforce and assimilate strategies, foster mental endurance, and better acclimate to official testing conditions.

In their investigation of factors related to SAT performance, Appelrouth, Zabrucky & Moore (2015) found significant effects of homework completion, instructional hours, practice and official testing, distribution of study, and timing of test preparation. When students began test preparation earlier in their junior year, completed more official and practice SATs, completed more individual and group hours of instruction, and completed a greater percentage of assigned homework, they scored higher on the SAT (Appelrouth et al., 2015). Although this study provided insight into the isolated factors related to successful SAT preparation, it did not establish a model of the most salient factors contributing to successful test preparation and how they directly and indirectly affect performance.

2. Factors Which May Impact Test Preparation Outcomes

In the current investigation we extend the research of Appelrouth et al., (2015), by examining direct and indirect influences as well as conditional direct and indirect influences on SAT scores. This model included the following variables:

2.1 Socioeconomic Status

Critics of the SAT have frequently suggested that the correlation between wealth and SAT scores ($r = .42$; Sackett, Kuncel, Arneson, Cooper, & Waters, 2009) is a challenge to the construct validity of the SAT, potentially reducing the SAT to a mere "wealth test" (Zwick, 2002). Affluent families use superior resources to secure educational advantages for their children through after-school tutoring and test preparation activities deemed "shadow education" (Buchmann et al., 2010). In this analysis we explored whether socioeconomic status (SES) acts as a moderator, resulting in conditional direct and indirect influences on higher test scores.

2.2 School Type

The culture of a high school may affect the manner in which its students approach SAT preparation. Not only do different types of schools allocate different levels of resources towards preparing students for college admission tests, but individual school cultures may influence the preparatory behaviors of families and students. Researchers have investigated the "private school effect," in which students attending private schools have purported academic advantages (Lubienski, Lubienski & Crane, 2008). In this analysis we investigated whether public or private school attendance has a moderating influence on direct and indirect contributions to higher SAT scores.

2.3 Gender

Despite their superior academic performance in high school and college, more ambitious course-taking and superior self-discipline (Duckworth & Seligman, 2006), females consistently underperform males on the SAT (Hannon, 2012; Nankervis, 2011). The College Board (2016) reported that for the class of 2016, males had an average three-section composite SAT score of 1494 (495 critical reading, 524 math, 475 writing) compared to 1474 for females (493 critical reading, 494 math, 487 writing). We explored whether gender exerts a moderating influence on the direct and indirect effects conducive to SAT preparation.

2.4 Timing of Test Preparation

Appropriate timing appears to be essential to successful test preparation (Devine-Eller, 2012; Turner, 2009). Students who extensively delay their preparation may limit the number of official tests available to them in advance of college application deadlines and compel students to "cram" their preparation into a shortened period of time (Turner, 2009). Devine-Eller (2012) found that students commencing test preparation as seniors may be starting too late and "the

savviest and most academically competitive [students] will have completed most of their prep in 11th grade (p. 475)." As judicious timing of SAT preparation plays a meaningful role in determining score gains we investigated the direct and indirect relationships between timing of SAT preparation and all other variables in the model.

2.5 Distributed Study

Researchers have consistently found benefits of distributing study over time (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Learning schedules with the greatest and most even distribution of lessons result in higher rates of transfer (Vlach & Sandhofer, 2012) as well as higher rates of encoding material into long-term memory (Rawson, Dunlosky, & Sciartelli, 2013). Combining the use of practice tests and spacing study over time leads to "successive relearning" which enhances student performance (Rawson et. al, 2013, p. 523). Positive effects of distributed study have been demonstrated in the domain of private tutoring (Elbaum, Vaugh, Hughes, & Moody, 2000) and in SAT preparation (Appelrouth et. al, 2015). In this analysis we further investigated the direct and indirect relations of distributed study to other variables influencing score gains on the SAT.

2.6 Time on Task

As researchers have found a relationship between time on task and academic gains (e.g., Andersen, Humlum, & Nandrup, 2016), SAT researchers have likewise noted the positive effect of instructional time on SAT score increases (Lilly & Montgomery, 2011). In this analysis we investigated the direct and indirect relationships between increased instructional time and other variables in the model.

2.7 Tutoring Format

Although many SAT researchers have failed to distinguish between one-on-one tutoring and group instruction, individualized instruction is related to larger academic gains in numerous domains (Bloom, 1984; Ireson, 2004). Tutoring is likely to be of higher quality and efficacy in a small group or individual format due to the "feedback-corrective process of mastery learning" (Bloom, 1984, p.7) and the individual tutor's ability to tailor instruction specifically to the needs of a particular student (Ireson, 2004). In their investigation, Applerouth et al., (2015) found that each hour of individual tutoring had a larger effect than a corresponding hour of group tutoring. We examined the direct and indirect relations between group and individual instruction and SAT preparation.

2.8 Homework

A positive correlation has been reported between time spent on homework and academic achievement (Cooper, Robinson, & Patall, 2006) as well as between completion of homework assignments and academic achievement (Cooper, Lindsay, Nye, & Greathouse, 1998). Appelrouth, et al. (2015) found that homework completion during test preparation was correlated positively with SAT score increases, and in this investigation we further examined the direct and indirect relations of homework to other SAT preparation variables.

2.9 Practice Tests

Researchers have found a powerful effect of practice tests on memory, learning, and academic performance (Arnold & McDermott, 2013; Roediger & Butler, 2011). Retrieval practice during the act of testing strengthens the memory of learned content for future retrieval much more than repeated restudy, and more practice tests yield greater performance gains (Arnold & McDermott, 2013, Roediger & Butler, 2011). Practice effects have been found to enhance performance for up to seven rounds of timed practice trials on the SAT. In their meta-analysis of 40 SAT-based studies, Kulik, Kulik, and colleagues found that a student would gain 20 points on an official SAT for one simulated SAT trial, and postulated that gains of 40-points on official SATs, resulting from four to six hours of practice trials—the time needed to administer two full practice SATs—would not be unusual (Kulik et al., 1984, p. 444). In this analysis we investigated the direct and indirect relations between practice SATs and other variables of SAT preparation.

2.10 Number of Official Tests

Score gains resulting from taking multiple practice SATs also occur from taking multiple official SATs (Vigdor & Clotfelter, 2003). Patterson, Mattern and Swerdzewski (2012) found that of 92,634 students who took the SAT multiple times, 64 percent of students achieved their highest single administration score on their last or final SAT. In this analysis we investigated the direct and indirect relations between official SATs and other variables of SAT preparation.

Aims of the Current Investigation

The primary aim of the current study is to expand upon the findings of Appelrouth, et al. (2015) by exploring the direct, indirect and conditional relationships among the factors that inform successful SAT preparation. By using this

approach, the salient factors of test preparation are incorporated into a single, functional model to demonstrate the mechanisms by which SAT preparation factors influence SAT gains.

The model hypothesizes a set of relations among those variables that result in higher SAT scores. Mediating and moderating effects are examined within the model. Mediation, an indirect effect, is the process by which an intermediary variable transmits the effect of a treatment or intervention from an independent variable to a dependent variable (Preacher, Rucker, & Hayes, 2007). Moderation, a conditional effect, is the process whereby a variable affects the direction or strength of the relation between an independent and dependent variable.

A path-analytic model is used to describe a series of mediating relationships between tutoring start time and SAT score increase. The path diagram illustrates the predicted relationships in the model, the manner in which potential mediators influence outcomes on the SAT. The conditional nature of relations within the model is addressed by examining selected moderator variables.

This study is the first of its kind to employ this statistical method in the field of high-stakes testing, adding to the growing body of research that employs mediation and moderated-mediation analyses to understand the mechanisms by which antecedent academic success factors influence outcomes (e.g., Arnold & McDermott, 2013; Marjoribanks, 2003). Selected variables in this analysis include time on task, practice and official tests, homework, and distribution of study.

