Spring 5-11-2015

Predicting Success In Preparing For High-Stakes Admissions Tests: A Moderated Mediation Analysis

Jed Appelrouth

ACCEPTANCE

This dissertation, PREDICTING SUCCESS IN PREPARING FOR HIGH-STAKES ADMISSIONS TESTS: A MODERATED MEDIATION ANALYSIS, by JED APPELROUTH, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree, Doctor of Philosophy, in the College of Education, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chairperson, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty.

_____
Karen M. Zabrucky, Ph.D.
Committee Chair

_____
DeWayne Moore, Ph.D.
Committee Member

_____
Nannette Evans Commander, Ph.D.
Committee Member

_____
Hongli Li, Ph.D.
Committee Member

_____
Date

_____
Laura D. Fredrick, Ph.D.
Chairperson, Department of Educational Psychology,
Special Education, and Communication Disorders

_____
Paul A. Alberto, Ph.D.
Dean
College of Education

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education's Director of Graduate Studies, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

_____

Jed I. Appelrouth

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Jed I. Appelrouth
2670 Shetland Drive
Decatur, GA 30033

The director of this dissertation is:

Dr. Karen M. Zabrucky
Department of Educational Psychology,
Special Education and Communication Disorders
College of Education
Georgia State University
Atlanta, GA 30303

CURRICULUM VITAE

Jed I. Appelrouth

ADDRESS: 2670 Shetland Drive
Decatur, GA 30033

EDUCATION:

Ph.D.	2015	Georgia State University Educational Psychology
Masters Degree	2007	Georgia State University Professional Counseling
Bachelors Degree	1998	University of Pennsylvania International Studies and Business

PROFESSIONAL EXPERIENCE:

2001-present	Chief Executive Officer Applerouth Tutoring Services Atlanta, GA
2005-2006	Career Counseling Intern Emory University Atlanta, GA
1999-2001	Business Analyst McKinsey and Company Atlanta, GA

PRESENTATIONS AND PUBLICATIONS:

Appelrouth, J. I., Moore, D., & Zabrucky, K. M. (2015). *Preparing students for college admissions tests: A moderated mediation analysis.* Paper submitted for presentation at the annual meeting of the Eastern Psychological Association, Philadelphia.

Appelrouth, J. I., Zabrucky, K. M., & Moore, D. (2015). *Preparing students for college admissions tests.* Paper submitted for presentation at the annual meeting of the Midwestern Psychological Association, Chicago.

Appelrouth, J. I. (2014, November). *Awaken Your Inner Therapist: Counseling Insights for Educational Consultants.* Paper presented at the semi-annual conference of the Independent Educational Consultant's Association (IECA), Orlando, Florida.

Appelrouth, J. I. (2014, June). *The new landscape of college admissions testing.* Paper presented at the annual conference of the Higher Educational Consultant's Association (IECA), Portland, Oregon.

Appelrouth, J. I. (2014, April). *What's coming next in the world of higher education?.* Paper presented at the semi-annual conference of the Independent Educational Consultant's Association (IECA), Indianapolis, Indiana.

Appelrouth, J. I. (2014, February). *The crossroads of educational testing.* Paper presented at the regional conference of the Southern Association of College Admissions Counselors (SACAC), Charleston, South Carolina.

Appelrouth, J. I. (2013, April). *Self-regulated learning strategies and transformational learning.* Paper presented at the semi-annual conference of the Independent Educational Consultant's Association (IECA), Chicago, Illinois.

Appelrouth, J. I. (2012, October). *Taming the beast of test anxiety.* Paper presented at the semi-annual conference of the Independent Educational Consultant's Association (IECA), Atlanta, GA.

Appelrouth, J. I. (2012, May). *Helping your students succeed on high stakes college admissions tests.* Paper presented at the eastern regional conference for the National College Advocacy Group (NCAG), Waltham, Massachusetts.

Appelrouth, J. I. (2012). *Get Your ACT Together: The Fabulous Guide to the ACT*. Atlanta, GA: Applerouth Tutoring Services.

Appelrouth, J. I. (2011, June). *The art of academic motivation.* Paper presented at the annual conference of the Higher Educational Consultant's Association (HECA), Vancouver, CA.

Appelrouth, J. I. (2008). *Ready, SAT, GO: The Fabulous Guide to the SAT*. Atlanta, GA: Applerouth Tutoring Services.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS

2014	Eastern Psychological Association
2014	Higher Educational Consultants Association
2007-2008	Learning Disabilities Association of Georgia
2005-Present	National Certified Counselor (NCC) National Board for Certified Counselors (NBCC)

PREDICTING SUCCESS IN PREPARING FOR HIGH-STAKES ADMISSIONS TESTS: A MODERATED MEDIATION ANALYSIS

by

JED I. APPELROUTH

Under the Direction of Dr. Karen M. Zabrucky

ABSTRACT

Research on preparing for standardized college admissions tests such as the SAT has largely been limited to studies exploring the relative effect size of test preparation. In several analyses and meta-analyses, investigators have demonstrated a positive effect of test preparation (e.g., Briggs, 2005; Kulik, Bangert-Drowns & Kulik, 1984; Lilly & Montgomery, 2011; Powers & Rock, 1999). Moving beyond the fundamental question of whether SAT prep influences test scores, researchers have only recently begun to explore the individual factors that inform successful test preparation. In their regression analysis of the salient factors of successful SAT preparation, Appelrouth, Moore, & Zabrucky (2014) 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 the factors involved in successful SAT preparation. It was hypothesized that there would be direct and indirect relationships between the factors of test preparation, and that some of these relationships would be moderated by student characteristics such as gender and socioeconomic status. Archival data from 1,933 students, provided by a private tutoring company, were analyzed. 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. Starting tutoring earlier junior year yielded a number of positive direct and indirect effects. 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. Though gender had no significant moderating effects, both school type and socioeconomic status moderated the relationship between start time and individual tutoring hours. 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.

INDEX WORDS: College Admissions, SAT, SAT Coaching, High-Stakes Tests, College Entrance Examinations, Test Preparation, Mediation, Moderation

PREDICTING SUCCESS IN PREPARING FOR HIGH-STAKES ADMISSIONS TESTS: A MODERATED MEDIATION ANALYSIS

by

JED I. APPELROUTH

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of

Doctor of Philosophy

in

Educational Psychology

in

The Department of Educational Psychology,
Special Education and Communication Disorders

in

the College of Education

Georgia State University

Atlanta, GA
2014

ACKNOWLEDGMENTS

It is a pleasure to thank those who have played an essential role in the completion of my dissertation and my degree. I owe the most profound gratitude to my committee chair and advisor Dr. Karen Zabrucky, without whose confidence and encouragement and understanding, I may well have strayed too far from the path to complete this significant task. Dr. Zabrucky worked with me through all the ups and downs of the last 7.5 years: the trials and tribulations of my work life, my personal life, and my academic life. Whenever I needed her, Dr. Zabrucky was there- like a reflection. She offered invaluable guidance and helped me find my way through all the phases of this degree. Dr. Zabrucky was the perfect mentor and advisor for me, and I remain in her debt.

I must thank Dr. DeWayne Moore who guided me through the statistics of my residency project and dissertation. Dr. Moore proved an invaluable teacher and guide through the challenging territory of higher analysis. When my statistical wheels were spinning, Dr. Moore provided just the right support to get me back on my feet and moving forward. He is the most skillful statistics teacher I have ever come across. I remain in his debt.

I must also thank Dr. Nannette Commander and Dr. Hongli Li who likewise passionately supported me throughout this process. They were rooting for my success from the beginning, as well as providing valuable guidance and insights. They helped improve and focus the research and strengthen the study. I am ever grateful for the role they played in the completion of this degree. Thank you all for this wonderful gift in my life!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACT	American College Testing
ETS	Educational Testing Service
GPA	Grade Point Average
PROCESS	Software Macro for SPSS created by Preacher and Hayes
PSAT	Preliminary SAT
NACAC	National Association for College Admissions Counseling
SAT	Formerly Scholastic Aptitude Test, currently no meaning
SES	Socioeconomic Status
TA	Test Anxiety

1. FACTORS PREDICTING SUCCESS ON THE SAT: A REVIEW OF THE LITERATURE

The college admissions process is more competitive than ever, and high school students and their families are going to great lengths to positively influence admissions outcomes. Students are taking increasingly strenuous course loads (College Board, 2012a) and are carefully selecting extracurricular activities to enhance their admissions portfolios. They are submitting applications to a greater number of colleges (Cohen, 2012), and tens of thousands of families are hiring professional coaches and consultants to ensure their children's applications and essays will stand out during the admissions process (Greisemer, 2012). Families are using their resources to help their children attain the SAT and ACT scores needed for both admissions and scholarships. Inside and outside of high school, students are preparing for these high-stakes admission tests, and a thriving test preparation industry has emerged to meet the growing demand for competitive admissions test scores (Buchmann, Condron, & Roscigno, 2010).

The SAT and ACT profoundly impact the college admissions process (Patterson, Mattern, & Swerdzewski, 2012), with over 88% of four-year colleges and universities placing moderate or considerable importance on admission test scores (NACAC, 2012). In their efforts to manage an ever-mounting tide of applications, admissions committees use SAT and ACT scores to compare applicants and infer the likelihood of student collegiate success (Buchmann et al., 2010; Lane, Kalberg, Mofield, Wehby, & Parks, 2009). Colleges also carefully attend to the average SAT and ACT scores of their incoming classes, as these scores factor directly into colleges' financial bond ratings (West-Faulcon, 2009) and their *U.S. News & World Report* collegiate rankings (U.S. News, 2012), which can impact the number and quality of applicants as well as alumni giving (Meredith, 2004; Morse, 2013). With so much riding on test scores, even marginal testing gains can affect a student's admissions outcomes. In a 2009 survey conducted

by the National Association of College Admission Counseling, more than one third of postsecondary institutions responded that an improvement of only 20-30 points on the SAT could "significantly improve a student's likelihood of admission" (Briggs, 2009).

Beyond helping a student gain admission to college, higher test scores can help students pay for college, as many colleges use test scores to determine who will receive financial aid. Hundreds of colleges and universities have established testing thresholds, coupled with GPA requirements, to award scholarship funding (Lilly & Montgomery, 2011; Morgan & Michaelides, 2005). Many colleges use sliding scales to establish aid thresholds, and, according to Dennis Trotter, Vice President and Dean of Admissions at Franklin and Marshall College, "a swing of 80 or 100 points on the SAT could mean the difference between the highest level scholarship or not receiving one at all" (Arenson, 2006, p. 18).

Given the profound impact of testing on the college admissions and scholarship allocation processes, it is not surprising that more students than ever before are taking high-stakes college admissions tests (Alon & Tienda, 2007; Buchman et. al, 2010). In 2000, 1,260,278 students took the SAT (College Board, 2007a) and 1,065,138 students took the ACT (ACT Inc., 2000). In 2013, 1,672,395 students took the SAT (College Board, 2014) and 1,799,243 students took the ACT (ACT Inc., 2013). This equates to a 49% net increase in the total number of test takers in a little over a decade. With the increasing prevalence and importance of college admissions testing, numerous researchers have turned their attention to these tests.

Research on the SAT

In the present literature review I will examine the influence of several variables on SAT performance. For decades, academic researchers have investigated how individual student

differences and aspects of test preparation affect SAT scores. Although in my empirical investigation (discussed in Chapter 2) I examine the student variables of gender and socioeconomic status, along with characteristics of test preparation, other researchers have examined additional student variables which warrant our attention. These variables, outside of the scope of the empirical study, are included in the first section of this literature review.

The Effects of Individual Student Differences on SAT Scores (outside the scope of the dissertation study)

Cognitive Factors

Given its initial construction as a modified IQ test (Lemann, 2000), the SAT has always been highly correlated with intelligence. Despite the SAT's numerous revisions during the last century, Frey and Detterman (2004) found that the modern SAT continues to correlate highly (.82) with *g*, the "general factor" of intelligence. Coyle and Pillow (2008) found that the current SAT, modified in 2005 with the addition of the Writing section and the removal of abstract elements such as analogies and quantitative comparisons, continues to correlate with *g* in the .78-.82 range.

Working memory, a factor of intelligence, is the ability to keep a mental representation active while engaged in additional processing (Engle, Tuholski, Laughlin, & Conway, 1999). Turner and Engle (1989) found significant correlations between measures of working memory and verbal (r = .28 to r = .34) and math (r = .26 to r = .33) scores on the SAT (p.135). In an investigation of the construct validity of the reading comprehension section of the SAT, Daneman and Hannon (2001) confirmed the significance of the SAT-working memory relationship, finding correlations between measures of working memory and SAT performance ranging from r = .30 to r = .53, depending upon the test-taking strategy employed.

Academic Self-Efficacy

Research studies and meta-analyses have confirmed the positive impact of self-efficacy, confidence in one's ability to succeed at a given task (Bandura, 1977), on academic performance (e.g., Bandura & Schunk, 1981; Chemers, Hu, & Garcia, 2001). Robust self-efficacy enhances student effort and resilience in the face of challenges (Pajares, 2002), and leads to greater use of cognitive and metacognitive problem solving strategies (Schunk, 2003). Robbins, Lauver, Le, Davis, Langley, and Carlstron (2004) found that academic self-efficacy was positively correlated with ACT/SAT performance ($r = .22$).

Test Anxiety

Some students experience test anxiety when facing high-stakes tests such as the SAT. Elliot and McGregor (1999) found evidence that test anxiety mediated the relationship between performance-avoidance goals and SAT performance. Students focused on avoiding failure on the SAT had a greater likelihood of experiencing test anxiety and achieving diminished SAT outcomes than did those students with alternate goal orientations.

Cognitive and Non-cognitive Factors

Hannon and McNaughton-Cassill (2011) invested how cognitive and non-cognitive factors in concert affect SAT performance. They found that knowledge integration, the ability to integrate prior knowledge from long-term memory with novel information acquired from text, accounted for 20% of the variance in SAT scores. Working memory and knowledge integration together accounted for 37.8% of the variance in SAT scores; when epistemic beliefs of learning and test anxiety were added to the analysis, these four variables accounted for 43.4% of the variance in SAT scores.

