Essays in Development Economics

Taehoon Kwon

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Rachel M. Heath, Chair

Michael Griffith

Jacob L. Vigdor

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Essays in Development Economics

Taehoon Kwon

Chair of the Supervisory Committee:
Associate Professor Rachel M. Heath
Department of Economics

Early-life education and health interventions have long-lasting impacts on our lifetime welfare. This dissertation consists of three essays in development economics for young people. Chapter 1 examines causal effects of the nationwide college admission policy reform in 2005 which introduces coarsening SAT grade structure and more competitive grading of high school GPA in Korea on high school students' time allocation behavior and welfare in a college admission game. I find no evidence of positive welfare effects. Chapter 2 extends the college admission policy implication by employing various non-parametric methods such as changes-in-changes and quantile difference-in-differences estimators introduced by Athey and Imbens (2006) to show the heterogeneous policy effects on students' leisure and study hours by outcome levels. Chapter 3[1] analyzes the behavioral effects of a social media fruit drink countermarketing messages. A randomized controlled trial among US Latinx parents with children age 0-5 found that social media countermarketing messages significantly reduced percent choosing fruit drink from online store in the intervention groups compared to the control group with car seat messages after the interventions.

[1] This chapter is based on Krieger et al. (2021) and under embargo until the journal publication decision is made and the submitted manuscript is published.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am just nothing but a lucky man who were able to reach the finish line during this unexpected pandemic thanks to so many people. I would like to express my sincere gratitude to my Chair Rachel Heath who always encouraged me to go forward with invaluable advice, Alan Griffith who accompanied me every step of the track, Jacob Vigdor for providing stunning insights and pinpoint coaching, and David Knight for his generous support and discussion. I am grateful to James Krieger for giving me an opportunity to join his research. I am honored to be a co-author with him and such a distinguished team: Christina Roberto, Rudy Ruiz, Lina Walkinshaw, and Jiali Yan. I thank Simon Reeve-Parker and Yu-chin Chen for their comprehensive supports, Jacques Lawarrée for sharing his valuable time throughout my darkest winter, and Haideh Salehi-Esfahani and Robert Halvorsen for having me as their assistant. I also would like to thank Ki Young Park and Yeon Seob Ha for their encouragement and motivation to pursue my doctoral degree. Financial support for my study was generously provided through Fulbright Graduate Study Award and Grover and Creta Ensley Fellowship. I wish to thank my parents-in-law Suk Ja Park and Pong Sik Choung for their unlimited supports. Thank my mom Yooja Kim, dad Kyoungcheon Kwon, and the big brother Jihoon Kwon for your love.

Finally, I am eternally grateful to the love of my life Hae Yun Choung for always being with me. You filled my every moment with smile, humor, and love to make me a better husband, father, and person since we first met.

DEDICATION

To my dear wife Hae Yun and our lovely son Eugene.

Chapter 1

COLLEGE ADMISSIONS, GRADING POLICY, AND STUDENTS' TIME ALLOCATION RESPONSES: THE CASE OF SOUTH KOREA

In this chapter, I study the effect of grading policy reform in college admissions competition on students' time allocation behavior and welfare. In the competition students play a zero-sum game where they only care about their relative ranking resulting in overinvestment of resources. The game participants face prisoner's dilemma and the result is an inefficient market outcome. Can SAT and GPA grading policy reduce high school students' college prep activities? This paper estimates causal effects of the nationwide education policy in 2005 which introduces coarsening SAT grade structure and more competitive grading of high school GPA in Korea on students' time allocation behavior in a college admission game. I model a student's welfare maximization behavior to investigate the policy effects on her college admission preparation activities. Empirical estimation is done by difference-in-differences method using repeated cross-sectional data. The result shows that the policy did not achieve the goal as the SAT taking students reduced leisure by a half hour per day and increased weekly private lessons by 1 hour and daily self study by 1.1 hour compared to the non SAT takers. Propensity score matching for difference-in-differences estimates and an alternative group setting also show the robust results.

1.1 Introduction

"Adolescents, just like adults, need time every day to unwind and interact with their peers. Too much pressure in schools might mean that students feel compelled to spend more time studying, leaving less time for these non-academic activities, at the expense of students' quality of life." - OECD (2017).

Teenagers in the modern society face a crucial trade-off between leisure and preparation of college. Leisure provides current direct utility with them. Also, quality leisure time for young people has been emphasized as not only a potential breeding ground for youth problems but also the possibilities for constructive development[1] in a broad range of positive outcomes (United Nations, 2006). At the same time, investment in college admission as an alternative use of time to leisure is believed to bring enormous returns in the labor market[2]. This implies that high school days are critical time for human capital investment (Bond et al., 2018). Zimmerman (2014) uses a regression discontinuity design to show the huge quarterly earnings difference by $1,593 between the two groups of students just above and below the admission cutoff for a university in Florida. Andrews et al. (2016) studies a case in Texas to find that the state's flagship college[3] graduation increases earnings significantly compared to the graduation from all other four-year public universities and public two-year colleges in Texas. In this context, estimating the high school student's lifetime welfare maximizing behavior who solves a problem of choosing her optimal leisure and admission investment may have economic implications on how to achieve individually and socially desirable outcomes.