Through this model, the following research questions are addressed: how is tutoring start time related to the number of instructional hours and the distribution of sessions? How is the number of group or individual tutoring hours related to practice and official tests? Is the distribution of sessions related to levels of homework completion?

Description of proposed direct and indirect influences on SAT score increases

To understand the relationships between the various factors involved in SAT preparation, we propose a theoretical model (see Figure 1) illustrating hypothetical direct and indirect relations.
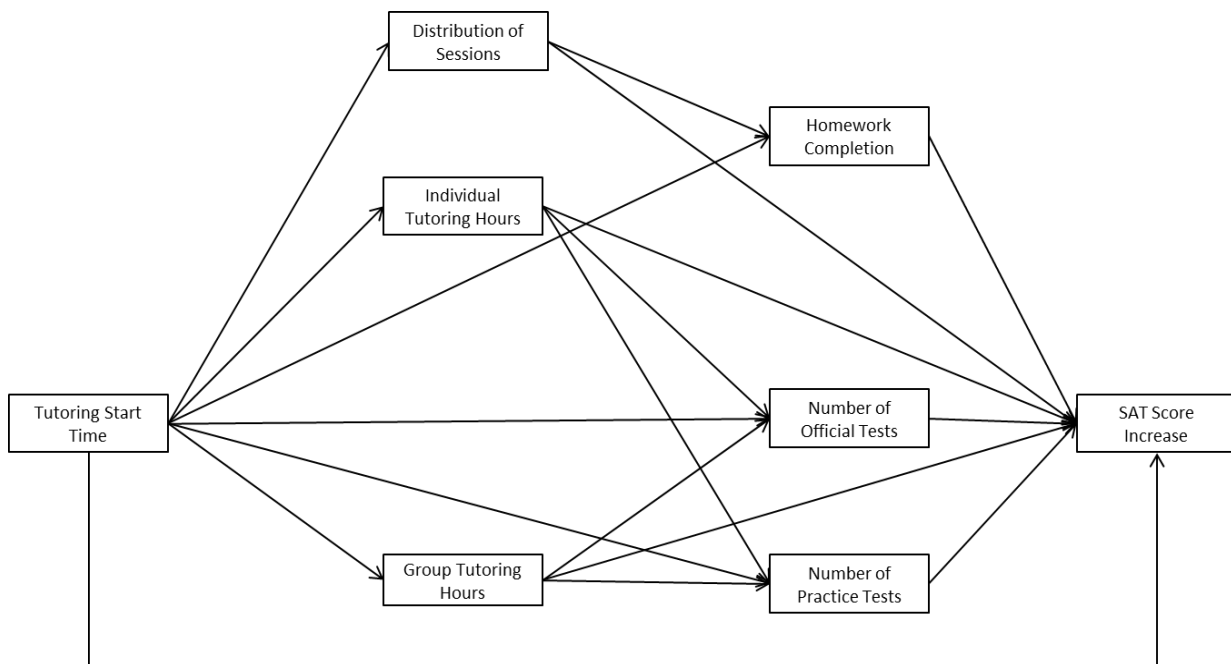

Figure 1. Theoretical model of direct and indirect effects

The initial, or exogenous, variable in our model is tutoring start time. From field observations and findings from other researchers (Devine-Eller, 2012), we inferred that *when* a student commences SAT preparation bears heavily upon direct and indirect influences, affecting multiple preparatory factors. From tutoring start time, the model moves sequentially to endogenous variables: distribution of sessions, individual tutoring hours and group tutoring hours, rate of homework completion and number of practice and official tests. We hypothesize that students who start

preparation later in their academic careers will be forced to condense their instructional sessions into shorter intervals; experience a net reduction in instructional time, practice and official tests; and lead to lower rates of homework completion. The model also examines hypothetical indirect effects, exploring whether more closely spaced tutoring sessions lead to decreases in homework completion, and consequently lower SAT scores. Similarly the model examines whether decreases in instructional time will lead to decreases in the number of practice and official tests, leading to lower SAT scores.

In addition to analyzing the hypothesized mediation pathways, conditional effects are investigated to determine which moderators influence direct paths, resulting in conditional direct and indirect influences. Will working mechanisms of the tutoring intervention differ as a function of specific student characteristics such as gender, SES, or school type? Will mediation only occur at certain values of the moderator variable? As an example, will the indirect effect of tutoring hours on practice tests be stronger in more affluent (higher SES) students and weaker in less affluent (lower SES) students?Direct and indirect influences and conditional effects are examined to help clarify the relationships between the various factors of SAT preparation.

3. Methods

3.1 Participants

Participants (from an archival database) were high school students whose families had secured the services of a test preparation center to help them prepare for the SAT. Between January 1, 2006, and August 1, 2013, demographic, attendance, participation, and SAT scores were routinely collected on all students who enrolled for test preparation classes. Participants lived in one of three major metropolitan areas where the tutoring center offered services: Atlanta, GA; Washington, DC; and New York, NY.

Participants were in their junior or senior year of high school and attended one of 168 private (n = 71) or public (n = 97) schools. Of the sample of 1,933 students, 1,246 attended private school and 687 attended public school. There were 1,014 females and 919 males in this group.These 1,933 students were selected because they fulfilled three criteria. All students had "baseline" SAT scores prior to beginning any preparation activities with the center: students either reported officially administered SAT scores or were administered a baseline SAT at the center. All students were engaged in individual or group SAT preparation with the center, but did *not* prepare for the ACT exam. Finally, following test preparation, students reported "post" intervention scores from officially administered SATs.

3.2 Measures

SAT baseline type: Baseline score data were captured in one of two ways. Of the 1,933 students in this study, 803 (41.54%) reported official SAT scores prior to initiating preparation activities with the center and 1,133 (58.46%) who had not previously taken the official SAT, were administered "unofficial" tests (College Board released SATs), at the center before the start of preparatory activities. The unofficial tests were administered under controlled conditions, using timing, breaks, and procedures closely approximating official testing conditions. One clear distinction between the two types of administrations is that the unofficial SATs lack the "experimental" section found on official SATs, a section which generally adds 25 minutes to the length of the test. Therefore, the unofficial SATs administered by the center were shorter in duration than the official tests.For the analysis, baseline type was coded as a dichotomous variable (0 = official, 1 = unofficial).

School Type: High school attendance data were collected from students or their families in an initial client intake procedure. Each high school was referenced against online databases and websites to determine if students attended public or private schools. For the statistical analysis, School Type was coded as a dichotomous variable (0 = private, 1 = public).

Gender: Gender data were collected from students or their families during the initial client intake procedure. For the statistical analysis, gender was coded as a dichotomous variable (0 = boy, 1 = girl).

Socioeconomic status: We used current property values of students' homes as a proxy for SES. During the client intake procedure, center employees recorded and inputted home addresses of students into the center database. Addresses were referenced against a national database of home values, www.Zillow.com, to determine the present value of these homes as of September 15, 2013. Families living in rented apartments were excluded from our analyses. In order to remove statistical effects of extreme outliers in this data set, we employed a natural logarithmic transformation and used the transformed values in all of our statistical analyses. For the statistical analysis on SES status, home property values were recoded as Socioeconomic Status (SES) using three variables (0 = low, 1 = middle, 2 = high).

Tutoring start time: We were interested in the effects of starting tutoring early in the junior year compared to waiting until later in the junior year or during the senior year. To establish a numeric value for starting early or late, we set the month of June at the conclusion of the junior year of high school as the base value. A student beginning test preparation in June of their junior year was assigned a value of 0. For students starting in any other month, we counted backward or forward, by months, from June. If a student began prepping in October of his/her junior year, that student received a value of -8 to signify that the student started preparing for the SAT eight months before June. If a student began prepping in July before their senior year, that student received a value of +1.