High School Curriculum and Extracurricular Activities

Curricular choices in middle and high school significantly affect SAT scores (Brody & Benbow, 1990; Rigor Boosts SAT Performance, 2010). Tracking the score gains of students who took the SAT in 7th grade and then again in the 11th and 12th grades, Brody & Benbow (1990) found that students who pursued rigorous verbal courses in high school, especially in foreign languages, achieved greater gains in SAT verbal scores, just as those students who pursued rigorous courses in math and science attained higher gains in SAT math scores. Correlational studies from the College Board (Rigor Boosts SAT Performance, 2010) confirmed that students who completed a more rigorous "core curriculum" scored 151 points higher on the SAT than did those students who pursued a less rigorous program of study. Likewise students who participated in honors or Advanced Placement courses attained significantly higher SAT scores than did their peers who enrolled exclusively in non-honors classes (Rigor Boosts SAT Performance, 2010). Participation in extracurricular activities both in and out of high school also contributed to higher SAT scores (Everson & Millsap, 2004; Gerber, 1996; Marsh & Kleitman, 2002). Researchers postulate that reasoning abilities which are assessed on tests like the SAT are developed in activities both inside and outside of the classroom (Everson & Millsap, 2004**).**

Race/Ethnicity

In 2013 the average three-section SAT score for all test-takers was 1497. Examining SAT scores through the lens of ethnicity reveals a profound and disturbing disparity. The average composite SAT score for Asian Americans was 1651, compared to 1576 for White students, 1353 for Latinos and 1278 for African Americans (College Board, 2014). The sizeable gaps between ethnic groups has troubled many policy makers and researchers. A number of researchers have investigated whether formal test preparation was a contributing factor to the

racial disparity on SAT performance. Although one might suspect that students from ethnic groups attaining the highest average scores would be the greater consumers of test-prep, perhaps as a response to historic score patterns, Black and Hispanic students were found to be *more* likely than comparable White students to engage in formal SAT preparation activities (Alon, 2010; Buchmann et al., 2010; Devine-Eller, 2012). With the lowest average SAT scores of any ethnic group, Black students used test preparation to a greater degree than did students from all other ethnic groups (Alon, 2010; Buchmann et al., 2010; Devine-Eller, 2012). Buchmann and colleagues found that Black student's elevated use of test-preparation did help reduce the score gap, but the effect size was small. One issue with this study is that it did not take into account the *quality* of the preparation offered to different students, which may be highly variable. Clearly, blanket participation in formal SAT preparation is not a panacea to resolve the racial disparities in SAT scores.

Calculator Use

Researchers have found small, but significant effects of increased calculator use on the SAT (Scheuneman, Camara, Cascallar, Wendler, & Lawrence, 2002). Students who score higher on the math section of the SAT are more likely to have calculators and to use them more frequently. Students who used scientific calculators outperformed those who used four-function calculators, and students who used graphing calculators had the strongest performance of any group.

The Effects of Individual Student Differences on SAT Scores (examined in the dissertation study)

Socioeconomic Status

Investigators have examined the effect of affluence on test preparation and academic outcomes. Compared to their more affluent peers, students of lower socioeconomic status (SES)

tend to perform at lower levels both in school and on high-stakes tests (Aikens & Barbarin, 2008; Baker & Johnston, 2010). Considering investment in education a "natural outgrowth of social capital" (Coleman, 1988), affluent families are more likely to use their superior resources to secure educational advantages for their children. Affluent students are more likely to partake in after-school tutoring and test preparation activities deemed "shadow education" (Buchmann et al., 2010; Grodsky, 2010; Sigal, 2010; Soo-yong, Schofer, & Kyung-keun, 2012).

Some critics of the SAT have argued that it is a de facto "wealth test," useful solely as a measure of SES (Zwick, 2002). A notable critic of the SAT, Alfie Kohn (2001, p. B12), argued that the only thing measured by the verbal section of the SAT is the "size of students' houses." Challenging this assumption, researchers have found that although SAT scores are clearly related to SES (r =.42), the relationship between SAT scores and post-secondary GPA is, to a great extent, independent of SES (Sackett, Kuncel, Arneson, Cooper, & Waters, 2009). When statistically controlling for SES, these researchers have found that the correlation between SAT scores and college grades drops only slightly, from r = .47 to r = .44, revealing that the predictive validity of the SAT is not a mere artifact of SES (Sackett et al., 2009). Other researchers (Zwick & Green, 2007) have found that when SAT scores are examined using a within-high-school analysis, versus a pooled across-high-school analysis, SAT scores have smaller associations with SES. Schools whose students have higher than average levels of SES tend to have higher than average SAT scores, but the SAT-SES relationship weakens greatly when examined through the lens of an individual high school (Zwick & Green, 2007). Within the context of a high school with a particular SES profile, student factors such as ethnicity and maternal and paternal education levels also contributed to SAT scores, independently of SES. Thus an SAT score is by no means a perfect proxy for SES.

Gender

The gender gap on the SAT has drawn the scrutiny of researchers for decades (e.g., Loewen, 1988; Mau & Lynn, 2001; Nankervis, 2011). Females consistently perform below males on the SAT, despite their superior academic performance—measured as GPA in high school and in college (Hannon, 2012; Horn, 1989; Loewen, 1988). The College Board (2014) reported that in 2013, males had an average three-section composite SAT score of 1510 (499 critical reading, 530 math, 481 writing) compared to 1486 for females (495 critical reading, 499 math, 492 writing). Males performed slightly better on critical reading, but significantly better on math, whereas females performed better on the writing section. The most pronounced math performance gap comes at the top of the SAT scoring pyramid: 9.6 percent of males—74,461 out of a total sample of 778,124—scored 700 or above, compared to only 5.2 percent of females —46,040 out of 886,337 (Women in Academia Report, 2013).

This underperformance by females on the SAT starkly contrasts with superior female performance in numerous other academic domains. Females typically exhibit more ambitious course-taking, a greater degree of self-discipline, and superior academic achievement during high school (Duckworth & Seligman, 2006). Duckworth and Seligman examined the fact that throughout elementary, middle and high school, girls earn higher grades than boys in all major subjects, though they score lower on many standardized tests including the SAT, ACT, and AP exams. Duckworth and Seligman suggested that superior self-discipline helped girls achieve consistently higher grades than boys. However, the attribute of superior self-discipline seemed to help girls "less on achievement tests and minimally on tests of intellectual aptitude" (2006, p. 205).

In light of their superior academic attainments in high school, to what can we attribute the persistent underperformance of female test takers on the SAT? While some critics have attacked the test as inherently biased, Wright, Palmer and Miller (1996) found no inherent gender bias in the SAT. Other researchers have looked to biological sex differences to explain the gender gaps on SAT scores, examining how established gender differences in visuospatial abilities and mental rotation resulting from differential in-utero testosterone exposure could affect test performance (Burton, Henninger & Hafetz, 2005; Halpern, Benbow, Geary, Gur, Hyde, & Gernsbacher, 2007).

Ramos and Lambating (1996) found a relationship between a student's risk-taking behavior, defined by the number of items that student omitted, and performance on the math section of the SAT. The researchers found that females omitted more items and exhibited less risk-taking behavior than males, which contributed to lower scores on the SAT.

Researchers have also explored how males and females approach problems on the math portion of the SAT. Gallagher (1992) found that when students could solve SAT math problems using either conventional or unconventional strategies, females favored conventional and algorithmic strategies, while males were more likely to use unconventional and insightful strategies. Females performed better on SAT math problems that were well-defined with more straightforward solutions such as computational and algebra problems; males outperformed females on problems that were less well-defined and required unconventional strategies such as estimation, logic and insight (Gallagher, 1992; Gallagher & De Lisi, 1994). Further analysis revealed that students who performed the best on the math section of the SAT were those with greater self-efficacy for math, due to their willingness to persist longer on problems that they could not solve immediately.

Gender differences in self-efficacy, perceptions about testing and anxiety are also cited as potential contributing factors to the gender gap in SAT scores (Byrnes & Takahira, 1993). Speth and Brown (1990) found that females may consider multiple choice tests like the SAT "more problematical or challenging, requiring more time and effort to prepare for" (p. 199). In some instances, perceptions of the SAT as a challenge give way to perceptions of the test as a threat, inducing anxiety. Females typically display a greater degree of academic anxiety (Altermatt & Kim, 2004) and test anxiety (Hembree, 1988) than boys. Consequently, this may impair performance on the SAT. In one analysis, female test anxiety and performance-avoidance goals accounted for *all* of the significant gender differences on the SAT (Hannon, 2012). Other researchers found evidence that stereotype threat impaired female performance on high-stakes standardized tests such as the SAT (Spencer, Steele, & Quinn, 1999; Danaher & Crandall, 2008). These findings suggest that a number of discrete variables may be contributing to the lower scores of females on the SAT.

The Effects of Test Preparation on SAT Scores (examined in the dissertation study)

The desire to attain a higher score on the SAT transcends gender, racial, and SES boundaries. Across the demographic spectrum, students and their parents, teachers and school administrators are going to great lengths to raise SAT scores. A thriving industry has emerged to assist students in this process by providing an unprecedented array of SAT preparatory options: books, flashcards, online programs, one-on-one tutoring, large groups and small groups—offered in their own high schools, through community organizations and commercial test preparation companies (Lilly & Montgomery, 2011). Low-income students can receive free or steeply-discounted tutoring through charitable foundations and non-profits (e.g., Boys and Girls Clubs),

whose initiatives are frequently supported by commercial test preparation companies (Devine-Eller, 2012).

As test preparation programs have proliferated, researchers have examined the efficacy of SAT preparation (Briggs, 2005; Kulik, Bangert-Drowns et. al, 1984; Lilly & Montgomery, 2011; Powers, 1993; Witt, 1993). Nearly all researchers have found that test preparation has a positive effect, though the debate endures regarding the magnitude of the effect. In a recent meta-analysis of 14 randomized studies of SAT preparation effectiveness, Lilly & Montgomery (2011) found that students who received SAT preparation interventions achieved gains over control groups of 23.5 points on the verbal subtest and 32.7 points on the math subtest for a combined score gain of 56.2 points.

Selection Bias Potentially Confounds Test Preparation Effects

When evaluating the impact of test preparation, many educational researchers have called attention to the issue of selection bias as a confounding research factor (Buchmann et al., 2010; Byun & Park, 2011; Devine-Eller, 2012; Messick & Jungeblut, 1981; Powers & Rock, 1999; Ryan, Ployhart, Greguras, & Schmit, 1998). Messick and Jungeblut (1981) raised concerns that "certain personal factors characteristic of students attending a particular coaching program, such as motivation or career aspirations, may be responsible, at least in part, for subsequent SAT performance that appears to be the result of the coaching experience" (p. 193).

Researchers have found that students who seek out SAT coaching differ from non-coached students in a variety of ways. Students who pursue SAT preparation tend to have higher grades in school (Messick & Jungeblut, 1981; Powers & Rock, 1999), higher academic aspirations (Buchmann et al., 2010) and higher baseline PSAT scores (Messick & Jungeblut, 1981). They tend to be enrolled in more rigorous classes (Buchmann et al., 2010), attend high

schools that helped them plan for college "very well" (Devine-Eller, 2012), and participate more in extra-curricular activities (Devine-Eller, 2012). Coached students are more likely to prepare for the SAT in a variety of ways, are somewhat more nervous about taking the SAT, and tend to place a higher degree of importance on attaining good scores than their uncoached classmates (Powers & Rock, 1999). Additionally, the parents of coached students tend to have greater incomes (Messick & Jungeblut, 1981), more formal education (Powers & Rock, 1999) and a greater degree of involvement in their children's education (Buchmann et al., 2010). Students from affluent families are more likely to use the most expensive forms of test preparation such as commercial courses and individual tutors (Buchmann et al., 2010), and females are more likely than males to use all forms of test preparation (Buchmann et al., 2010).

Distinguishing Between Different Forms of Test Preparation

Rubenstein (2003) noted that many investigations into test preparation have mistakenly lumped together different forms of preparation, failing to distinguish between 40-hour intensive programs and those consisting of a single two-hour after-school session. The failure to make these distinctions potentially confounds research results, as different forms of SAT preparation vary greatly in their efficacy, their cost, and their accessibility (Buchmann et al., 2010; Sigal, 2010).

Using data drawn from the National Education Longitudinal Study, Buchmann and colleagues (2010) found a great variance in the manner in which students approach SAT preparation. Twenty-seven percent of a sample of 8,150 SAT-takers used no preparation and 40 percent used nothing beyond books, videos, or software. For students participating in organized coaching, a high school course was the "highest" level of preparation for 15 percent of the sample, an external group course offered by a commercial test preparation company was the

highest level for 11 percent of the sample, and a private tutor was the highest level for seven percent of the sample. Nearly half of all students in the sample (46 percent) who reported using any form of test preparation used two or more types of preparation, and only two percent exclusively used private tutoring or group classes offered by a commercial test preparation company. In regards to effectiveness, combining different forms of preparation significantly improved test scores (Buchmann et al., 2010), and students working with private tutors or taking commercially available group classes outside of their schools had the largest score gains (see also, Sigal, 2010).

Which Factors Contribute to the Efficacy of Test Preparation?

Researchers have begun to shift from questioning whether test preparation works to asking *why* it works and *how* to make it more effective. In Bond's (2008) analyses of test preparation programs, programs that emphasized test wiseness, (e.g., coaching test-taking fundamentals such as time allocation and guessing strategies) could diminish test anxiety and increase self-efficacy for testing, thereby enhancing score gains. Powers (1993) conceptualized the benefits of SAT coaching as two-fold: direct benefits through learning strategies and test-content as well as indirect benefits through enhanced confidence and diminished anxiety. Scruggs & Mastropieri (1992) delineated four common attributes of successful test preparation programs: baseline pre-tests, effective modeling of problem-solving strategies by trained instructors, accurately calibrated practice tests and timed post-tests. Rubenstein (2003) similarly attributed successful preparatory outcomes to instruction in basic skills, test strategies, problem-solving skills, and taking full-length practice tests.