[1] Furthermore, adolescent development has health implications throughout life (WHO).

[2] American Community Survey (2018) states that college graduates earned 75% more than high school graduates. Quality education in college improves labor productivity by providing access to information sources (Thomas et al., 1991), strengthening ability to decipher new information (Schultz, 1975) ability to learn (Rosenzweig, 1995), but also helps to join a more productive social network.

[3] University of Texas at Austin and Texas A&M University at College Station.

There exists a general concern that students and their families are investing a lot of time and money in college prep activities. Bound et al. (2009) find that US students response to growing college admission competition by taking more challenging courses, being involved more resume activities, or investing in signals of ability such as SAT scores. If college admissions competition is zero-sum game where students do care about their relative ranking, the ongoing unproductive arms race may lead society to an inefficient outcome as the game participants are facing prisoner's dilemma in that the market outcome is not the welfare maximizing result. College admission policy may intervene to affect students and parents incentives to move away from the market failure. Would the implementation of the new grading systems for SAT and GPA enhance high school students' welfare or boost admission competition? How would student's optimal resources allocation change in response to the grading policy? To answer these questions, I focus on a Korean education policy reform applied to the 2008 college freshmen cohort which aimed to reduce unnecessarily competitive admission race[4]. Since most Korean colleges examine high school GPA and SAT score of applicants in determining admission decisions, a nationwide education policy coarsening the SAT grading scale and strengthening importance of high school GPA may affect students' welfare by affecting their time and financial resources allocation behavior.

A number of studies models college admissions problem by using mechanism. Hafalir et al. (2018) model college admissions contest as a Bayesian game to solve the equilibrium where heterogeneous students choose a college and a single dimensional effort level to maximize their utilities from the college admissions. On the admission supply side study, Che and Koh (2016) develop a matching market model in the context of college admissions where colleges make admissions decisions to reach equilibria given the students' dominant strategies. Avery et al. (2014) design a non-price competition model to analyze strategic behaviors among

[4] *2008 College Entrance Examination Reform*, Ministry of Education, 2004.

colleges to attract highly qualified students.

I examine the nature of the fierce college competition in the framework of game of status (Hopkins and Kornienko (2004); Moldovanu et al. (2007)). In this view, the result of the competition can be socially undesirable in a sense that there exists over-expenditure in academic signaling activities such as GPA, SAT score, or resume to get better college admissions than the socially optimal level of investment. With limited household resource, time and money spending on those signaling efforts sacrifice students' leisure and consumption which would have enhanced directly teenagers' utility. Since the number of college admissions are limited, each applicant cares about her relative position, *status*, on the distributions of admission criteria such as SAT and GPA. According to the descending order of the applicants' academic performance (or SAT score and GPA), college admissions are assigned as the *status prizes* to those who are classified above the admission cutoff. A student's strategy is a mapping from time and financial resources to admission investment expenditure. Students want to achieve higher quality of admissions which allow better job market opportunities by signaling good quality of labor productivity. The game is described by a simultaneous move game under incomplete information. The equilibrium of this college admission game is similar to the symmetric Nash equilibrium in Hopkins and Kornienko (2004)[5] in the sense that the result is a Pareto-inferior outcome where students invest *too much* to the admission compared to the *socially* desirable outcome.

In the literature, however, study of teenagers' time use related to the college admission game is rare. Becker (1965) and Gronau (1977) introduce the time budget constraint of households to show the utility maximizing household's time allocation behavior. While the

[5] In this prisoners' dilemma in a non-cooperative one-shot admission game, the Nash equilibrium strategy of individual student (or her caregivers) is to invest her maximum amount of resource to get better admissions as long as incentive to marginal investment is greater than the marginal disutility or cost of effort or financial resource. Similar equilibrium concept in a *status* game where students care about their relative rank in the class is also described by Dubey and Geanakoplos (2010).

effect of non-working time use such as sleep on adult labor productivity has been recently studied (Gibson and Shrader, 2018), only little about young students' time allocation behavior related to adolescent welfare in response to the college admission policy change is investigated. My study models the lifetime welfare maximizing high school student's leisure allocation behavior in the context of Korean college admission competition. In this model, student's welfare is an increasing function of time use on *leisure*[6] and on college admission quality which takes time use on admission preparatory activities as input factors. As there is a trade-off between time use on current leisure and on study for the future, the optimal time allocation is strongly influenced by the college admission policy. This study also considers more choice variables such as students' time on leisure and admission preparatory activities to empirically capture the detailed changes in students' behavior connecting them to students' welfare. Most papers introduced above assume a simplified single dimensional effort choice by college applicants. My model expands student's choice variables to analyze the multi-dimensional policy effects.