Distribution of sessions: To calculate the average amount of time that elapsed between tutoring sessions, we established a global measure. We divided the number of days that transpired between the initial and the final tutoring sessions, either group or private, by the total number of sessions. This value provided the average amount of time that elapsed between sessions and provided a measure of the distribution of study. Due to extreme positive skew and kurtosis in this data set, we transformed this data using a natural logarithmic transformation and used the transformed values in all analyses.

Individual and group tutoring hours: The number of hours of individual, one-on-one tutoring, measured to the quarter hour, was recorded from session notes maintained by the center as were the number of hours of group tutoring, involving groups with fewer than 20 students.

Homework completion: Tutors recorded homework completion data at the beginning of every group or individual tutoring session using a 100-point scale at 5-point increments. A score of 100 indicated that all assigned homework was completed. A score of 0 indicated that no homework assigned was completed. Tutors estimated the value for partial completion. The average homework completion rate for all sessions was calculated for each student.

Number of practice and official tests: Students enrolled in either group or individual sessions were strongly encouraged to take mid-term practice tests, which replicated the conditions of the baseline tests. The center administered practice tests released by the College Board to better prepare students for official tests. The practice test count is the number of interim, nonofficial practice tests taken by each student. We also recorded the number of official SAT tests reported by the student or his/her family.

SAT score increase: A common practice of university admissions departments is to "superscore" the SAT (Patterson et. al, 2012). This involves adding the highest section scores (critical reading, math and writing) from various SAT administrations to create a single composite score. We adopted the practice of super-scoring to calculate student score increases. If a student took a baseline assessment with the center, this score was used as the baseline score. If a student had already taken a single official SAT prior to beginning preparation, this score was used as the baseline. If a student had taken multiple SAT administrations before beginning tutoring, we used the pre-preparation "super-score" as the baseline. To determine the net gains from tutoring, we subtracted the baseline score from the final super-scored SAT, calculated from the highest section scores achieved on all official SATs.

The data used in the current investigation were the same data collected by Appelrouth, et al. (2015), the archival data routinely collected by a private tutoring center between January 1, 2006, and August 1, 2013, and maintained within the tutoring center's internal database. We originally exported the center's data, assigning each student a unique numeric identifier and removing all personal information. Nominal variables were dummy-coded. We used SPSS to screen the data for (i) outliers (standardized residuals and leverage values), and (ii) multicollinearity (tolerance). Preliminary analyses, which included all of the measures as predictors, revealed fairly normal data that did not have excessive multicollinearity (tolerance values >.20). Twelve data points were found to be outliers, with standardized residuals greater than 3.0 or less than -3.0, or leverage values greater than .03, and we removed them from the data set before conducting our analyses.

Descriptive statistics and bivariate correlations. In the current investigation, means, standard deviations, ranges, skew and kurtosis values were calculated for all variables as were the means and standard deviations for the score increases of discrete groups within the sample. Bivariate correlations were conducted to determine the relationships between study variables.

Analyses of direct and indirect effects. All statistical analyses were conducted using Mplus v7.3. Results are considered to be significant if the p-value in a given analysis is equal to or lower than 0.05. Several indirect, mediated, influences on SAT increase are depicted in the model. Given our large sample size, we used normal theory tests available in Mplus to test the indirect effects.

Conditional effects. Conditional (i.e., moderated) effects were examined for selected variables—gender, socioeconomic status, school type, and baseline type— within the model. In examining the potential for moderated

effects, we examined whether the direct effect of a predictor on an outcome was contingent upon the values of the moderator. To calculate conditional indirect effects, in which indirect, or mediated effects, were tested at different values of the moderator, we used a normal theory approach available in Mplus. The model with proposed moderators is presented in Figure 2.
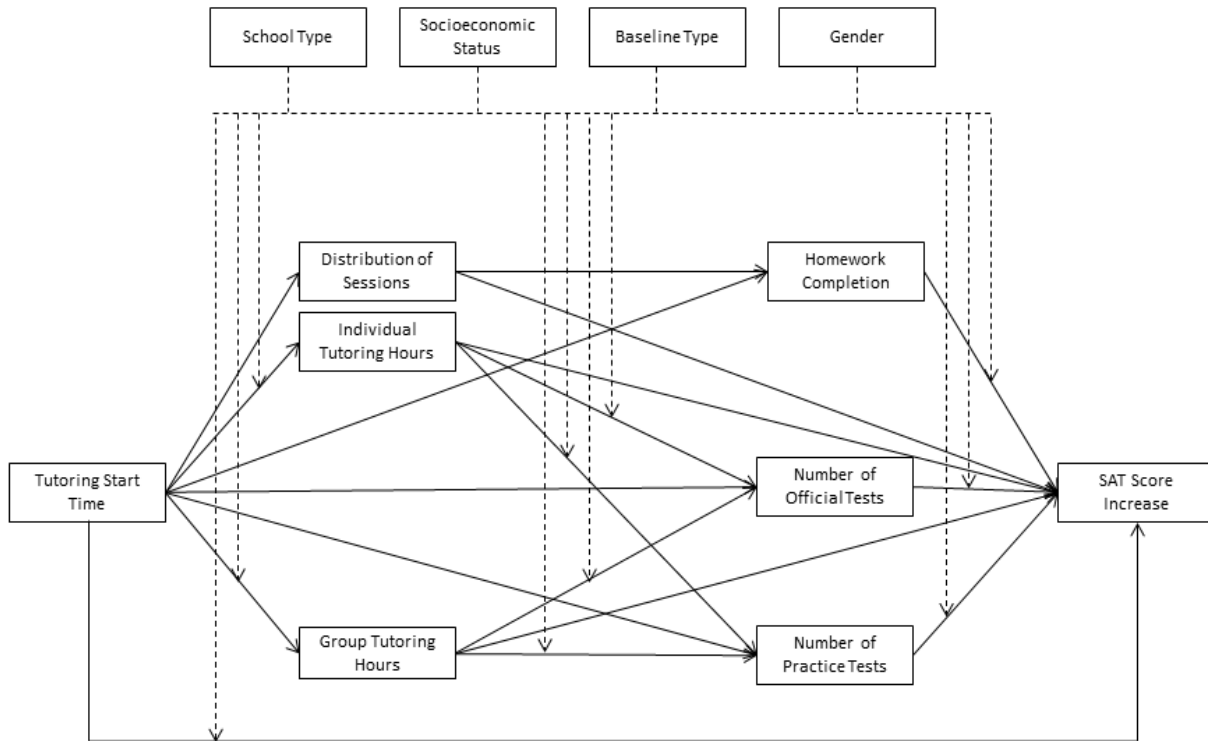

Figure 2. Theoretical model of direct, indirect and conditional effects

3.3 Descriptive Statistics

Table 1 shows the means, standard deviations, ranges, skew and kurtosis for each of the measures. As Table 1 reveals, there is good variability for most of the dependent variables as evidenced by the moderate to high standard deviations. The values for skew and kurtosis suggest that the distributions for the majority of variables do not deviate severely from normal, staying within the range of ± 3. Two variables with the strongest kurtosis, home property value and distributed sessions were adjusted using natural logarithmic transformations, thereby removing the excessive skewness in the distributions.

Table 1. Means, standard deviations, ranges, skew and kurtosis for all variables

	M	SD	Range	Skew	Kurtosis
Home property value	$749,996	$674,299	$45,529 - $15,101,670	7.64	121.45
Home property value (adjusted)	13.32	0.62	10.73 - 16.53	.18	1.51
Individual tutoring hours	11.85	9.31	0.00 - 50.50	.84	0.73
Group tutoring hours	7.89	10.77	0.00 - 67.50	1.09	0.72
Number of practice tests	2.33	1.52	0 - 10	.91	1.85
Number of official tests	2.01	0.85	1 - 6	.71	0.48

SAT tutoring start time	-4.03	3.95	-11 - 10	.23	-0.57
Distribution of sessions	14.21	12.80	0 - 175.28	4.23	31.72
Distribution of sessions (adjusted)	2.50	0.65	0 - 5.17	0.18	0.97
Homework completion	79.30	22.17	0.00 - 100	-1.60	2.75
SAT score increase	200.13	111.08	0 - 620	.42	-0.17

Table 2 shows the means and standard deviations for the SAT increases achieved by groups of students categorized by gender, school type, and baseline type. The data revealed that students taking unofficial baseline tests achieved score increases 70.57 points higher than students who took official baseline tests. Private school students attained score increases that were 22.66 points above public school students. Females attained score increases 3.29 points above males.