Researchers have identified several factors that may contribute to positive outcomes of SAT test preparation:

Time on Task

The significant positive relationship between time on task and academic gains is well established (Seifert & Beck, 1984). Increased time on task is correlated with enhanced academic outcomes and has been shown to diminish academic anxiety (Guida, Ludlow & Wilson, 1985). Researchers who have investigated SAT coaching programs have found that the size of score effects is associated with greater contact time in coaching programs (Appelrouth, Zabrucky & Moore, 2014; Messick & Jungeblut, 1981; Lilly & Montgomery, 2011). Messick and Jungeblut found a positive relationship between the size of SAT score increases and the number of hours in a preparation class. Lilly and Montgomery found this effect only for math scores and found that preparation beyond eight hours for the verbal section of the SAT did not yield any additional score gains. In a meta-analysis of SAT preparation effectiveness, Lilly & Montgomery (2011) found that duration of coaching was a factor in the efficacy of preparation programs, whereby math scores were significantly increased in programs with more than eight hours of instruction.

Using Effective Instructional Formats

Different forms of SAT preparation vary in their effectiveness. Some students prepare for the SAT independently using books, flashcards, or online resources. Other students prepare for the SAT working with a private tutor, while others prepare in a small group or a large group setting. Researchers have only recently begun to attend to differences in instructional formats, student-teacher ratios, tutor quality, and the context of instruction (Ireson, 2004). In her review of the efficacy and prevalence of tutoring, Ireson proposes that researchers should distinguish between the various forms of tutoring—large group, small group, and individual (one-on-one) tutoring—due to their varying levels of effectiveness. In the domain of SAT instruction,

researchers have found evidence that individual tutoring is significantly more effective than group tutoring at raising SAT scores (e.g., Appelrouth et al, 2014).

Other researchers have provided evidence pertaining to the efficacy of individual and small group tutoring. Walberg's (1984) meta-analysis of the factors which predict academic success found that tutoring had one of the largest instructional effects on academic attainment, larger than individual factors such as home environment and SES, and educational variables such as assigned homework, class size, teacher expectations and individualized instruction. In their meta-analysis of 65 studies exploring the efficacy of individual and small-group tutoring programs, Cohen, Kulik, & Kulik (1982) found that students who engaged in individual or small-group tutoring achieved educational attainments averaging .40 standard deviations above those of students engaged exclusively in standard classroom instruction. Elbaum, Vaughn, Hughes, & Moody, (2000), in a meta-analysis of 29 studies exploring the efficacy of individual tutoring for reading, found an identical effect size: on average, students in the individual tutoring condition performed at a level .40 standard deviations higher than that of students in the comparison group, who did not receive tutoring.

Bloom (1984) found a larger effect size of individualized tutoring: his research indicated that students in the individual tutoring condition performed at a level 2.0 standard deviations above that of students in the standard classroom learning condition. In his evaluation of individual versus group instruction, Bloom found that the average student who was tutored individually outperformed 98 percent of the students in **a** control class who were taught using group instruction. Ninety percent of the individually tutored students attained the "summative achievement" reached by only the highest 20 percent of students under conventional instructional conditions (Bloom, p. 5).

Bloom (1984) attributed the success of individual tutoring to the efficacy of the "feedback-corrective process of mastery learning (p.7)." The elevated level of feedback present in the individual tutoring format corresponds with an elevated level of inquiry, as more questions are asked by both the tutor and the tutee than are asked in a conventional classroom setting (Graesser & Person, 1994). Ireson (2004) found that personalization, the tutor's ability to tailor instruction specifically to the needs of an individual student and address that student's particular weaknesses, was a major contributing factor to the efficacy of individual tutoring.

According to Bloom (1984), the style of instruction found in the individual tutoring format, can be transferred to the classroom setting. When this "feedback-corrective" method of instruction was used systematically in a group setting, the achievement of the average student was one standard deviation above that of the average student in a control class, even when both classes were taught by the same teacher, using nearly identical instructional material (Bloom, 1984). By this analysis, it is not the teacher, nor the time on task, but the instructional method that makes individual tutoring effective. Other researchers have found evidence that the quality of the tutor does play a significant role in the academic outcome for the student. The degree of training of the tutor is a significant predictor of a student's academic success in the individual tutoring condition (Elbaum et al., 2000).

Although most of the research conducted on tutoring explores the direct effect of tutoring on academic performance, a number of researchers have found that the benefits of tutoring transcend academic gains. Following the completion of the tutoring intervention, compared to students who received instruction in other formats, individually-tutored students demonstrated the most positive changes in attitudes and interests (Bloom, 1984). Likewise, compared to students who did not participate in tutoring, by the conclusion of the school year, those students

who participated in small, after-school tutoring groups, consisting of two to four students, reported decreases in test anxiety, increased academic motivation and enhanced academic self-concept (Mischo & Haag, 2002).

Timing of Test Preparation

Turner (2009) identified "judicious timing of test preparation" as one of the most important features of an effective test preparation program. Well-spaced preparation, incorporating regular breaks between sessions, was found to be superior to cramming (Turner, 2009). Devine-Eller (2012) investigated the effects of optimal timing and found that to increase scores, preparation "must be undertaken with the proper amount of lead time—not so little that it is rushed, and not so much that the skills training atrophies" (p. 463). The National Association for College Admissions Counseling recommends that students commence test preparation in 11th grade (National Association for College Admission Counseling, 2008), and Devine-Eller warns that students who begin test preparation earlier may be "jumping the gun' given their lack of requisite knowledge. Conversely, she acknowledges that many high school seniors who are engaging in test preparation for the first time may have started too late. According to Devine-Eller, although nearly fifty percent of 12th graders prepare for the SAT, "the savviest and most academically competitive will have completed most of their prep in 11th grade. This interpretation is supported by [her] findings that 12th graders whose families are in the highest income quartiles are each a third less likely to prep" (p. 475) than students of lower levels of SES. Devine-Eller's findings regarding the benefits of early preparation are supported by Appelrouth et al., (2014), who found a significant effect on score increases of starting SAT preparation earlier in the junior year.

Benefits of Distributed Study

Researchers have found that students benefit from spaced practice conditions, distributing learning events across time, rather than massing events in close succession (Baldwin, & Ford, 1988). In a meta-analysis of 317 experiments, Cepeda, Pashler, Vul, Wixted, & Rohrer (2006) found that learners consistently demonstrated increased long-term performance on spaced, rather than massed learning schedules. In another meta-analysis of the distribution of practice effects, Donovan & Radosevich (1999) found spaced practice conditions superior to massed practice. They found that individuals in spaced practice conditions outperformed those in massed practice conditions by almost one half of a standard deviation. They also found that the optimal interstudy interval appears to be related to the type of task being learned. To optimize learning, more complex tasks require longer rest periods.

Rawson, Dunlosky, & Sciartelli (2013) found spaced study to be highly potent for enhancing learning and memory. Combining the use of practice tests and spacing study over time led to a system of "successive relearning" which enhanced student performance (p. 523). Elbaum et al. (2000) examined the impact of distributed study in private tutoring and found that learning intervals that were too long weakened performance. Interventions lasting up to 20 weeks, with a mean instruction time of 63 hours, had a mean weighted effect size of 0.65. In contrast, interventions lasting longer than 20 weeks, with a mean instruction time of 61 hours, had a mean weighted effect size of 0.37. This suggests that care must be taken to find the optimal interval for spacing study, as intervals that are either too short or too long can diminish academic outcomes. In the domain of SAT preparation, Appelrouth et al. (2014) found a benefit of distributing study, with a positive effect of 8.96 SAT points gained for every additional day

between tutoring sessions. By their analysis, cramming was inferior to distributing sessions across time.

Taking Practice Tests

Taking practice tests is one of the most potent strategies for enhancing learning (Rawson & Dunlosky, 2012). For decades, researchers have found evidence that administering practice tests enhances retention, overall learning and final testing outcomes (Appelrouth et al., 2014; Arnold & McDermott, 2013; Izawa, 1968; Johannessen & Kahn, 2001; Kulik, Bangert-Drowns et al., 1984; Kulik, Kulik, & Bangert-Drowns, 1984; Rawson & Dunlosky, 2012; Rawson et al., 2013; Reeve & Lam, 2007; Roediger & Butler, 2011). Retrieving information during a test acts as a powerful mnemonic enhancer, strengthening the memory of learned content for future retrieval events (Arnold & McDermitt, 2013; Roediger & Butler, 2011). Additionally, the feedback from practice tests helps students better calibrate their comprehension and minimize potential overconfidence in the absence of such feedback (Rawson & Dunlosky, 2012).

Retrieval practice through practice tests provides greater long-term retention than does repeated study (Roediger & Butler, 2011). Full-length practice SATs, according to Rubenstein (2003), help students to practice and better assimilate strategies they have learned, enhance mental endurance and acclimate to official testing conditions. Practice, absent corrective feedback, however, may not be as beneficial for students. When students choose a "lure" or wrong answer on a practice SAT, if they do not receive corrective feedback, they will be more inclined to choose that same wrong answer on subsequent tests (Roediger & Butler, 2011). Corrective feedback has been shown to enhance testing outcomes and rates of retention (Kang, McDermott, & Roediger, 2007).

Researchers have defined practice effects as "improvements in cognitive test performance due to repeated evaluation with the same or similar test materials" (Duff, Callister, Dennett, & Tometich, 2012, p. 1117). Researchers have found evidence that test-taker characteristics impact the degree to which test-takers will benefit from practice effects. Reeve & Lam (2007) have demonstrated that when it comes to practice effects, "the rich get richer" (p.228). Students with higher baseline scores benefit more from practice and achieve greater score gains than do their peers with lower baseline scores. Kulik, Kulik, et al., (1984) also found that after taking a single practice SAT, lower-ability students experience smaller testing gains compared to middle- and upper-level students. To account for this finding, researchers posit that higher ability students may learn more from experience, potentially a reflection of higher IQ and more efficient learning, and their motivation levels may be bolstered by the positive reinforcement from practicing (Reeve & Lam, 2007).

Educational researchers have provided evidence that more retrieval attempts may lead to greater testing gains (Appelrouth et al., 2014; Arnold & McDermott, 2013; Rawson & Dunlosky, 2012; Roediger & Butler, 2011). In a meta-analysis of 40 studies, Kulik, Kulik, et al., (1984) found that effect sizes increased with the number of practice tests given, demonstrating score gains up to seven rounds of practice. Their results suggest that students would score 20 points higher on an official SAT, after taking one simulated SAT test. Given four to six hours of practice trials, the time needed to administer two full practice SATs, it would not be unusual for students to achieve gains of 40 points on the SAT (Kulik, Kulik, et al., 1984, p. 444). Appelrouth et al. (2014) found that when the effects of practice SATs were isolated in a regression model, each practice SAT contributed 5.16 points to a student's final SAT increase, a smaller effect than that proposed by Kulik, Kulik et. al.

Taking the SAT Multiple Times

Just as students benefit from taking multiple practice tests, so too do they benefit from taking the official SAT multiple times (Vigdor & Clotfelter, 2003). Buchmann and colleagues (2010) found that roughly 85 percent of students take the SAT once or twice, and 15 percent take the test three or more times. The rate of retaking is higher for students applying to selective institutions, and those students who take the test multiple times are generally rewarded with higher scores (Vigdor & Clotfelter, 2003). Vigdor & Clotfelter provided evidence that the tendency to achieve higher scores upon re-examination was not simply a result of selection bias of the pool of retakers. Rather it was a reflection of the benefits derived from multiple testing administrations, including increased test familiarity and substantive knowledge gains. Supporting this hypothesis, Patterson, Mattern and Swerdzewski (2012) found that of the 92,634 students (of their sample of 150,377) who took the SAT more than once, 64 percent of examinees achieved their highest single administration score on their latest (final) SAT. Appelrouth et al. (2014) corroborated this finding, demonstrating that students preparing for the SAT gained an average of 22.62 points for every official SAT administration they took.

In addition to the cumulative score gains which result from taking multiple official SATs, another benefit to students is that many college admissions departments will use the highest section scores from each individual section of the SAT, even if these scores were attained on different testing dates (Vigdor & Clotfelter, 2003). Therefore, even if students do not attain their highest score in one sitting on their final SAT, their highest section scores from all of their test dates will be combined into one "super-score," providing a significant benefit to test re-takers.

General Discussion and Conclusions

In a culture that rewards performance on high-stakes admissions tests such as the SAT, it is important that we as a society understand the various factors that influence these test scores. When SAT scores impact both college admission decisions and scholarship outcomes (Patterson et al., 2012), it is important that all students have a fair chance to succeed on the SAT. Policy makers must attend to entrenched and enduring score gaps among different types of students (e.g., Aikens & Barbarin, 2008; Hannon, 2012; Mickelson, 2006). High school students, their families, teachers and administrators, need to understand what they can individually do to positively affect SAT scores.

In many cases, environmental differences and resource disparities circumscribe the preparatory opportunities available to different groups of students. Despite these differences, students can still make personal choices that have been shown to enhance SAT scores. By engaging in a more challenging high school curriculum and participating in extra-curricular activities, in and out of school (Brody & Benbow, 1990; Gerber, 1996), students can positively influence their SAT scores.

The greatest opportunity for students to enhance their test scores lies in their using effective preparatory strategies for the SAT. Researchers have provided compelling evidence that test preparation positively affects scores (e.g., Lilly & Montgomery, 2011). Students have an incredible array of preparatory choices to make regarding how they prepare, when they prepare, and how much time they invest in their preparation (e.g., Buchmann et al., 2010; Lilly & Montgomery, 2011). Researchers have found that increased contact hours greatly impact the efficacy of preparation, particularly on the math section of the SAT. By putting in more time on

task, students increase the likelihood that they will achieve score increases on the SAT (Messick & Jungeblut, 1981).

By pursuing preparatory formats that optimize score gains, such as smaller groups or individualized preparation, students could increase their likelihood of achieving significant score increases (e.g., Appelrouth et al., 2014; Ireson, 2004). When these preparatory options are available in school or in after-school programs, student participation should be encouraged.

By judiciously timing preparation so it leads up to the official testing administration, students can also increase their scores (e.g.,Turner, 2009). Likewise, students, parents, teachers and administrators need to understand that cramming is not an effective strategy to prepare for the SAT. Students need to distribute their practice over time to facilitate greater learning and retention (e.g., Cepeda et al., 2006). Students need to reinforce their learning through the use of practice tests, one of the most potent forces for enhancing learning (Rawson & Dunlosky, 2012). By increasing the number of SAT practice tests they take, students have a greater chance of optimizing their score gains on the SAT (Arnold & McDermott, 2013).

Students also benefit from taking multiple official SATs, as each testing administration acts as an additional practice/learning event (Appelrouth et al., 2014; Vigdor & Clotfelter, 2003). Additional SAT administrations can give students a greater degree of comfort with the official test and allow them to perform at a higher level in the high-stakes, stressed condition. Finally, students should be encouraged to practice with advanced calculators and become comfortable using them frequently on the math section of the SAT, as higher calculator usage correlates with higher math scores (Scheuneman et al., 2002).

If students, parents and teachers understand and attend to the relationship between motivation, self-efficacy, anxiety and SAT performance, this can create a beneficial academic

climate for improving test scores. Instructors can conduct interventions to help raise student's self-efficacy for the SAT, such as providing mastery experiences for students (Bandura, 1977), using test materials appropriately calibrated to students' current level of ability. In the domain of high stakes testing, as in other academic domains, if students develop higher levels of self-efficacy, this can lead to a reduction in anxiety, and enhanced motivation (Bandura, 1997). This in turn, can shift student perceptions of the SAT from being a threat, to being a challenge, which has been shown to enhance academic outcomes (Pajares, 1996). If students are led to believe that they have the ability to influence their SAT scores, this self-belief will lead to many positive behaviors and outcomes (Bandura, 1977; Dweck, 2006).

Future research could further explore the relation between self-belief and SAT performance and the relative effects of cognitive and non-cognitive variables on SAT performance. Another related line of inquiry involves studying whether targeted interventions to mitigate test anxiety in female and minority students could shrink the gender and racial testing gaps, building off Hannon's (2012) finding that test anxiety and goal orientation accounted for all of the significant gender differences on the SAT. Perhaps anxiety reduction could help level the playing field for many students.

We are only recently beginning to understand which factors contribute to performance on the SAT; a great deal remains to be discovered. It is through this current line of inquiry that we can begin to find the answers that will enable us to create highly effective preparatory programs, empowering more students to attain SAT score thresholds for admissions and scholarships. This in turn will allow us to take strides to reduce the SAT score gaps that persist across gender and racial/ethnic lines and create more equity in the domain of admissions testing.

References

ACT Inc. (2000). *Selections from the 2000 National Score Report*. Retrieved July 31, 2013, from http://www.act.org/newsroom/data/2000/t1.html

ACT Inc. (2013). *ACT Profile Report: National*. Retrieved November 13, 2014, from http://www.act.org/newsroom/data/2013/pdf/profile/National2013.pdf

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: the contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100*(2), 235–251.

Alon, S. (2010). Racial differences in test preparation strategies. *Social Forces (University Of North Carolina Press), 89*(2), 463-474.

Altermatt, E. R., & Kim, M. E. (2004). Getting girls de-stereotyped for SAT exams. *Education Digest: Essential Readings Condensed for Quick Review, 70*(1), 43-48.

Appelrouth, J., Zabrucky, K. M., & Moore, D. (2014). *Preparing Students for College Admissions Tests*. Manuscript submitted for publication.

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

Baker, M., & Johnston, P. (2010).The impact of socioeconomic status on high-stakes testing reexamined. *Journal of Instructional Psychology, 37*(3), 193-199.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*, 63-105.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review 84*(2): 191-215.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*(3), 586-598.

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

Bond, L. (2008). Teaching to the test: Coaching or corruption. *The New Educator, 4*(3), 216-223.

Briggs, D. C. (2005). Meta-Analysis: A case study. *Evaluation Review, 29*(2), 87-127.

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

Brody, L. E., & Benbow, C. P. (1990). Effects of high school coursework and time on SAT scores. *Journal of Educational Psychology, 82*(4), 866-875.

Buchmann, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces 89*(2), 435-482.

Burton, L. A., Henninger, D., & Hafetz, J. (2005). Gender differences in relations of mental rotation, verbal fluency, and sat scores to finger length ratios as hormonal indexes. *Developmental Neuropsychology, 28*(1), 493-505.

Byrnes, J. P., & Takahira, S. (1993). Explaining gender differences on SAT-math items. *Developmental Psychology, 29*(5), 805-810.

Byun, S., & Park, H. (2011). The academic success of East Asian American youth: the role of shadow education. *Sociology of Education, 85*(1), 40-60.

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.

Chemers, M. M., Hu, L., & Garcia, B. F. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology, 93*(1): 55-63.

Cohen, S. (2012). The three biggest lies in college admission. *Forbes* online version. Retrieved September 15, 2013, from http://www.forbes.com/sites/stevecohen/2012/09/29/the-three-biggest-lies-in-college-admission

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

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94,* 95–120.

College Board. (2007a). *A historical view of subgroup performance differences on the SAT reasoning test*. Retrieved November 15, 2013, from http://research.collegeboard.org/sites/default/files/publications/ 2012/7/researchreport-2006-5-historical-view-subgroup-performance-sat.pdf

College Board. (2007b). SAT percentile ranks: Critical reading, mathematics, & writing percentile ranks by gender and ethnic groups. Retrieved November 15, 2013, from http://www.collegeboard.com/prod_downloads/highered/ra/sat/SAT_CR_mathematics_writing_percentile_ranks_gender_ethnic_groups.pdf

College Board. (2012a). *Annual AP program participation 1956-2012*. Retrieved November 15, 2013, from http://media.collegeboard.com/digitalServices/pdf/research/2012_annual_participation.pdf

College Board. (2014). *2014 college-bound seniors-total group profile report*. Retrieved November 13, 2014, from https://secure-media.collegeboard.org/digitalServices/pdf/sat/TotalGroup-2014.pdf

Coyle, T. R., & Pillow, D. R. (2008). SAT and ACT predict college GPA after removing g. *Intelligence, 36*, 719–729.

Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*(6), 1639-1655.

Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology: General, 130*, 208–223.

Devine-Eller, A. (2012). Timing matters: Test preparation, race, and grade level. *Sociological Forum, 27*(2), 458-480.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84 (5),* 795-805.

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.

Duff, K., Callister, C., Dennett, K., & Tometich, D., (2012). Practice effects: a unique cognitive variable. *The Clinical Neuropsychologist, 26 (7),* 1117–1127.

Dweck, C. S. (2006). *Mindset.* New York, NY: Random House.

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

Elliot, J. A., & McGregor, H. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 74*, 628–644.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal Of Experimental Psychology: General, 128*(3), 309-331.

Everson, H. T., & Millsap, R. E. (2004). Beyond individual differences: exploring school effects on SAT scores. *Educational Psychologist, 39*(3), 157-172.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science, 15*, 373–378.

Gallagher, A. M. (1992). *Sex differences in problem-solving strategies used by high-scoring examinees on the SAT–M. Educational Testing Service Report No. 92-2*. Princeton, NJ: Educational Testing Service.

Gallagher, A. M., & De Lisi, R. (1994). Gender differences in scholastic aptitude test--mathematics problem solving among high-ability students. *Journal of Educational Psychology, 86*(2), 204-212.

Gerber, S. B. (1996). Extracurricular activities and academic achievement. *Journal of Research and Development in Education*, 30, 42–50.

Graesser, A. C., & Person, N. K. (1994). "Question asking during tutoring." *American Educational Research Journal, 31*(1), 104-137.

Greisemer, N. (2012). Nine key ways independent educational consultants support students and families. *Examiner.com* online magazine. Retrieved November, 15, 2013 from

http://www.examiner.com/article/9-key-ways-independent-educational-consultants-support-students-and-families

Grodsky, E. (2010). Learning in the shadows and in the light of day: A commentary on "Shadow Education, American Style: Test Preparation, the Sat and College Enrollment." *Social Forces, 89 (2),* 475-481.

Guida, F., Ludlow, L., & Wilson, M. (1985). The mediating effect of time-on-task on the academic anxiety/achievement interaction: A structural model. *Journal of Research & Development in Education, 19*(1), 21-26.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J., & Gernsbacher, M. (2007). The science of sex differences in science and math. *Psychological Science in the Public Interest (Wiley-Blackwell), 8*(1), 1-51.

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.

Hannon, B., & McNaughton-Cassill, M. (2011). SAT performance: Understanding the contributions of cognitive/learning and social/personality factors. *Applied Cognitive Psychology, 25,* 528-535.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*(1), 47-77.

Horn, J. M. (1989). Truth, gender, and the SAT. *Academic Questions, 3*(1), 35.

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

Izawa, C. (1968). Effects of reinforcement, neutral and test trials upon paired-associate acquisition and retention. *Psychological Reports, 23*, 947–959.

Johannessen, L. R., & Kahn, E. A. (2001). *How to prepare students for high-stakes tests (and still live with your conscience!).* Paper presented at the annual fall conference of the Illinois Association of Teachers of English Elmhurst, IL.

Kulik, J. A., Bangert-Drowns, R. L., & Kulik, C. C. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin, 95*(2), 179-188.

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.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19,* 528 –558.

Kohn, A. (2001). Two cheers for an end to the SAT. *Chronicle of Higher Education, 47*(26), B12.

Lane, K. L., Kalberg, J. R., Mofield, E., Wehby, J. H., & Parks, R. J. (2009). Preparing students for college entrance exams: Findings of a secondary intervention conducted within a three-tiered model of support. *Remedial and Special Education, 30*(3), 3-18.

Lemann, N. (2000). *The big test: the secret history of the American meritocracy*. New York, NY: Farrar, Straus and Giroux.

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.

Loewen, J. W. (1988). Gender bias in SAT items. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Marsh, H. W., & Kleitman, S. (2002). Extracurricular school activities: the good, the bad, and the nonlinear. *Harvard Educational Review, 72*(4), 464-514.

Mau, W., & Lynn, R. (2001). Gender differences on the Scholastic Aptitude Test, the American College Test and college grades. *Educational Psychology: An International Journal of Experimental Educational Psychology, 21*(2), 133-36.

Messick, S., & Jungeblut, A. (1981). Time and method in coaching for the SAT. *Psychological Bulletin, 89*(2), 191-216.

Meredith, M. (2004). Why do universities compete in the ratings game? An empirical analysis of the effects of the *U.S. News and World Report* college rankings. *Research in Higher Education, 45*(5), 443-461.

Mickelson, R. (2006). Segregation and the SAT. *Ohio State Law Journal, 67*(1), 157-199.

Mischo, C., & Haag, L. (2002). Expansion and effectiveness of private tutoring. *European Journal of Psychology of Education, 18*(3), 263-273.

Morgan D. L., & Michaelides, M. P. (2005). Setting cut scores for college placement. *College Board Research Report No 2005-9*, 1-12. Retrieved April 28, 2013 from https://professionals. collegeboard.com/profdownload/accuplacer-setting-cut-scores.pdf

Morse, R. (2013). Study: rankings affect student applications. *US News and World Report* online version. Retrieved November, 20, 2013 from http://www.usnews.com/education/blogs/college-rankings-blog/2013/10/03/study-rankings-affect-student-applications

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 (2012). *State of college admission.* Arlington, VA: NACAC. Retrieved July 31, 2013, from http://www.nacacnet.org/research/research-data/Documents/2012SOCA.pdf

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543-578.

Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory Into Practice, 41*(2) 116.

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. (1993). Coaching for the SAT: A summary of the summaries and an update. *Educational Measurement: Issues and Practice, 3*, 24-30.

Powers, D. E., & Rock, D. A. (1999). Effects of coaching on SAT 1: Reasoning Test Scores. *Journal of Educational Measurement, 36 (2),* 93-118.

Ramos, I., & Lambating, J. (1996). Gender differences in risk-taking behavior and their relationship to SAT-mathematics performance. *School Science & Mathematics, 96*(4), 202-207.

Rawson, K. A., & Dunlosky J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning*?*. *Educational Psychology Review, 24*(3):419-435.

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.

Reeve, C. L., & Lam, H. (2007). The relation between practice effects, test-taker characteristics and degree of g-saturation. *International Journal of Testing, 7*(2), 225–242.

Rigor Boosts SAT Performance. (2010). *Electronic Education Report, 17*(18), 2-3.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261-288.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20 –27.

Ryan, A., Ployhart, R. E., Greguras, G. J., & Schmit, M. J. (1998). Test preparation programs in selection contexts: self-selection and program effectiveness. *Personnel Psychology, 51,* 599-621.

Rubenstein, J. (2003). Test preparation: What makes it effective? In *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.

Scheuneman, J., Camara, W. J., Cascallar, A. S., Wendler, C., & Lawrence, I. (2002). Calculator access, use, and type in relation to performance on the SAT 1 reasoning test in mathematics. *Applied Measurement In Education, 15*(1), 95.

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation." *Reading and Writing Quarterly*: Overcoming Learning Difficulties 19(2): 159-173.

Scruggs, T. E., & Mastropieri, M. A. (1992). *Teaching Test-Taking Skills: Helping Students Show What They Know*. Brookline, MA: Brookline Books.

Seifert, E. H., & Beck Jr., J. J. (1984). Relationships between task time and learning gains in secondary schools. *Journal of Educational Research, 78*(1), 5.

Sigal, A. (2010). Racial differences in test preparation strategies. *Social Forces, 89 (2),* 463-474.

Soo-yong, B., Schofer, E. & Kyung-keun, K. (2012). Revisiting the role of cultural capital in East Asian educational systems: The case of South Korea. *Sociology of Education, 85*(3), 219–239.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28.

Speth, C., & Brown, R. (1990). Effects of college students' learning styles and gender on their test preparation strategies. *Applied Cognitive Psychology*, (4) 189-202.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.

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.

U.S. News (2012). America's best colleges 2012: National universities: Top schools. Retrieved November 20, 2013, from http://www.usnews.com/education/best-colleges/articles/2012/09/11/methodology-undergraduate-ranking-criteria-and-weights-2

Vigdor, J. L., & Clotfelter, C. T., (2003). Retaking the SAT. *The Journal of Human Resources, 38 (1),* 1-33.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership,4,* 119-30.

West-Faulcon, K. The river runs dry: When title VI trumps state anti-affirmative action laws. *University of Pennsylvania Law Review, 157 (4),* 1075-1160.

Witt, E. A. (1993, April). *Meta-Analysis and the effects of coaching for aptitude tests*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta GA.