Table 2. Means and standard deviations for SAT score increases

Gender	Baseline Type	School type		Total
		Private *M (SD)*	Public *M (SD)*	*M (SD)*
Female	Official	167.13 (109.42) *n* = 232	162.91 (99.15) *n* = 175	165.32 (105.02) *n* = 407
	Unofficial	238.86 (111.50) *n* = 404	200.64 (91.00) *n* = 203	226.08 (106.55) *n* = 607
	Total	212.70 (115.93) *n* = 636	183.17 (96.58) *n* = 378	201.69 (110.00) *n* = 1014
Male	Official	155.46 (103.03) *n* = 238	147.41 (95.32) *n* = 158	152.25 (99.98) *n* = 396
	Unofficial	234.19 (108.36) *n* = 372	231.26 (109.15) *n* = 151	233.35 (108.49) *n* = 523
	Total	203.48 (112.96) *n* = 610	188.38 (110.44) *n* = 309	198.40 (112.29) *n* = 919
Total	Official	161.22 (106.28) *n* = 470	155.56 (97.51) *n* = 333	158.87 (102.71) *n* = 803
	Unofficial	236.62 (109.96) *n* = 776	213.70 (100.15) *n* = 354	229.44 (107.47) *n* = 1130
	Total	208.18 (114.53) *n* = 1246	185.52 (103.00) *n* = 687	200.13 (111.08) *n* = 1933

3.4 Direct Relationships Within the Model

We examined the relationships between different variables within the model, looking first at the direct relationships, which are found in Table 3. The model had good fit based on standard fit indices (chi-square= 102,29 (16), CFI = .97, RMSEA = .053 (.043-.063), SRMR = .025). Between the independent variable, start time, and the dependent variable, SAT score increase, multiple direct and indirect relations were found. Significant direct relations were reported

between tutoring start time and session distribution, individual and group tutoring hours, homework completion, number of official and practice tests and total SAT increase. All direct effects in the model are depicted in Figure 3.

Table 3. Analysis of direct effects within the model

| Path	B	SE
Start time to distribution of sessions	-0.06**	<.01
Start time to individual tutoring hours	-0.39**	0.05
Start time to group tutoring hours	-0.59**	0.06
Start time to homework completion	-0.42**	0.14
Start time to number of official tests	0.05**	0.01
Start time to number of practice tests	-0.06**	0.01
Start time to SAT score increase	-7.09**	0.67
Distribution of sessions to homework completion	-0.45	0.84
Distribution of sessions to SAT score increase	16.71**	3.72
Individual tutoring hours to number of official tests	0.03**	<.01
Individual tutoring hours to number of practice tests	0.07**	<.01
Individual tutoring hours to SAT score increase	3.10**	0.33
Group tutoring hours to number of official tests	0.01**	<.01
Group tutoring hours to number of practice tests	0.07**	<.01
Group tutoring hours to SAT score increase	1.94**	0.28
Homework completion to SAT score increase	0.13	0.10
Number of official tests to SAT score increase	13.72**	2.79
Number of practice tests to SAT score increase	6.88**	1.69

**$p < .01$
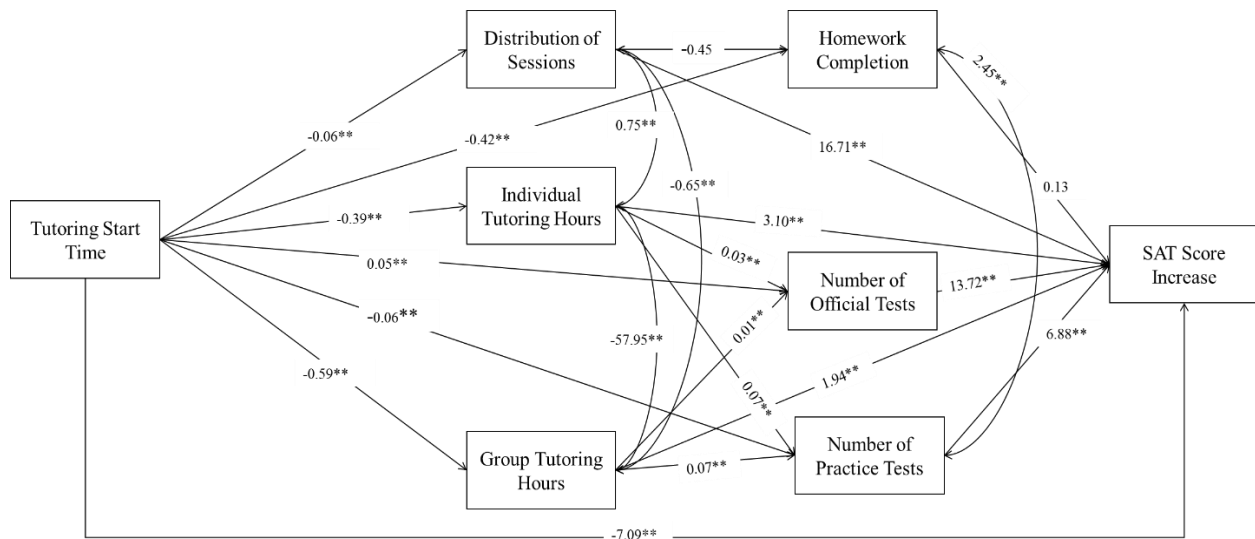
Figure 3. Model with direct effects and covariances estimated (unstandardized regression coefficients)

Note: Covariances were included based on modification indices.

Starting tutoring earlier junior year yielded many positive effects. For every month prior to June of Junior year that students began SAT preparation, students had tutoring sessions that were more distributed by .06 days; completed .39 additional hours of individual tutoring, .59 additional hours of group tutoring and .42% more of their assigned homework; took .06 more practice tests and increased their SAT scores by 7.09 points. One unexpected finding was that starting preparation earlier junior year had a slightly negative direct effect upon the number of official tests a student took, although the indirect effects were positive.

Having a greater distribution of tutoring sessions resulted in greater score increases. Likewise, an increase in individual tutoring hours, group tutoring hours, practice tests and official tests all led directly to increased SAT scores. Each additional official test yielded a 13.72 point SAT increase, compared to gains of 6.88 for a practice test, 3.10 for an individual tutoring hour, and 1.94 points for a group tutoring hour. Additional group and individual tutoring hours predicted an increase in both practice and official tests.

3.5 Indirect Relationships Within the Model

Multiple indirect effects between start time and score increase are depicted in the model (see Figure 3). These indirect effects identify the influence that start time exerts on score increase through the mediators. All tests of these indirect effects are shown in Table 4. The model contains six instances of 2-path mediation (see Table 4) in which a single variable mediates the relationship between start time and SAT score increase; five of these are statistically significant. Session distribution mediated the relationship between start time and SAT score increase in that score increases attributed to earlier start times are partially due to the increase in session distribution which resulted from the earlier start times. The number of tutoring hours also mediated the relationship between start time and score increase with indirect effects through individual tutoring (1.21) and group tutoring (1.14). Number of tests mediated the relationship between start time and score increase with positive indirect effects through practice tests (0.44), but negative indirect effects through official tests (-0.69).