Women in Academia Report (2013). *The gender gap in SAT scores*. (2013). Retrieved December 20, 2013, from http://ezproxy.gsu.edu/login?url=http://search.proquest.com.ezproxy.gsu.edu/docview/1436954484?accountid=11226

Wright, R. E., Palmer, J. C., & Miller, J. C. (1996). An examination of gender-based variations in the predictive ability of the SAT. *College Student Journal, 30*(1), 81-84.

Zwick, R. (2002). Is the SAT a 'Wealth Test'?. *Phi Delta Kappan, 84*(4), 307-311.

Zwick, R., & Green, J. G. (2007). New perspectives on the correlation of SAT scores, high school grades, and socioeconomic factors. *Journal of Educational Measurement, 44,* 23– 45.

2. PREDICTING SUCCESS IN PREPARING FOR HIGH-STAKES ADMISSIONS TESTS: A MODERATED MEDIATION ANALYSIS

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 alike. Our collective 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, 2012). As application numbers continue to surge at colleges and universities across the country (Cohen, 2012), 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; Morgan & Michaelides, 2005).

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 1,672,395 students who took the SAT in 2013 (College Board, 2014) engaged in formal test preparation offered in their high schools, through community organizations, or through commercial test preparation companies (Buchmann, Condron, & Roscigno, 2010; Lilly & Montgomery, 2011). These test preparation programs vary greatly in format, instructional method and efficacy. Some of these formal test preparation programs have drawn criticism for their lack of efficacy, their focus on "relentless drilling" and their lack of research-based methodology (Allensworth, Macarena & Ponisciak, 2008; Xie, 2013). There is a

clear need for a better understanding of how to craft more 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 (Briggs, 2005; Kulik, Bangert-Drowns et. al, 1984; Lilly & Montgomery, 2011; Powers, 1993); 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 their 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 a factor in the efficacy of SAT preparation programs, whereby math scores were significantly increased in programs with more than eight hours of instruction. Scruggs & Mastropieri (1992) and 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 into the factors that inform successful SAT preparation, Appelrouth, Zabrucky & Moore (2014) 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., 2014). Although this study provided insight into the isolated factors predicting successful SAT preparation, it did not establish a model of the most salient factors of successful test preparation and how they directly and indirectly affect SAT performance.

Factors Which May Impact Test Preparation Outcomes

The current investigation extends the research of Appelrouth et al., (2014), by examining direct and indirect influences as well as conditional direct and indirect influences on SAT scores. This model included the following variables:

Socioeconomic Status

Critics of the SAT frequently point to the correlation between wealth and SAT scores (r =.42; Sackett, Kuncel, Arneson, Cooper, & Waters, 2009) as a challenge to the construct validity of the SAT, potentially reducing the SAT to a mere "wealth test" (Kohn, 2001; Zwick, 2002). Affluent families use their superior resources to secure educational advantages for their children through after-school tutoring and test preparation activities deemed "shadow education" (Buchmann et al., 2010; Grodsky, 2010). Additionally, affluent families frequently have access to superior social capital (Coleman, 1988), which may inform how they approach processes such as college admissions and test preparation. In this analysis I explored whether socioeconomic status (SES) acts as a moderator, resulting in conditional direct and indirect influences on higher test scores.

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 their students for college admission tests, but the culture in each school 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). To explain differences in performance between public and private school students, researchers have proposed selection bias, differences in socioeconomic status and parental education and involvement, coupled with the ability of private schools to refuse admissions to particular students as contributing factors (i.e., Gamoran, 1996; Lubienski et al., 2008). In this analysis I investigated whether public or private school attendance has a moderating influence on direct and indirect contributions to higher SAT scores.

Gender

Despite their superior academic performance in the domains of high school and college, their more ambitious course-taking and superior self-discipline (Duckworth & Seligman, 2006), females consistently underperform males on the SAT (Hannon, 2012; Horn, 1989; Loewen, 1988; Nankervis, 2011). The College Board (2014) reported that in 2013, males had an average composite SAT score of 1510 (499 critical reading, 530 math, 481 writing) compared to 1486 for females (495 critical reading, 499 math, 492 writing). I explored whether gender exerts a moderating influence on the direct and indirect effects conducive to successful SAT preparation.

Timing of Test Preparation

Appropriate timing is one of the most fundamental components of successful test preparation (Devine-Eller, 2012; Turner, 2009). If students wait too long to begin SAT preparation, they will be at a significant disadvantage in terms of the number of potential test dates available to them before they must submit college applications and test scores. Additionally, starting late may force students to condense their sessions and effectively "cram" their preparation into a shorter period of time (Turner, 2009). The National Association for

College Admissions Counseling encourages students to begin SAT test preparation in 11th grade (National Association for College Admission Counseling, 2008). Devine-Eller (2012) investigated the timing effects of SAT preparation and found that high school seniors who are engaging in test preparation for the first time may have started too late. She posits that "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 such a meaningful role in determining score gains, I investigated the direct and indirect relationships between timing of SAT preparation and all other variables in the model.

Spaced Versus Distributed Study

Researchers have consistently found benefits of spacing, or distributing, study over time (Baldwin, & Ford, 1988). Two meta-analyses have demonstrated consistent and sizeable performance gains from the distribution of practice (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Donovan & Radosevich, 1999). When comparing three learning schedules—massed, clumped (i.e., clustered) and spaced—the spaced schedules, which had the greatest and most even distribution of lessons, resulted 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 led to a system of "successive relearning" which enhanced 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, 2014). In this analysis I further investigated the direct and indirect relations of distributed study to other variables influencing score gains on the SAT.

Time on Task

Academic gains have been found to correlate strongly with time on task (Seifert & Beck, 1984), and SAT researchers have likewise noted the positive effect of instructional time on SAT score increases (Messick & Jungeblut, 1981; Lilly & Montgomery, 2011). In this analysis I investigated the direct and indirect relationships between increased instructional time and other variables in the model.

Tutoring Format

Although many SAT researchers have ignored one-on-one tutoring and limited the scope of their investigations to the efficacy of group instruction, individualized instruction has been found to be related to larger academic gains in numerous domains (Bloom, 1984; Ireson, 2004). Tutoring is likely to be of much 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 into SAT increases, Applerouth et al., (2014) found that each hour of individual tutoring had a larger effect than a corresponding hour of group tutoring. I examined the direct and indirect relations between group and individual instruction and other variables of SAT preparation.

Homework

When researchers have investigated the effects of homework upon the academic outcomes of high school students, 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). Other researchers (Duckworth & Seligman, 2006; Xu, 2011) have found

gender effects, in which girls reported statistically significant higher scores in homework completion than did boys, potentially due to girls' higher levels of self-discipline. Appelrouth, et al. (2014) found homework completion correlated positively with SAT score increases, and in this investigation I further examined the direct and indirect relations of homework to other SAT preparation variables.

Practice Tests

Researchers have found a powerful effect of practice tests on memory, learning, and academic performance (Arnold & McDermott, 2013; Johannessen & Kahn, 2001; Kulik, Kulik, & Bangert-Drowns, 1984; Rawson, Dunlosky, & Sciartelli, 2013; Roediger & Butler, 2011). Forced retrieval during the act of testing strengthens the memory of learned content for future retrieval (Arnold & McDermitt, 2013) much more than repeated restudy (Roediger & Butler, 2011). More practice tests yield greater performance gains (Arnold & McDermott, 2013, Roediger & Butler, 2011), particularly on the SAT where practice tests have been shown to enhance performance for up to seven rounds of timed practice trials (Kulik, Kulik, et al., 1984). 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, Kulik, et al., 1984, p. 444). In this analysis I investigated the direct and indirect relations between practice SATs and other variables of SAT preparation.

Number of Official Tests

The 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 from their original sample of 150,377 students, of the 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 I 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. (2014) 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-

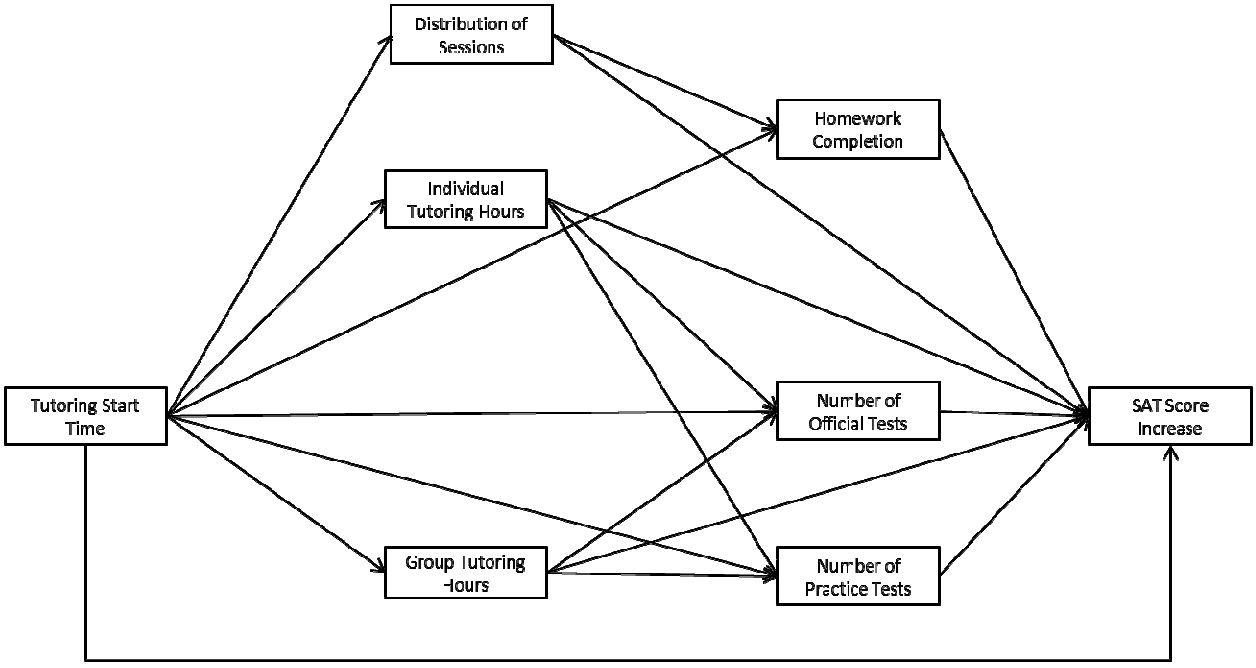
mediation analyses to understand the mechanisms by which antecedent academic success factors influence outcomes (Arnold & McDermott, 2013; Carroll, 1963; Marjoribanks, 2003; Walberg, 1981). Selected variables in this analysis include time on task, practice and official tests, homework, and distribution 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 increase

To understand the relationships between the various factors involved in SAT preparation, I 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), I inferred that *when* a student commences SAT preparation bears heavily upon direct and indirect influences. If a student delays SAT preparation until late in the student's junior year, this will not only constrain the number of practice and official tests available to that student before having to submit college applications and final test scores, but will also impact the number of tutoring hours and the spacing of those hours. I have hypothesized that the longer a student waits to begin preparation, the more condensed will be the distribution of tutoring sessions. Because so many effects appear to stem from tutoring start time, it is the initiating, or exogenous, variable of our theoretical model.

From tutoring start time, the model moves sequentially to three endogenous variables: distribution of sessions, individual tutoring hours and group tutoring hours. Researchers have found benefits of distributed sessions (Cepeda et. al, 2006; Turner, 2009), and I hypothesize that students who start preparation later in their academic careers will be forced to condense their instructional sessions into shorter intervals. I hypothesize that by reducing the distribution of sessions, this may affect homework completion, a factor in academic performance (Cooper et. al, 1998; Cooper et. al, 2006). Students forced to "cram" their study into a more condensed time period may be unable to complete all of their assigned homework. As a result, session distribution may affect homework completion, which will, in turn, affect final SAT score increase.

Additionally I hypothesized that starting tutoring later may lead to a net reduction of instructional time in both group and individual tutoring hours. As researchers have found

instructional time relates to SAT score gains (Messick & Jungeblut, 1981; Lilly & Montgomery, 2011), the drop in hours would directly depress the score increases.

I hypothesized that any changes to instructional time would also affect the number of practice and official SATs, both of which may ultimately affect SAT score increase. Students are encouraged to take practice tests as part of their test preparation, to reinforce concepts learned during instruction time and provide feedback regarding levels of attained mastery. Researchers have found direct effects of practice SAT tests on score increases (Kulik, Kulik, et. al, 1984). When students take fewer hours of SAT instruction, I hypothesize this will lead to a decreased number of practice tests, thereby depressing score increases.

In similar fashion, as students start tutoring later and take fewer hours of instruction, I hypothesized that they will take fewer official SATs. I hypothesized that the number of official SATs students take will be directly affected by how early or late they begin preparation: a student starting to prepare for the SAT earlier in their junior year would theoretically take more tutoring hours and engage in more practice and official tests than a student starting to prepare during the senior year, due to time constraints and application deadlines. As depicted in the model, the number of official and practice tests will directly impact the ultimate SAT score increase.

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 students (lower SES)? Direct and indirect influences and conditional effects are examined to help clarify the relationships between the various factors of SAT preparation.

Methodology

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. They all had "baseline" SAT scores prior to beginning any preparation activities with the center. These students either reported officially administered SAT scores or were administered a baseline SAT at the center. Participants also engaged in individual or group SAT preparation with the center, but did *not* prepare for the ACT exam. Finally, following test preparation, they reported "post" intervention scores from officially administered SATs.

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. One thousand and one hundred and thirty students (58.46%) had not

previously taken the official SAT and, thus, 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 that closely approximated 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 were coded as a dichotomous variable (0 = official, 1 = unofficial).

School Type: High school attendance data were collected from students or their families in the 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 were 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 were 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 the 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, 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 or senior years. 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 session, 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. As an example, a student who completed four sessions within 30 days would have a distributed session score of 7.50, while a student who completed four sessions within 60 days would have a score of 15.00. Due to extreme outliers 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, Mattern, & Swerdzewski, 2012). This involves adding the highest section scores (critical reading, math and writing) from various SAT administrations to create a single composite score. For example, if a student scored a 650 on the critical reading section of the SAT in January, a 700 on the math section in March, and a 730 on the writing section in May, and these three section scores were the highest attained during the three official administrations, the student's "super-scored" SAT would be 650 + 700 + 730 = 2080. 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.

Procedures and data analyses

The data used in the current investigation were the same data collected by Appelrouth, et al. (2014), 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.

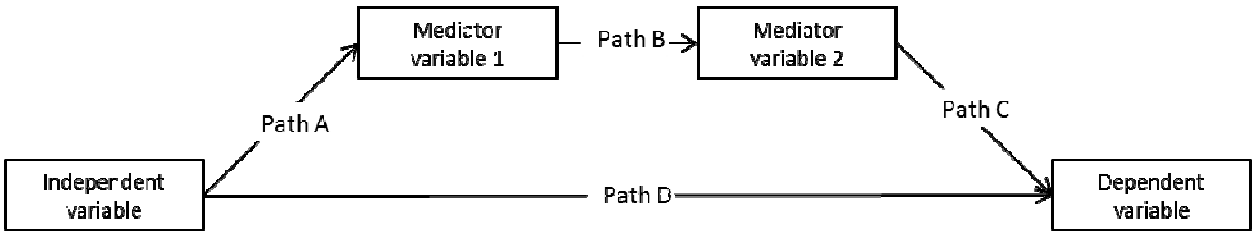
Analyses of direct and indirect effects

All statistical analyses were conducted using the IBM SPSS Statistics software; results are considered to be significant if the p-value in a given analysis is equal to or lower than 0.05. A series of regression analyses was conducted to estimate the direct effects in the theoretical

model. To be conservative, each direct effect was estimated, controlling for other variables in the model.

Several indirect, mediated, influences on SAT increase are depicted in the model.

Figure 2. Mediation model



There are multiple approaches available for testing indirect effects. Bootstrapping resampling procedures, which estimate a confidence interval for the indirect effects and function independently of the distribution of the product, have become a best practice (Preacher et. al, 2007), although with large sample sizes normal theory approaches to testing the significance of the indirect effect (e.g., the Sobel test), yield essentially the same results. In this analysis I used the bootstrapping procedure, available in PROCESS, to test the significance of the indirect effects.

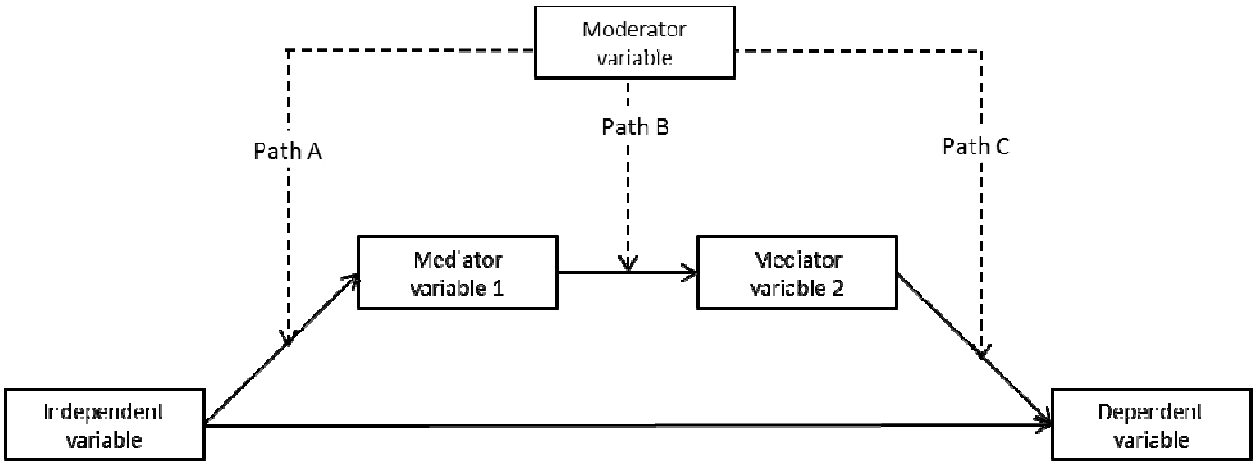
To illustrate the statistical procedures used to measure the significance of indirect relationships within the model, let's analyze one example of 3-path mediation, as illustrated in Figure 1. In the model, the independent variable (IV) is tutoring start time and the dependent variable (DV) is SAT score increase. To determine whether a stage-1 mediator, individual tutoring hours (M1), and a stage-2 mediator, number of official SATs (M2), mediate the indirect relationship between the IV and DV, I estimated the paths between these variables from three regression models. I estimated the initial pathway (a) between the IV and M1, the secondary pathway (b) between M1 and M2, and the tertiary pathway (c) between M2 and the DV, testing

each regression coefficient for statistical significance. The indirect effect is the product of the a, b, and c paths. Bootstrapped (N=1000) estimates of the product were obtained and a 95 percent confidence interval was established. Confidence intervals that do not include zero indicate that the indirect effect is statistically significant.

Conditional effects

Conditional effects, or moderated effects, were examined for selected variables—gender, socioeconomic status, school type, and baseline type— as moderators of the direct effects within the model. In examining the potential for moderated effects, I examined whether the direct effect of a predictor on an outcome was contingent upon the values of the moderator.

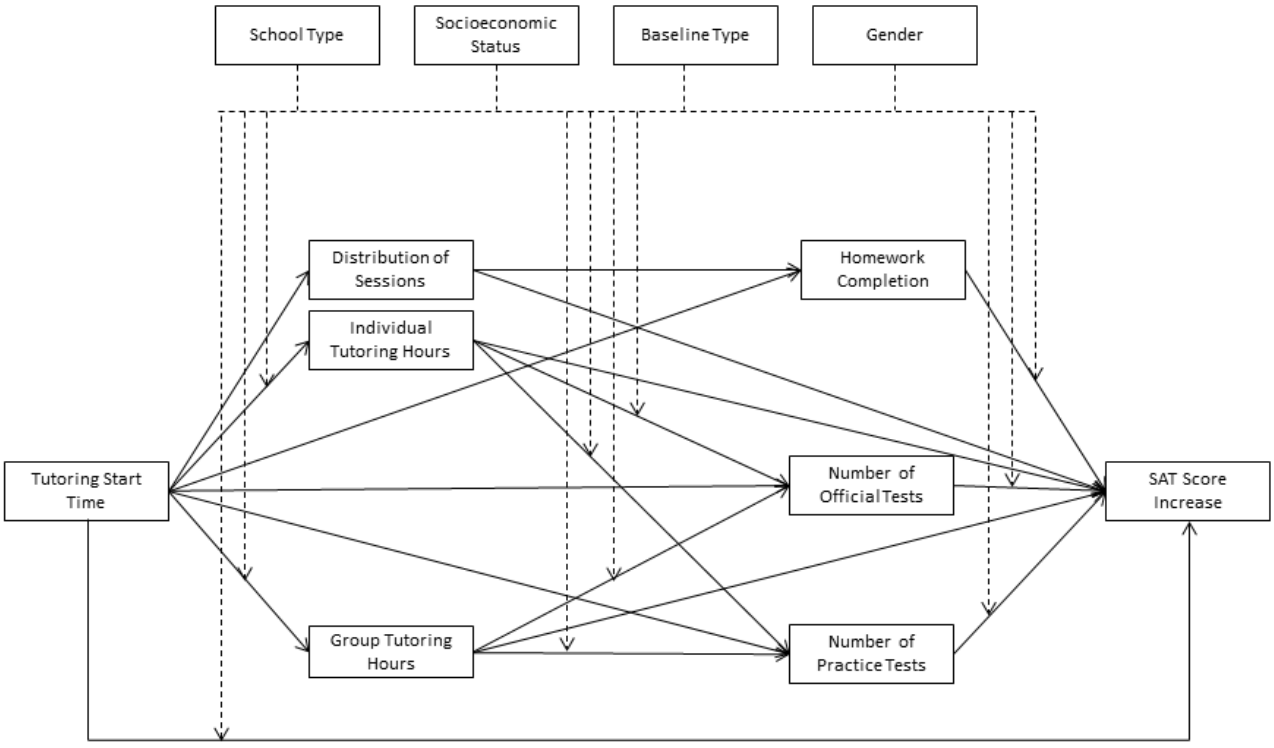
Figure 3.Moderated mediation model



To calculate conditional indirect effects, in which indirect, or mediated effects, were tested at different values of the moderator, a normal theory approach was used. This approach was used due to limitations of the PROCESS Macro for the current model. Using a Sobel test, a product of coefficients procedure, significance was measured using large sample t values, or z-values. Although there is a risk of bias using the Sobel test, due to the potential of asymmetrical

distributions with smaller samples, given our relatively large sample, this was not an issue. The model with proposed moderators is presented in Figure 4.

Figure 4. Theoretical model of direct, indirect and conditional effects



Results

Descriptive Statistics

Table 1 shows the means, standard deviations, ranges, skew and kurtosis for each of the measures.

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

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 far 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 2 shows the means and standard deviations for the SAT increases achieved by groups of students categorized by gender, school type, and baseline type.

Table 2. Means and standard deviations for SAT score increases

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

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.

Direct relationships within the model

I examined the relationships between different variables within the model, looking first at the direct relationships, which are found in Table 3.

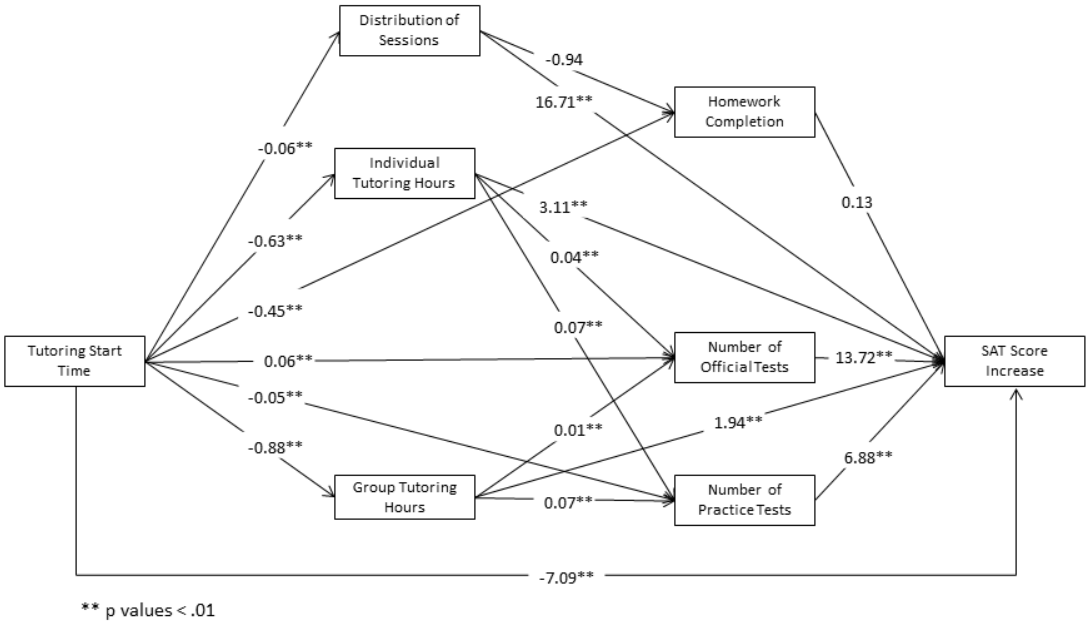
Table 3. Analysis of direct effects within the model

Path	*B*	*SE*
Start time to distribution of sessions	-0.06**	<.01
Start time to individual tutoring hours	-0.63**	0.05
Start time to homework completion	-0.45**	0.15
Start time to number of official tests	0.06**	0.01
Start time to number of practice tests	-0.05**	0.01
Start time to group tutoring hours	-0.88**	0.05
Start time to SAT score increase	-7.09**	0.67
Distribution of sessions to homework completion	-0.94	0.85
Distribution of sessions to SAT score increase	16.71**	3.72
Individual tutoring hours to number of official tests	0.04**	<.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

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

The model begins with tutoring start time as the independent variable (IV) and concludes with SAT score increase as the dependent variable (DV). Between the IV and DV there are direct and indirect relations. Significant direct relations were reported between tutoring start time and the following variables in the model: session distribution, individual tutoring hours, group tutoring hours, homework completion, number of official tests, number of practice tests and total SAT increase. All direct effects in the model are depicted in Figure 5.

Figure 5: Model with direct effects estimated (unstandardized regression coefficients)



Starting tutoring earlier junior year yielded many positive effects. As I established June of Junior year as the zero point for the variable start time, for every month prior to June that students began SAT preparation, they had tutoring sessions that were more distributed by .06 days, completed .63 additional hours of individual tutoring, .88 additional hours of group tutoring, .45% more of their assigned homework, took .05 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, 3.11 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.

Indirect relationships within the model

Multiple indirect effects between start time and score increase are depicted in the model (see Figure 5). 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.

Table 4. Analysis of indirect effects within the model

Two-path mediation

Variable Sequence	*B*	*SE*	Bootstrap 95% CI
Start time (A) to Distribution of sessions (B) to SAT score increase (C)	-1.02	0.24	[-1.53, -0.59]*
Start time (A) to Homework completion (B) to SAT score increase (C)	-0.06	0.05	[-.20, .02]
Start time (A) to Individual tutoring hours (B) to SAT score increase (C)	-1.70	0.26	[-2.26, -1.23]*
Start time (A) to Group tutoring hours (B) to SAT score increase (C)	-1.28	0.21	[-1.72, -0.88]*
Start time (A) to Number of official tests (B) to SAT score increase (C)	0.83	0.18	[0.49, 1.22]*
Start time (A) to Number of practice tests (B) to SAT score increase (C)	-0.32	0.10	[-0.58, -0.15]*

Three-path mediation

Variable Sequence	*B*	*SE*	Bootstrap 95% CI
Start time (A) to Distribution of sessions (B) to Homework completion (C) to SAT score increase (D)	0.01	0.01	[-.01, .05]
Start time (A) to Individual tutoring hours (B) Number of official tests (C) to SAT score increase (D)	-0.21	0.06	[-0.33, -0.12]*
Start time (A) to Individual tutoring hours (B) Number of practice tests (C) to SAT score increase (D)	-0.36	0.10	[-0.57, -0.18]*
Start time (A) to Group tutoring hours (B) Number of official tests (C) to SAT score increase (D)	-0.08	0.03	[-0.15, -0.04]*
Start time (A) to Group tutoring hours (B) Number of practice tests (C) to SAT score increase (D)	-0.43	0.11	[-0.66, -0.23]*

* *indicates confidence interval does not include zero and is significant at p <.05*

2-path mediation

The model contains six instances of 2-path mediation in which a single variable mediates the relationship between start time and SAT score increase; five of these are statistically significant. The results from the 2-path mediation analysis are provided in Table 4.