Table 4. Analysis of indirect effects within the model

Two-path mediation

Variable Sequence	B	SE
Start time (A) to Distribution of sessions (B) to SAT score increase (C)	-0.99**	0.23
Start time (A) to Homework completion (B) to SAT score increase (C)	-0.05	0.05
Start time (A) to Individual tutoring hours (B) to SAT score increase (C)	-1.21**	0.21
Start time (A) to Group tutoring hours (B) to SAT score increase (C)	-1.14**	0.20
Start time (A) to Number of official tests (B) to SAT score increase (C)	0.69**	0.16
Start time (A) to Number of practice tests (B) to SAT score increase (C)	-0.44**	0.12

Three-path mediation

Variable Sequence	B	SE
Start time (A) to Distribution of sessions (B) to Homework completion (C) to SAT score increase (D)	0.003	0.01
Start time (A) to Individual tutoring hours (B) Number of official tests (C) to SAT score increase (D)	-0.18**	0.05
Start time (A) to Individual tutoring hours (B) Number of practice tests (C) to SAT score increase (D)	-0.19**	0.05
Start time (A) to Group tutoring hours (B) Number of official tests (C) to SAT score increase (D)	-0.05*	0.02
Start time (A) to Group tutoring hours (B) Number of practice tests (C) to SAT score increase (D)	-0.27**	0.07

* $p < .05$ **$p < .01$

3.6 3-Path Mediation

We examined five instances of 3-path mediation within the model (see Table 4), four of which reveal significant mediation of the effects of tutoring start time on score increase. Increases in SAT scores were associated with earlier start times due to the indirect effects through individual tutoring and official SATs (.18), through individual tutoring and practice SATs (.19), through group tutoring and official SATs (.05), and through group tutoring and practice SATs (.27).

3.7 Conditional Direct Relationships Within the Model

Interaction tests were conducted to examine the moderating role of socioeconomic status, school type, baseline type and gender in the model. Table 5 shows the results of the tests of the interactions for statistically significant moderators, and the conditional coefficients. Figure 4 displays the conditional effects in the context of the model. Our analysis revealed that gender was not a moderating factor for any relations depicted in the model. School type was a significant moderator of the relationship between start time and individual tutoring hours. The coefficients from Table 5 indicate that compared to private school students ($B = -0.53$, $p < .01$), public school students show a smaller effect ($B = -.16$, $p < .05$) of start time on individual tutoring hours. When private school students start tutoring one month earlier in their junior year, this translated to a larger gain in the number of individual tutoring hours (.53 hours) than it did for their public school counterparts (.16 hours). The interaction coefficient ($B = .37$, $p < .01$) is the difference between the two values (.i.e., 0.53 - 0.16 = .37), and the significance test revealed that these values were significantly different.

Table 5. Analysis of conditional direct effects within the model

| Path	Interaction		Conditional Estimates		
	B	*SE*	Category	*B*	*SE*
School type moderating the path between start time and individual hours	0.37**	0.09	Private	-0.53**	0.06
			Public	-0.16*	0.08
School type moderating the path between start time and group hours	0.39**	0.10	Private	-0.73**	0.07
			Public	-0.34**	0.09
School type moderating the path between homework completion and score increase	0.43*	0.21	Private	0.01	0.12
			Public	0.43*	0.17
Socioeconomic status moderating the path between start time and individual hours	-0.52**	0.10	Low	-0.24**	0.06
			High	-0.76**	0.09
Socioeconomic status moderating the path between start time and group hours	-0.27	0.14	Low	-0.55**	0.07
			High	-0.82**	0.12
Baseline type moderating the path between official tests and score increase	16.95**	5.92	Official	20.30**	4.76
			Unofficial	37.25**	3.99

4* $p < .05$ **$p < .01$
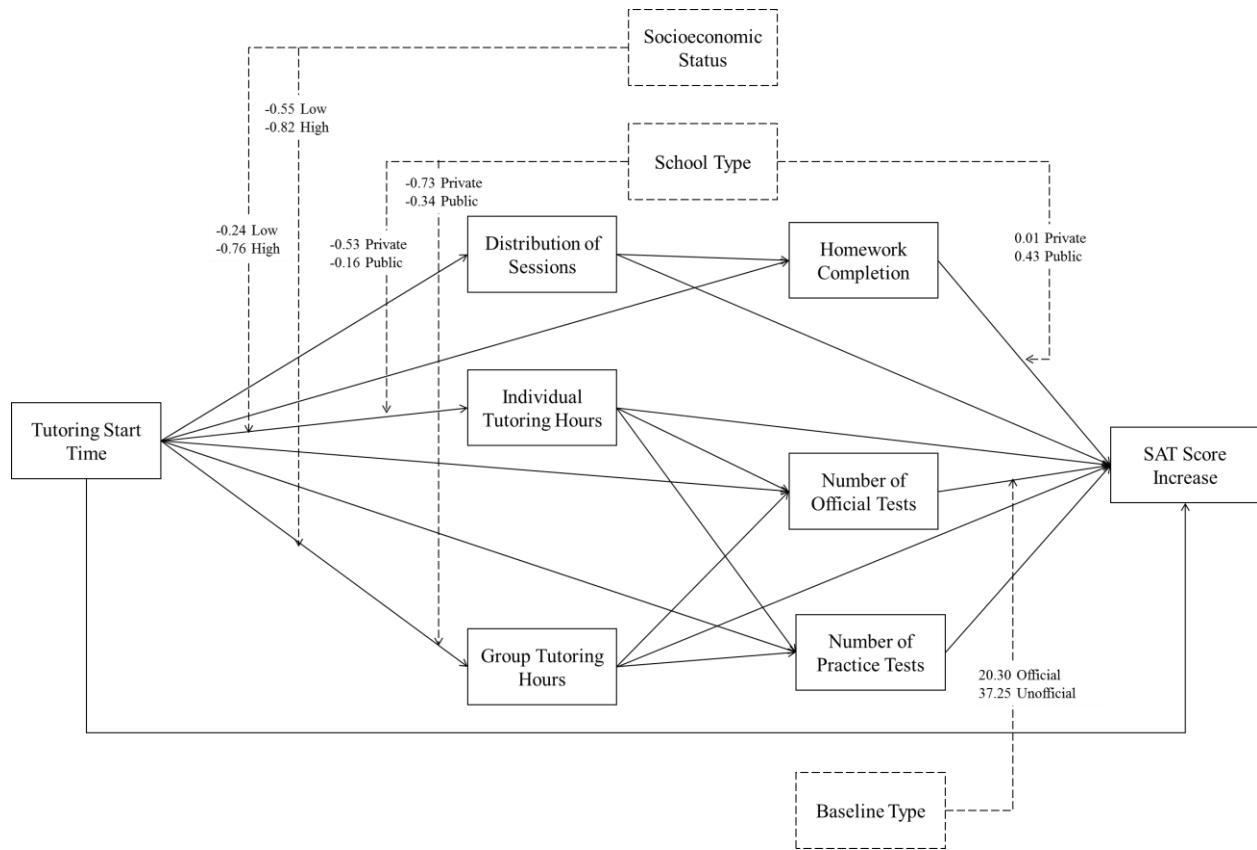
Figure 4. Model with conditional values

School type exerted a significant moderating effect on the relationship between tutoring start time and the number of group tutoring hours. Starting tutoring a month earlier in the school year led to an increase of .73 group hours for private school students ($B = -0.73$, $p < .01$) compared to an increase of only .34 hours for students in public school ($B = -0.34$, $p < .01$). School type also exerted a significant moderating effect upon the relationship between homework completion and score increase. Homework completion had a much greater effect on SAT score increase for public school students ($B = 0.43$, $p < .05$) than for students in private school ($B = 0.01$, ***p > .05***), whereby the benefit of completing 100% of assigned homework would be 42 points higher for students in public school.

Socioeconomic status (SES) moderated the relationship between start time and individual and group tutoring hours. Initial tests indicated that the moderating effects of SES were not significant between low and medium SES but were significant between low and high and medium and high SES. Therefore, the results of the moderating effects of SES are presented for low versus high SES only. Earlier start times had a stronger effect on the number of individual tutoring hours for high SES students ($B = -0.76$, p $< .01$) than for low SES ($B = -0.24$, p $< .01$) students.Similarly, earlier start times had a stronger effect on the number of group tutoring hours for high SES students ($B = -.82$, p $< .01$) than for low ($B = -0.55$, p $< .01$) SES students.

Baseline type was the final significant moderator in the model, moderating the path between the number of official tests and score increases. Students who began preparing for the SAT by taking an unofficial practice test gained 37.25 points ($B = 37.25$, $p < .01$) for each official SAT they took. In contrast, students who had already completed an official SAT in advance of commencing test preparation gained only 20.30 points for each official SAT ($B = 20.30$, $p < .01$).

3.8 Conditional Indirect Relationships Within the Model

Table 6 presents the conditional 2-path indirect effects within the model. Where Table 5 presented the conditional direct effects, Table 6 integrates these effects and reports the conditional indirect effects. The mediating effect of individual hours between start time and SAT increase was moderated by school type. Private school students who started SAT preparation a month earlier in the junior year gained 1.64 points ($p < .01$) through increased individual

tutoring, compared to public school students who only gained .50 points ($p < .05$). Similarly, private school students who started SAT preparation a month earlier gained 1.41 points ($p < .01$) through increased group tutoring, compared to public school students who gained .65 points ($p < .01$).

Table 6. Analysis of conditional 2-path indirect effects within the model

Conditional indirect relations	Category	B	SE
Moderator: School Type			
Start time to individual hours to SAT increase	Private	-1.64**	0.26
	Public	-0.50*	0.24
Start time to group hours to SAT increase	Private	-1.41**	0.25
	Public	-0.65**	0.19
Start time to homework completion to SAT increase	Private	-0.003	0.05
	Public	-0.18	0.09
Moderator: Socioeconomic Status			
Start time to individual hours to SAT increase	Low	-0.75**	0.20
	High	-2.35**	0.38
Start time to group hours to SAT increase	Low	-0.90**	0.18
	High	-1.92**	0.34
Moderator: Baseline Type			
Start time to homework completion to SAT increase	Official	-1.36**	0.36
	Unofficial	-2.50**	0.41

* $p < .05$ **$p < .01$

The moderating effect of SES was similar to that of school type. High SES students who started SAT preparation a month earlier in the junior year gained 2.35 points ($p < .01$) through increased individual tutoring, compared to low SES students who only gained .75 points ($p < .01$). Likewise, starting a month earlier yielded high SES students a larger score gain ($B = 1.92$, $p < .01$) through the indirect effects of increased group tutoring hours than it did their low SES ($B = .90$, $p < .01$) peers.

Baseline type exerted a moderating influence on the indirect effect from start time to homework completion to SAT increase. Students with an unofficial SAT baseline who started preparation a month earlier in the junior year gained 2.50 points ($p < .01$) through increased homework completion, compared to students with an official baseline SAT who only gained 1.36 points ($p < .01$).

Table 7 presents the conditional 3-path indirect effects within the model. School type moderated the indirect effect from start time to individual hours to official tests to SAT increases. By starting SAT preparation a month earlier in the junior year, private school students gained .24 points ($p < .01$) through increased individual tutoring and increased official SAT tests. In comparison, public school students only gained .07 points ($p > .05$) through this pathway. The moderating effect of school type significantly influenced three additional 3-path relations within the model. By starting SAT preparation a month earlier in the junior year, private school students gained .25 points ($p < .01$) through increased individual tutoring and practice tests, .06 points ($p < .05$) through increased group tutoring and official tests, and .33 points ($p < .01$) through increased group tutoring and practice tests. In comparison, through these same pathways, public school students only gained .08 points ($p > .05$), .03 points ($p < .05$); and .15 points ($p < .01$) respectively. Private school students gain more from early start times.

Table 7. Analysis of conditional 3-path indirect effects within the model

Conditional indirect relations	Category	B	SE
Moderator: School Type			
Start time to individual hours to official tests to SAT increase	Private	-0.24**	0.06
	Public	-0.07	0.04
Start time to individual hours to practice tests to SAT increase	Private	-0.25**	0.07
	Public	-0.08	0.04
Start time to group hours to official tests to SAT increase	Private	-0.06*	0.03
	Public	-0.03*	0.01
Start time to group hours to practice tests to SAT increase	Private	-0.33**	0.09
	Public	-0.15**	0.06
Start time to session distribution to HW completion to SAT increase	Private	0.00	0.003
	Public	0.01	0.02
Moderator: Socioeconomic Status			
Start time to individual hours to official tests to SAT increase	Low	-0.11**	0.04
	High	-0.35**	0.09
Start time to individual hours to practice tests to SAT increase	Low	-0.11**	0.04
	High	-0.36**	0.10
Start time to group hours to official tests to SAT increase	Low	-0.04*	0.02
	High	-0.09*	0.04
Start time to group hours to practice tests to SAT increase	Low	-0.21**	0.06
	High	-0.45**	0.12
Moderator: Baseline Type			
Start time to individual hours to official tests to SAT increase	Official	-0.28**	0.08
	Unofficial	-0.51**	0.09
Start time to individual hours to practice tests to SAT increase	Official	-0.11	0.06
	Unofficial	-0.02	0.06

$* p < .05 ** p < .01$

The moderating influence of SES in many ways mirrored that of school type. By starting SAT preparation a month earlier in the junior year, high SES students gained .35 points ($p < .01$) through increased individual tutoring and official tests, compared to low SES students who gained only .11 points ($p < .01$). By starting SAT preparation a month earlier in the junior year, high SES students gained .36 points ($p < .01$) through increased individual tutoring and practice tests, .09 points ($p < .05$) through increased group tutoring and official tests, and .45 points ($p < .01$) through increased group tutoring and practice tests. In comparison, through these same pathways, low SES students

only gained .11 points ($p < .01$), .04 points ($p < .05$); and .21 points ($p < .01$) respectively. High SES students gain more from early start times.

Baseline type exerted a moderating influence on the indirect effect from start time to individual hours to official tests to SAT increases. Students with an unofficial SAT baseline who started preparation a month earlier in the junior year gained .51 points ($p < .01$) through increased an increase in individual hours and official tests, compared to students with an official baseline SAT who only gained .28 points ($p < .01$).

4. Discussion

This research serves as an important first step toward empirically examining the relationship among the various factors that inform successful SAT preparation. The structural model developed for this study addresses the question of how preparatory factors such as time on task, distribution of study, and practice effects mediate the relationship between beginning preparation early and attaining higher SAT scores, and how student characteristics such as gender, school type, and SES influence test-preparation variables.

Building upon the analysis of Appelrouth et al. (2015), our model confirmed that starting tutoring earlier junior year leads directly to more distributed tutoring sessions, more private and group tutoring, greater homework completion, more practice tests and higher SAT scores. Unexpectedly, starting earlier had a slightly negative effect upon the number of official tests a student will take. This finding is difficult to interpret, and counterintuitive as one would anticipate that additional preparatory time would facilitate additional SAT administrations. However the effect is so small, .05 of a test gained per month, that it lacks practical significance.

Other variables also exerted a positive direct effect upon SAT score increases: session distribution, individual and group tutoring hours, practice and official tests. Increases in group and individual tutoring hours positively affected the number of practice and official tests.

Results from the present study reveal numerous intervening variables that help explain how SAT start time relates to final SAT score increases. The results of mediation analyses in the current study provide evidence that session distribution, individual tutoring hours, group tutoring hours, homework completion, number of official tests, and number of practice tests mediate the relationship between tutoring start time and total SAT increases. Thus, gains which appear to derive from early start times must be partially attributed to the greater distribution of sessions, increase in contact hours, greater homework completion and more practice and official tests.

Significant 3-path mediation effects were also found in the model. The number of individual tutoring hours and official SAT tests mediated the relationship between start time and score increases, as did the number of individual hours and practice tests. Similarly, the number of group tutoring hours and official SAT tests mediated the relationship between start time and score increases, as did the number of group hours and practice tests. Some of the benefits of beginning tutoring earlier in a student's junior year can be attributed to a combination of increased tutoring hours coupled with increased practice and official tests.