Session distribution mediated the relationship between start time and SAT score increase. To interpret this finding, as students have higher values on start time, meaning that they started preparing for the SAT later in their junior year, their SAT scores decreased indirectly as a function of having SAT sessions that were less distributed. Session distribution therefore mediated the relationship between SAT start time and SAT score increase. In addition to the SAT increase directly attributed to starting earlier in the year, an additional increase was attributed to the greater distribution of sessions that resulted from having more time to prepare. These "total" effects are the sum of the direct effects and all indirect effects (Hayes, 2013).

Individual and group tutoring hours mediated the relationship between start time and score increase. Thus, additional increases in SAT scores were associated with earlier start times due to the indirect effects through individual tutoring (1.70) and group tutoring (1.28).

Similarly, official and practice tests mediated the relationship between start time and score increase. Additional SAT increases were positively associated with earlier start times due to the indirect effects through practice tests (0.32); however, SAT increases were negatively associated with earlier start times due to the indirect effects through official tests (-0.83). As a reminder, in this analysis, positive coefficients indicate a smaller increase in SAT scores and negative coefficients indicate a larger increase in SAT scores, due to the influence of start time. Thus the positive *B* value of .83 from table 4 translates to a decrease of .83 SAT points.

3-path mediation

I analyzed the indirect effects in the model, examining 3-path mediation. The model has five instances of 3-path mediation. Upon analysis utilizing the PROCESS SPSS macro, four of the five pathways yielded bootstrap confidence intervals that did not contain zero; in other words, with a 95% confidence these pathways revealed significant mediation of the effects of

tutoring start time on score increase. The results from the 3-path mediation analysis are provided in Table 4.

Individual tutoring hours and official SAT tests mediated the relationship between start time and score increase. Increases in SAT scores were associated with earlier start times due to the indirect effects through individual tutoring and official SATs (.21). Likewise, individual tutoring hours and practice SATs mediated the relationship between start time and score increase. Increases in SAT scores were associated with earlier start times due to the indirect effects through individual tutoring and practice SATs (.36).

Group tutoring hours and official SAT tests mediated the relationship between start time and score increase. Increases in SAT scores were associated with earlier start times due to the indirect effects through group tutoring and official SATs (.08). Likewise, group tutoring hours and practice SATs mediated the relationship between start time and score increase. Increases in SAT scores were associated with earlier start times due to the indirect effects through group tutoring and practice SATs (.43).

Conditional direct relationships within the model

Table 5. Analysis of conditional direct effects within the model

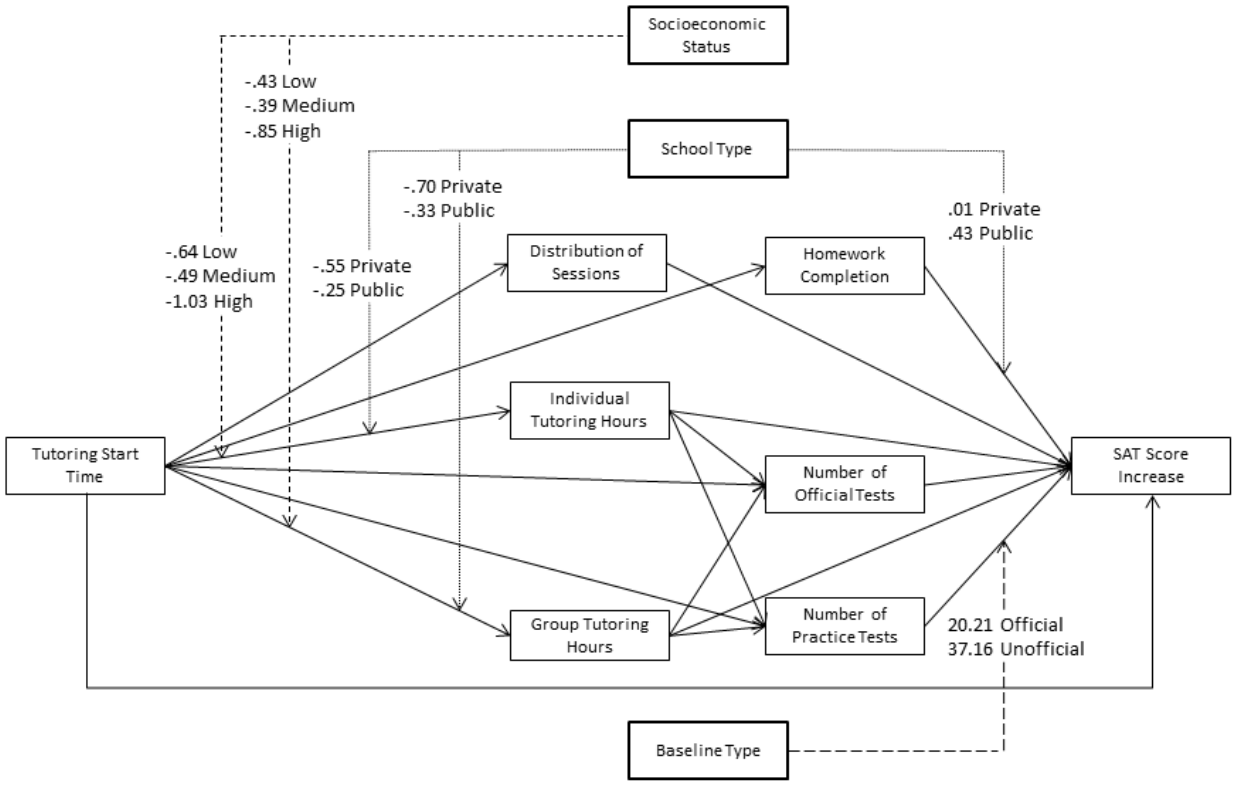
Path	Interaction		Conditional Estimates		
	B	*SE*	Category	*B*	*SE*
School type moderating the path between start time and individual hours	.30**	0.08	Private	-0.55**	0.05
			Public	-0.25**	0.06
School type moderating the path between start time and group hours	.37**	0.09	Private	-0.70**	0.06
			Public	-0.33**	0.08
School type moderating the path between homework completion and score increase	.43*	0.21	Private	0.01	0.12
			Public	0.43*	0.17
Socioeconomic status moderating the path between start time and individual hours	-0.04	0.09	Low	-0.43**	0.08
			Medium[a]	-0.39**	0.06
	-0.46**	0.09	High	-0.85**	0.08
Socioeconomic status moderating the path between start time and group hours	-0.15	0.11	Low	-0.64**	0.09
			Medium	-0.49**	0.07
	-0.54**	0.11	High	-1.03**	0.09
Baseline type moderating the path between official tests and score increase	16.95**	5.94	Official	20.21**	4.79
			Unofficial	37.16**	3.98

[a] Medium served as the baseline or reference condition

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

In order to examine the moderating role of socioeconomic status, school type, baseline type and gender, regression analyses with interaction terms were performed using the PROCESS macro (Hayes & Preacher 2012) and the general linear model in SPSS. Table 5 shows the results of the tests of the interactions for statistically significant moderators, and the conditional coefficients, also known as simple slopes. Figure 6 displays the conditional effects in the context of the model.

Figure 6: Model with conditional values



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 tutoring hours. The coefficients from Table 5 indicate that compared to private school students (B = -0.55, p < .01), public school students show a smaller effect (B = -.25, p < .01) 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 (.55 hours) than it did for their public school counterparts (.25 hours). The interaction coefficient (B = .30, p < .01) is the difference between the two values (.i.e., 0.55- 0.25 = .30), and the significance test revealed that these values were significantly different.

The moderating effect of school type also held for the relationship between tutoring start time and the number of group tutoring hours. Compared to private school students (B = -0.70, p < .01), public school students showed a smaller effect (B = -0.33, p < .01) of start time on group tutoring hours. When private school students started tutoring a month earlier in the school year, this translated to a larger increase (an additional .70 hours) in group tutoring hours than for their public school counterparts who only added an extra .33 hours. This analysis revealed that starting earlier led to a greater increase in overall tutoring hours for students in private school than for those in public school.

School type likewise exerted a significant moderating effect upon the relationship between homework completion and score increase. Compared to private school students (B = 0.01, p > .05), public school students showed a greater effect (B = 0.43, p < .05), of homework completion on score increase. Homework had a stronger effect on score increases for students in public school than in private school.

Socioeconomic status (SES) moderated the relationship between start time and both individual and group tutoring hours. Earlier start times had a stronger effect on the number of individual tutoring hours for high SES students (B = -0.85, p < .01) than for middle (B = -0.39, p < .01) or low SES (B = -0.43, p < .01) students. Similarly, earlier start times had a stronger effect on the number of group tutoring hours for high SES students (B = -1.03, p < .01) than for middle (B = -0.49, p < .01) or low (B = -0.64, p < .01) SES students. When more affluent students started SAT preparation earlier, this led to a greater increase in the number of tutoring hours they completed than for their less affluent peers.

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 37.16 points ($B = 37.16$, $p < .01$) for each official SAT they took. In contrast, students who began test preparation after they had already taken an official SAT only gained 20.21 SAT points ($B = 20.21$, $p < .01$) for each subsequent official SAT they took. 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.

Conditional indirect relationships within the model

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	-1.71**	0.24
	Public	-.78**	0.20
Start time to group hours to SAT increase	Private	-1.36**	0.23
	Public	-.64**	0.18
Start time to homework completion to SAT increase	Private	-0.0045	0.05
	Public	-0.19	0.10
Moderator: SES			
Start time to individual hours to SAT increase	Low	-1.15**	0.22
	Medium	-1.61**	0.21
	High	-2.11**	0.29
Start time to group hours to SAT increase.	Low	-.72**	0.16
	Medium	-1.01**	0.16
	High	-1.32**	0.22
Moderator: Baseline type			
Start time to homework completion to SAT increase	Official	1.21**	0.35
	Unofficial	2.23**	0.44

*p < .05; **p < .01

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, giving us moderated mediation.

School type had a moderating effect upon the indirect effect from start time to individual hours to SAT increases. Private school students gained more points by starting earlier in the year than their public school counterparts. By starting SAT preparation a month earlier in the junior year, private school students gained 1.71 points ($p < .01$) through increased individual tutoring, compared to public school students who only gained .78 points ($p < .01$).

School type had a similar conditional effect upon the indirect effect from start time to group hours to SAT increases. By starting tutoring a month earlier in their junior year, private school students gained 1.36 points through increased group tutoring, compared to a gain of only .64 points for public school students.

School type exerted a moderating influence on the indirect effect from start time to homework completion to SAT increase. By starting tutoring a month earlier in their junior year, public school students gained .19 ($p > .05$) points through indirect effects of homework completion, compared to a negligible gain of only .0045 ($p > .05$) for private school students.

SES exerted a moderating influence on the indirect effects from start time to individual and group hours to SAT increases. By starting tutoring a month earlier in their junior year, high SES students gained more points through the indirect effects of increased individual tutoring hours ($B = 2.11$, $p < .01$) than middle ($B = 1.61$, $p < .01$) or low ($B = 1.15$, $p < .01$) SES students. Likewise, starting a month earlier yielded high SES students a larger score gain ($B = 1.32$, $p < .01$) through the indirect effects of increased group tutoring hours than it did their middle SES (B

= 1.01, $p < .01$) or low SES ($B = .72$, $p < .01$) peers. Starting earlier conferred greater benefits on more affluent students through the indirect effects of increased tutoring hours.

Baseline type exerted a moderating influence on the indirect effect from start time to homework completion to SAT increases. By initiating tutoring a month earlier in their junior year, students who took an unofficial SAT baseline had greater gains ($B = 2.23$, $p < .01$) through the indirect effect of homework completion on score increases than did students who took an official SAT baseline test ($B = 1.21$, $p < .01$).

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	-.30**	0.07
	Public	-.14**	0.04
Start time to individual hours to practice tests to SAT increase	Private	-.26**	0.07
	Public	-.12**	0.04
Start time to group hours to official tests to SAT increase	Private	-.10**	0.02
	Public	-.05**	0.02
Start time to group hours to practice tests to SAT increase	Private	-.34**	0.09
	Public	-.16**	0.05
Start time to session distribution to HW completion to SAT increase	Private	0.001	0.01
	Public	0.02	0.02
Moderator: SES			
Start time to individual hours to official tests to SAT increase	Low	-.20**	0.05
	Medium	-.29**	0.06
	High	-.37**	0.08
Start time to individual hours to practice tests to SAT increase	Low	-.18**	0.05
	Medium	-.25**	0.06
	High	-.33**	0.09
Start time to group hours to official tests to SAT increase	Low	-.07**	0.02
	Medium	-.09**	0.02
	High	-.11**	0.03
Start time to group hours to practice tests to SAT increase	Low	-.26**	0.07
	Medium	-.32**	0.08
	High	-.39**	0.10
Moderator: Baseline type			
Start time to individual hours to official tests to SAT increase	Official	-.44**	0.11
	Unofficial	-.81**	0.11
Start time to individual hours to practice tests to SAT increase	Official	-.18**	0.05
	Unofficial	-.33**	0.05

*p < .05; **p < .01

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

School type had a moderating effect upon 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 .30 points ($p < .01$) through increased individual tutoring and increased official SAT tests. In comparison, public school students only gained .14 points ($p < .01$) through this pathway.

The moderating effect of school type influenced four additional 3-path relations within the model. By starting SAT preparation a month earlier in the junior year, private school students gained .26 points ($p < .01$) through increased individual tutoring and practice tests compared to public school peers who only gained .12 points ($p < .01$). Starting preparation a month earlier yielded private school students a gain of .10 points ($p < .01$) through increased group tutoring and official tests compared to public school peers who only gained .05 points ($p < .01$). It also yielded private school students a gain of .34 points ($p < .01$) through increased group tutoring and practice tests compared to public school peers who only gained .16 points ($p < .01$). The three path models confirmed the findings from the two-path models. Starting test preparation earlier generally conferred greater benefits upon private school students.