To examine the moderating role of socioeconomic status, school type, baseline type and gender, interaction terms were included. No conditional effects were found for gender, indicating that gender was not a moderating factor for any relations depicted in the model. School type was a significant moderator of the relationship between start time and individual and group tutoring hours. When private school students started tutoring earlier in their junior year, this translated into a larger gain in the number of individual and group tutoring hours than it did for their public school counterparts. Each additional month of SAT preparation would yield .53 hours of individual tutoring and .73 hours of group tutoring for a student in private school, compared to .16 hours of individual and .34 hours of group tutoring for a public school student. In a similar fashion, school type exerted a significant moderating effect upon the relationship between homework completion and score increases. Completing more of the assigned SAT homework had a stronger effect on score increases for students in public school than for those in private school. The net benefit of completing 100% of assigned homework would be 42 points higher for a student in public school, a practically meaningful increase in terms of college admissions and scholarships.

As with school type, SES moderated the relationship between start time and both individual and group tutoring hours. Starting SAT preparation a month earlier led to an increase in .76 individual and .82 group hours for a high SES student and an increase of only .24 individual hours and .55 group hours for a low SES student, indicating the increased tendency of more affluent students to procure more expensive and individualized forms of SAT preparation.

Baseline type moderated the path between the number of official tests and score increases. Students who began preparing for the SAT by taking an unofficial practice test gained 37.25 points for each official SAT they took, compared to students with an official SAT baseline who gained only 20.30 SAT points. It is plausible that students

who began preparation after already participating in an official SAT administration, gaining valuable experience in the process, derived less benefit from additional official tests than their counterparts who had yet to engage in an official SAT administration. Alternately, students who started tutoring before taking an official SAT may have different characteristics than those who began tutoring after attaining a presumably less-than satisfactory score on an official SAT.

Combining the effects of mediation and moderation identified several conditional indirect effects, which reflected the dependency of mediation on the levels of student characteristics. In examining the 2-path conditional mediation, school type has a moderating effect upon the indirect effect from start time to individual and group hours to SAT increase. Private school students gained more points by starting earlier in the year than their public school counterparts through the mechanism of increased private and group tutoring.

SES exerted a moderating influence on indirect effects from start time to individual and group hours to SAT increases. High SES students who started preparation a month earlier in the junior year gained 2.35 points through increased individual tutoring and 1.92 points through increased group tutoring compared to low SES students who only gained .75 hours and .90 hours, respectively. Starting early gave a bigger advantage to more affluent students through the mechanism of increases in private and group tutoring.

Baseline type exerted a moderating influence on the indirect effect from start time to homework completion to SAT increases. When starting tutoring earlier in the year, students who took an unofficial SAT baseline had greater gains through the indirect effect of homework completion than did students who took an official SAT baseline test. This unexplained finding warrants further investigation.

In examining 3-path conditional mediation in the model, school type had a moderating effect upon the indirect effect from start time to individual hours to official tests to SAT increase in that private students saw larger effects. Private school students also saw larger SAT score gains through increased individual tutoring and practice SAT tests, increased group tutoring and official tests, and increased group tutoring and practice tests. The model suggested that private school students derive greater indirect benefits from starting test preparation early.

The moderating influence of SES in many ways mirrors that of school type. By starting SAT preparation earlier high SES students saw larger SAT score gains through increased individual tutoring and increased official SAT tests, increased individual tutoring and practice tests, increased group tutoring and official tests, and increased group tutoring and practice tests. The three path models add evidence that starting earlier generally conferred greater benefits upon more affluent students through increases in tutoring hours and practice and official tests.

Baseline type exerted a moderating influence on the indirect effect from start time to individual hours to official tests to SAT increases. Students who took an unofficial SAT baseline had greater score gains through the indirect effect of increased individual hours and official tests than did students who took an official SAT baseline test. This contributes support to the finding that students who take a preliminary official SAT may derive smaller gains through tutoring and testing.

5. Strengths and Limitations

The present study has several notable strengths, the first of which involves having access to a data set from a commercial test preparation company, an occurrence infrequently found in the literature. The novelty of the data and the multiplicity of variables measured allow us to make a unique contribution to the field. Additionally, extending the research to a conditional mediation analysis deepened the understanding of the relationship among the various factors related to SAT preparation.

Perhaps the greatest limitation of the data also lies in its source. As this data set comes from a sample that is not representative of the general public, the generalizability of the data is limited. Optimally, we would examine a random sample of students and a control group. However, this is challenging given the nature of using data from a commercial operation and the lack of an easily identifiable control group.

Other limitations of the data involve using home value as a proxy for SES. In many studies, income is used as a proxy for SES, and may be a more valid measure. Additionally, our division of students into high, medium and low SES categories was somewhat arbitrary as nearly all students in the data set would generally qualify as middle to high SES. Most clients with the disposable income needed to hire outside tutors do not fall into the lower levels of SES.

Another limitation of the study involved the measure of homework completion, which was not a standardized measure. Instructors in the company were permitted to assign different quantities of SAT homework to their students based on the circumstances of the tutoring, student characteristics and student schedules. Additionally, the reliability of the assessment of homework completion was not verified. Instructors were asked to determine the rate of

homework completion and assign a value between 0 and 100. The imprecision of this measurement and the lack of standardization of homework assigned render this variable less robust, potentially introducing an element of measurement error. In future studies, a better measure is needed to address the effect of homework completed on changes in SAT scores.

Finally, the non-experimental research design and the data-analytic strategy of the current study limit the ability to interpret the results as causal relationships. The correlational nature of the analysis limits causal inferences, particularly regarding the predicted links among mediators, which may be confounded by reverse causation. Mediation analysis cannot reliably act as a substitute for the experimental analysis of causal mechanisms. Alternative mechanisms may account for the relations found between variables, an inherent limitation of all cross-sectional path models. Additional intervening variables or candidate mediators could be introduced into the model to better explain the relationships. A more complete framework would include other effects, factors of preparation and student characteristics. Thus it is necessary to explore the model within a broader framework, to replicate the results using different samples, contexts, and variables, before drawing firm conclusions regarding the general validity of the model.

6. Theoretical and Practical Implications

The findings of the current study contribute to the existing research in practical ways. First, the findings provide a greater understanding of how to design an effective SAT preparation program. As an increasing number of schools and non-profits are creating their own test preparation programs, it would benefit the administrators and instructors of these programs to learn which factors contribute to successful test preparation outcomes. In this study, we have highlighted some constructive features of a successful test preparation program, such as starting earlier junior year, incorporating timed practice tests, properly spacing out sessions over time, encouraging students to take the SAT multiple times, and providing adequate instructional time.

The identification of mediators within the model helps to clarify which aspects of the tutoring intervention are the core elements, which best predict higher scores. Likewise understanding the effects of moderation helps to identify how different student populations may respond to different aspects of the tutoring intervention.

Mediation outcomes in the model suggest numerous ways to enhance SAT scores. One could directly target specific aspects of the intervention such as building in more practice effects, more contact hours, and more distributed sessions. Alternately, one could indirectly increase practice effects and contact hours through moving the preparation earlier in the school year. Considering the indirect effects of starting earlier, the benefits of early preparation remain compelling. Many school systems offer SAT preparation relatively late to their students, even starting preparation during the senior year of high school. This analysis provides empirical evidence pointing to the importance of beginning SAT preparation earlier, which will ultimately lead to higher SAT increases for students through direct and indirect pathways.

Our investigation into potential moderators revealed that gender yielded no conditional effects, thus administrators need not attend to this factor when constructing a tutoring program. Administrators would be wise to consider that school type and levels of student affluence influence tutoring intervention. When private school students start tutoring earlier in their junior year, this translates to a larger gain in the number of individual and group tutoring hours and official and practice tests than it does for public school students. Leveraging potentially greater financial resources, private school students increase their consumption of individual and group tutoring when they start tutoring earlier. Administrators could help offset this effect for public school students by increasing the number of contact hours in their SAT preparation programs. Administrators could also offer more structured practice tests to public school students to help them attain the same gains from practice effects experienced by private school students.

Administrators could benefit from understanding the moderating influence of SES on tutoring outcomes. When more affluent students start SAT preparation earlier, they see larger score gains through increased individual tutoring, group tutoring, official and practice tests, and through combinations of those factors. To help compensate for this, administrators working with lower SES students should build in increased contact time and practice effects into SAT preparatory programs.

Students who have not taken an official SAT before starting preparation seem to derive greater benefit from both practice and official tests than do students who take an official SAT before engaging in preparation. There may be meaningful differences between students who proactively initiate test preparation before sitting for an official test and those who reactively initiate test preparation after taking an official test. Students who start preparing before taking an official test should plan on building in more practice effects to help them gain the benefits of additional exposure to testing conditions.

Findings from this study have broad implications for anyone interested in designing an effective SAT preparatory course and for individuals aspiring to maximize their own SAT scores. Students preparing for the SAT could benefit from understanding the relationships between starting preparation early and the other factors for successful test preparation: contact hours, practice effects, homework completion and session distribution. When students understand the model for successful SAT preparation, they can better structure their preparation in the most economical and efficient way to optimize their scores.

7. Areas for Future Inquiry

This study focused on building a model of effective test preparation by examining a limited number of preparatory factors. It would be beneficial to extend this inquiry by integrating additional preparatory factors into the model including cognitive variables such as motivation, self-efficacy, and anxiety. The investigation of moderators could be expanded to examine SES using parameters for low, middle, and high SES more in line with nation norms. To establish findings that could generalize to a broader audience, the sample would need to be expanded and become more inclusive. Finally, future research could explore the extent to which our current findings generalize to preparation for other high-stakes assessments such as the ACT.

Declaration of Conflicting Interests

The senior author of this paper, Jed I. Appelrouth, is the owner of the private tutoring center providing archival data and adhered to all procedures and instructions provided to him by the Institutional Review Board of his university.

References

Andersen, S. C., Humlum, M. K., & Nandrup, A. B. (2016). Increasing instruction time in school does increase learning. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), 7481-7484. https://doi.org/10.1073/pnas.1516686113

Appelrouth, J. I., & Zabrucky, K. M. (2017). Preparing for the SAT: A review. *College and University Journal, 92*(1), 1-15.

Appelrouth, J., Zabrucky, K. M., & Moore, D. (2015). Preparing students for college admissions tests. *Assessment in Education: Principles, Policy & Practice*, 1-18.

Arenson, K. W. (2006). Colleges say SAT mistakes may affect scholarships. *New York Times*, March 26, 18.

Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 940-945. https://doi.org/10.1037/a0029199

Bloom, B. S. (1984). The search for methods of group instruction as effective as one-to-one tutoring. *Educational Leadership, 41*(8), 4-18.

Briggs, D. C. (2005). Meta-Analysis: A case study. *Evaluation Review, 29*(2), 87-127. https://doi.org/10.1177/0193841X04272555

Briggs, D. C. (2009). *Preparation for college admission exams* (2009 NACAC Discussion Paper). Arlington, VA: National Association for College Admission Counseling.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354-380. https://doi.org/10.1037/0033-2909.132.3.354

College Board. (2016). *2016 college-bound seniors-total group profile report*. Retrieved October 13, 2016, from https://secure-media.collegeboard.org/digitalServices/pdf/sat/total-group-2016.pdf

Cooper, H, Lindsay, J. J., Nye, B., & Greathouse, S. (1998). Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement. *Journal of Educational Psychology, 90*, 70-83. https://doi.org/10.1037/0022-0663.90.1.70

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987-2003. *Review of Educational Research, 76*, 1-62. https://doi.org/10.3102/00346543076001001

Devine-Eller, A. (2012). Timing matters: Test preparation, race, and grade level. *Sociological Forum, 27*(2), 458-480. https://doi.org/10.1111/j.1573-7861.2012.01326.x

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, *98*, 198-208. https://doi.org/10.1037/0022-0663.98.1.198

Elbaum, B., Vaugh, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Education Psychology*, *92*(4), 605-619. https://doi.org/10.1037/0022-0663.92.4.605

Green, J., Brown, R., & Education Commission of the States. (1983). Student Achievement in Public and Private Schools. *Issuegram 16.*

Hannon, B. (2012). Test anxiety and performance-avoidance goals explain gender differences in SAT-V, SAT-M, and overall SAT scores. *Personality & Individual Differences, 53*(7), 816-820. https://doi.org/10.1016/j.paid.2012.06.003

Hoover, E. (2016). Record-breaking numbers of applicants? Don't gloat. *Chronicle of Higher Education, 62*(30), A8.

Ireson, J. (2004). Private tutoring: how prevalent and effective is it? *London Review of Education, 2*(2), 109-122. https://doi.org/10.1080/1474846042000229458

Kulik, J. A., Kulik, C. C., & Bangert-Drowns, R. L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal, 21*, 435-447. https://doi.org/10.3102/00028312021002435

Lilly, J., & Montgomery, P. (2011). Systematic reviews of the effects of preparatory courses on university entrance examinations in high school-age students. *International Journal of Social Welfare, 21*, 3-12.

Marjoribanks, K. (2003). Learning environments, family contexts, educational aspirations and attainment: A moderation-mediation model extended. *Learning Environments Research, 6*(3), 247-265. https://doi.org/10.1023/A:1027327707647

Nankervis, B. (2011). Gender inequities in university admission due to the differential validity of the SAT. *Journal of College Admission,* (213), 24-30.

National Association for College Admission Counseling. (2015). *2015 State of College Admission.* Retrieved October 29, 2016, from http://www.nacacnet.org/research/research-data/Documents/2015SoCA.pdf

Patterson, B.F., Mattern, K.D., & Swerdzewski, P. (2012). Are the best scores the best scores for predicting college success? *Journal of College Admission, 3*, 35-45.

Powers, D. E., & Rock, D. A. (1999). Effects of coaching on SAT 1: Reasoning Test scores. *Journal of Educational Measurement, 36*(2), 93-118. https://doi.org/10.1111/j.1745-3984.1999.tb00549.x

Preacher, K., Rucker, D., & Hayes, A. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 185-227. https://doi.org/10.1080/00273170701341316

Rawson, K.A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: improving performance on course exams and long-term retention. *Educational Psychology Review, 25*, 523-548. https://doi.org/10.1007/s10648-013-9240-4

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20-27. https://doi.org/10.1016/j.tics.2010.09.003

Rubenstein, J. (2003). Test preparation: what makes it effective? *Measuring up: Assessment issues for Teachers, Counselors, and Administrators,* 397-414. Greensboro, NC: ERIC Counseling and Student Services Clearinghouse.

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin, 135*(1), 1-22. https://doi.org/10.1037/a0013978

Turner, S. L. (2009). Ethical and appropriate high-stakes test preparation in middle school: Five methods that matter. *Middle School Journal, 41*(1), 36-45. https://doi.org/10.1080/00940771.2009.11461702

Vigdor, J. L., & Clotfelter, C. T. (2003). Retaking the SAT. *The Journal of Human Resources, 38*(1), 1-33. https://doi.org/10.2307/1558754

Vlach, H., & Sandhofer, C. (2012). Distributing learning over time: the spacing effect in children's acquisition and generalization of science concepts. *Child Development, 83*(4), 1137-1144. https://doi.org/10.1111/j.1467-8624.2012.01781.x

Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly, 10*, 196-218. https://doi.org/10.1080/15434303.2012.721423

Zwick, R. (2002). Is the SAT a 'wealth test'?. *Phi Delta Kappan, 84*(4), 307-311. https://doi.org/10.1177/003172170208400411