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 .37 points ($p < .01$) through increased individual tutoring and official tests compared to medium SES peers who gained .29 points ($p < .01$) and low SES students who gained only .20 points ($p < .01$). Starting preparation a month earlier yielded high SES students a gain of .33 points ($p < .01$) through increased individual tutoring and practice tests compared to medium SES peers who gained .25 points ($p < .01$) and low SES students who only gained .18 points ($p < .01$). It also yielded high SES students a gain of .11 points ($p < .01$) through increased group tutoring and official tests

compared to medium SES peers who gained .09 points ($p < .01$) and low SES students who only gained .07 points ($p < .01$). Finally, the effect of starting a month earlier yielded high SES students a gain of .39 points ($p < .01$) through increased group tutoring and practice tests compared to medium SES peers who gained .32 points ($p < .01$) and low SES students who only gained .26 points ($p < .01$). The three path models confirmed the findings from the two-path models: starting earlier generally conferred greater benefits upon more affluent students.

Baseline type exerted a moderating influence on the indirect effect from start time to individual hours to official tests to SAT increases. By initiating tutoring a month earlier in their junior year, students who took an unofficial SAT baseline had greater gains ($B = 0.81$, $p < .01$) through the indirect effect of increased individual hours and official tests on SAT increase than did students who took an official SAT baseline test ($B = 0.44$, $p < .01$). Similarly, starting tutoring a month earlier in their junior year yielded students who took an unofficial SAT baseline a greater gain ($B = 0.33$, $p < .01$) through the indirect effect of increased individual tutoring hours and practice tests than did students who took an official SAT baseline test ($B = 0.18$, $p < .01$). Students who took a preliminary official SAT derived smaller gains through tutoring and testing. Other student differences not accounted for in our analyses may explain some of these effects.

Discussion

The principal goal of this study was to examine the relationships between factors that predict successful preparation for the SAT. Using archival data collected from a commercial SAT preparation service, in the present study I extend prior research by Appelrouth et al. (2014) by testing mediators of intervention effects and by examining whether mediation is moderated by student characteristics. This investigation provides the first moderated-mediation statistical

analysis of factors involved in high-stakes admissions testing, extending our understanding of the relationships between salient factors of a successful SAT preparation program.

Researchers have previously examined the effects of components of SAT preparation in isolation. We have ample evidence that practice tests are helpful (e.g., Kulik, Bangert-Drowns et al., 1984), that contact time contributes to SAT score increases (e.g., Lilly & Montgomery, 2011), and that the timing of SAT preparation is a meaningful predictor of performance on the SAT (Devine-Eller, 2012; Turner, 2009). But we do not have a working model of how these factors operate in concert. This research serves as an important first step toward empirically examining the relationship among the various factors, which 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. The model also examines how student characteristics such as gender, school type, and SES influence test-preparation variables.

Researchers such as Turner (2009) and Devine-Eller (2012) have identified proper timing of test preparation as a key factor to successful outcomes. Devine-Eller found that the most academically competitive students complete the majority of their SAT preparation in their junior year. Similarly, Appelrouth et al., (2014) found that students who begin SAT preparation earlier in their junior year attain greater SAT score increases than those who wait until later in the junior year to begin preparation. These findings supported the decision in the current study to establish the start time of SAT preparation as the endogenous variable of the path model.

Results from the present study reveal numerous intervening variables that help account for the relationship between SAT start time and final SAT score increases. The results of mediation

analyses in the current study provided 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.

Direct Effects

Starting tutoring earlier junior year leads 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 direct 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, .06 of a practice test gained per month, that it lacks practical significance.

Starting preparation earlier had significant positive direct effects upon SAT scores as did the following variables: session distribution, individual and group tutoring hours, practice and official tests. Group and individual tutoring hours positively predicted both practice and official tests.

Indirect Effects

Multiple indirect effects between start time and score increase as well as significant 2-path mediation relations were found in the model. Session distribution mediates the relationship between start time and SAT score increase as do individual and group tutoring hours and official and practice tests. What appears to be an SAT increase attributed to starting earlier in the year must be partially attributed to the greater distribution of sessions, more contact hours, and more practice tests.

Similarly, official and practice tests independently mediated the relationship between start time and score increase. An unexpected finding was that starting tutoring earlier led to fewer official tests, whereas starting tutoring earlier led to an increase in the number of practice tests. The effect of each official test on SAT scores is nearly double that of each practice test. Therefore an earlier start led to SAT score gains mediated by an increase in practice tests, but these gains have been offset by a paradoxical decrease in official SATs. This effect warrants further investigation.

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. Some of the benefits of beginning tutoring earlier in a student's junior year can be attributed to a combination of increased individual tutoring hours coupled with increased practice tests. This effect also applies for 3-path mediation involving group tutoring and practice and official tests, although the effect of group hours was smaller than that of individual tutoring hours.

Conditional Effects

In order to examine the moderating role of socioeconomic status, school type, baseline type and gender, regression analyses with interaction terms were conducted. 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 both 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. School type likewise 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, although the effect size is so small it has no practical significance.

Socioeconomic status (SES) moderated the relationship between start time and both individual and group tutoring hours. When more affluent students started SAT preparation earlier, this will led to a greater increase in the number of tutoring hours they completed than for their less affluent peers.

Baseline type is 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 37.16 points for each official SAT they took, compared to students with an official SAT baseline who gained only 20.21 SAT points. It is quite possible that 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.

Conditional Indirect Effects

Combining the effects of mediation and moderation led to the conditional indirect effect (moderated mediation), which reflected the dependency of mediation on the levels of student characteristics. In examining the 2-path moderated mediation relations, school type has a moderating effect upon the indirect effect from start time to individual hours to SAT increases. Private school students gained more points by starting earlier in the year than their public school counterparts. School type had a similar conditional effect upon the indirect effect from start time to group hours to SAT increases. School type exerted a moderating influence on the indirect effect from start time to homework completion to SAT increases.

SES exerted a moderating influence on indirect effects from start time to individual and group hours to SAT increases. By starting tutoring earlier, high SES students gained more points through the indirect effects of increased individual and group tutoring hours than did medium and low SES students. Starting earlier conferred greater benefits on more affluent students through the indirect effects of increased tutoring hours.

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 on score increase than did students who took an official SAT baseline test. This unexplained finding warrants further investigation.

In examining 3-path moderated 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 which private school students saw larger gains from earlier start times. Private school 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. Finally, earlier start times yielded private school students a small, but statistically insignificant increase over public school students through increased session distribution and homework completion. The three path models confirmed 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 confered greater benefits upon more affluent students.

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. A similar effect existed in that students who took an unofficial SAT baseline had greater score gains through the indirect effect of increased individual hours and practice 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 derive smaller gains through tutoring and testing.

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 from a regression analysis to a moderated 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 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, other factors of preparation and student characteristics. Thus it is necessary to explore further the

model considered 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.

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 has important implications, for it helps clarify which aspects of tutoring intervention are the core elements and which best predict higher test scores when isolating other factors. By drawing attention to the core elements, this can allow administrators and individual students to focus on the effective components of tutoring intervention and minimize focus on those aspects having smaller effects once mediation is considered. It is also meaningful to understand the effect of moderation, to know that different mechanisms might be operating in different ways for students with specific characteristics. This could lead to a better understanding of how to best prepare certain student populations for success on the SAT.

Based on mediation outcomes, there seem to be numerous ways to enhance scores. One could target specific aspects of the intervention such as building in more practice effects, more

contact hours, and more distributed sessions. One could also approach the attainment of higher scores indirectly through earlier start times, which correlate with more distributed sessions, more contact hours, and more practice effects. The mediation analyses reveal that what appears to be an SAT increase attributed to starting earlier in the year must be partially attributed to the greater distribution of sessions, increased contact hours and practice tests. Considering the partial mediation of the 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.

This investigation contributes to our understanding of how student characteristics such as gender and SES and school type influence SAT preparation. This information could be useful for educators seeking to craft a successful SAT preparation program for a particular population of students. Being aware that student differences influence the effectiveness of interventions can help administrators construct more well-designed tutoring programs.

As conditional effects were not found for gender—females and males responded in a similar fashion to the tutoring intervention— administrators would not need to 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 that the moderating influence of SES mirrors in many ways that of school type. By starting SAT preparation earlier high SES students see larger SAT score gains through increased individual tutoring, group tutoring, official and practice tests, and through those factors in combination. Starting earlier confers greater benefits on more affluent students through indirect effects. To help balance this, administrators working with students from lower SES groups 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 are meaningful differences between students who initiate test preparation before sitting for an official test—proactive preparation—and those who initiate test preparation after taking an official test—reactive preparation. Those who start preparing before taking an official test should plan on building in more practice effects to help them gain the benefits of more exposure to testing conditions.

Findings from this study have broad implications for test preparation programs, high school teachers, and college-bound students. Beyond the benefits to educators hoping to craft successful SAT coaching programs, findings from this study could have important implications for individual high school students aspiring to optimize their own SAT scores for college scholarships and admissions. 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. If students can understand the model for successful SAT preparation, they can better structure their preparation in the most economical way to achieve their greatest potential SAT score increase

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. Investigators could also examine the extent to which non-cognitive variables such as motivation, self-efficacy, and anxiety factor into successful performance on the SAT. 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

Allensworth, E., Macarena, C., & Ponisciak, S. (2008). *From high school to the future: ACT Preparation- too much, too late*. Chicago, IL: Consortium on Chicago School Research. Retrieved July 31, 2013, from http://ccsr.uchicago.edu/sites/default/files/publications/ACTReport08.pdf

Anderson, E. S., & Keith, T. Z. (1997). A longitudinal test of a model of academic success for at-risk high school students. *Journal of Educational Research, 90*(5), 259-268.

Appelrouth, J., Zabrucky, K. M., & Moore, D. (2014). *Preparing Students for College Admissions Tests*. Manuscript submitted for publication.

Arenson, K. W. (2006, March 26). Colleges Say SAT Mistakes May Affect Scholarships. *New York Times*. 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.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*, 63-105.

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.

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

Buchmann, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces 89*(2), 435-482.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record,* 64, 723-733.

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.

Cohen, S. (2012). The three biggest lies in college admission. *Forbes* online version. Retrieved September 15, 2013, from http://www.forbes.com/sites/stevecohen/2012/09/29/the-three-biggest-lies-in-college-admission

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94,* 95–120.

College Board. (2014). *2014 college-bound seniors-total group profile report*. Retrieved November 13, 2014, from https://secure-media.collegeboard.org/digitalServices/pdf/sat/TotalGroup-2014.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.

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.

Devine-Eller, A. (2012). Timing matters: Test preparation, race, and grade level. Sociological Forum, 27(2), 458-480.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. Journal of Applied Psychology, 84 (5), 795-805.

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.

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.

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

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

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

Grodsky, E. (2010). Learning in the shadows and in the light of day: A commentary on "Shadow Education, American Style: Test Preparation, the SAT and College Enrollment." *Social Forces, 89 (2),* 475-481.

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.

Horn, J. M. (1989). Truth, gender, and the SAT. *Academic Questions, 3*(1), 35.

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

Johannessen, L. R., & Kahn, E. A. (2001). *How to prepare students for high-stakes tests (and still live with your conscience!).* Paper presented at the annual fall conference of the Illinois Association of Teachers of English Elmhurst, IL.

Kapler, I. V., Cepeda, N. J., & Weston, T. (2012). Spacing, testing, and feedback: helping students overcome forgetting. *Education Canada, 52*(1), 31-34.

Kulik, J. A., Bangert-Drowns, R. L., & Kulik, C. C. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin, 95*(2), 179-188.

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.

Kohn, A. (2001). Two cheers for an end to the SAT. *Chronicle of Higher Education, 47*(26), B12.

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.

Loewen, J. W. (1988). Gender bias in SAT items. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Lubienski, S., Lubienski, C., & Crane, C. (2008). Achievement differences and school type: the role of school climate, teacher certification, and instruction. *American Journal of Education, 115(*1), 97-138.

Marjoribanks, K. (2003). Learning Environments, Family Contexts, Educational Aspirations and Attainment: A Moderation-Mediation Model Extended. *Learning Environments Research, 6*(3), 247-265.

Morgan D. L., & Michaelides, M. P. (2005). Setting cut scores for college placement. *College Board Research Report No 2005-9, 1-12.* Retrieved April 28, 2013 from https://professionals. collegeboard.com/profdownload/accuplacer-setting-cut-scores.pdf

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 (2008). *Guiding the Way to higher education: step-by-step to college workshops for students*. Arlington, VA: NACAC. Retrieved July 31, 2013, from http://www.nacacnet.org/research/PublicationsResources/ Marketplace/Documents/LateHS.pdf

National Association for College Admission Counseling (2012). *2012 State of College Admission.* Retrieved September 10, 2014 from http://www.nacacnet.org/research/research-data/Documents/2012SOCA.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. (1993). Coaching for the SAT: A summary of the summaries and an update. *Educational Measurement: Issues and Practice, 3*, 24-30.

Powers, D. E., & Rock, D. A. (1999). Effects of coaching on SAT 1: Reasoning Test Scores. *Journal of Educational Measurement, 36 (2),* 93-118.

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

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20 –27.

Rubenstein, J (2003). Test Preparation: What makes it effective? In *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.

Scruggs, T. E., & Mastropieri, M. A. (1992). *Teaching Test-Taking Skills: Helping Students Show What They Know*. Brookline, MA: Brookline Books.

Seifert, E. H., & Beck Jr., J. J. (1984). Relationships between task time and learning gains in secondary schools. *Journal of Educational Research, 78*(1), 5.

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.

Vigdor, J. L., & Clotfelter, C. T., (2003). Retaking the SAT. *The Journal of Human Resources, 38* (1), 1-33.

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.

Walberg, H. J. (1981). A psychological theory of educational productivity. In F. H.Farley and N.Gordon (Eds.), *Psychology and education*, 81– 110. Chicago: National Society for the Study of Education.

Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly, 10*, 196–218.

Zwick, R. (2002). Is the SAT a 'Wealth Test'?. *Phi Delta Kappan, 84*(4), 307-311.