My paper contributes to the empirical evidence of the demand side behaviors to the nationwide college admission policy changes by Korean government given the colleges' strategies. This study analyzes the college admissions game with a similar view to Bodoh-Creed and Hickman (2018). They model the college admissions market as a contest where students compete for admission by endogenously choosing the level of human capital to accrue prior to a rank-order admissions contest given the colleges ranking. Olszewski and Siegel (2016) also design a contest with many heterogeneous players and prizes to characterize equilibrium and Olszewski and Siegel (2019) show that a coarse performance-disclosure policy can be a Pareto improving policy which benefit all students. In contrast with their model prediction, this paper empirically evaluates the coarse performance-disclosure policy effect in Korea to-

[6]In this study, leisure is defined as free time away from study excluding time use on sleeping at night and having meals.

gether with the relative grading policy in GPA to show that both rule changes may have negative effects on student's welfare.

Although there is an enormous literature on the education policy impact on students' performance or effort incentive, a relatively small literature examines policy effect on student's welfare and resource management behavior captured by their specific time use data. Grau (2018) builds a rank-order tournament structural model where high school students decide their level of effort and whether or not to take the college admissions test and empirically evaluates that an increase in GPA weight increases students' average effort. This paper finds the similar effect an increased GPA weight while my model assumes that SAT taking decision is not a choice variable but exogenously determined by student's preference and family characteristics. There exists a study which estimates the effect of the new SAT coarse grading policy that my study is interested. Han et al. (2016) show mixed results that the students' efforts devoted to the SAT Korean have decreased, but those devoted to the SAT English have increased. However, they do not provide a clear interpretation of the mixed results by SAT subjects. They do not explicitly account for the effects of a change in GPA grading policy from absolute to relative grading which had implemented at the same time. Also, they calculate the marginal benefit of a SAT score by linking each SAT cutoff for a college admission and the average earning of the 50-year-old alumni of the college. This approach is only valid when students put efforts expecting the distribution of average annual salaries along the college ranking to be fixed in the future and this mid-career salary is the most significant determinant of optimal effort level in a student's utility function.

In this chapter, I estimate the comprehensive effect of the grading policy by considering changes of grading system in both SAT and GPA to more precisely identify the aggregate effect on student's admission preparation behavior. I model that high school students decide how much time they would spend on each academic and leisure activity to maximize their expected lifetime utility during the three-year of race to college admission. I show that coarse grading SAT system increases ex-ante uncertainty of student's return to investment and may result in larger expenditure on admission activities. Also, I shows that changing GPA grading scheme from absolute grading to relative grading provides incentive to study harder because achieving the highest status (grade, ranking) becomes more competitive within a cohort as the relative grading restricts the number of students in each grade category.

I empirically investigate the model's predictions about the policy effects on student's resource allocation behavior using Korean Education and Employment Panel (KEEP) Survey data. My difference-in-differences model compares SAT takers group and non-takers group in two academic cohorts before and after the policy changes in 2005. The result shows that the policy did not achieve the goal as the SAT taking students reduced leisure by a half hour per day and increased weekly private lessons by 1 hour and daily self study by 1.1 hour compared to the non SAT takers. Propensity score matching for difference-in-differences estimates also show the robust results. This implies that there was no policy achievement of reducing students' study burden due to the increased rank uncertainty in SAT and increased competitiveness in GPA.

The remainder of this chapter proceeds as follows. Section 1.2 provides policy background. Section 1.3 sets up the model. Section 1.4 describes data. Section 1.5 introduces the empirical strategy. Section 1.6 presents the empirical results. Section 1.7 checks robustness of empirical model and results. Section 1.8 concludes.

1.2 Background

1.2.1 High School Education and College Admission Policy in Korea

In this seciton, I briefly introduce the Korean education system for understanding of the policy background. Korean government manages the highly standardized courses and evaluation system from elementary to high school education. Every Korean parent has a legal obligation to send her five to seven years old children to an elementary school and a middle school until they graduate therefrom[7]. After six years of elementary school and three years of middle school, most middle school graduates decide to go to private or public high schools. In 2019, 463,130 middle school graduates out of 464,717 (99.7%) continued their education at high school. Seventy percent of middle school graduates choose to go to general high school while more than a quarter selects vocational high school. The rest of them go to the specialized high schools in natural science, foreign language, or art and music are special cases under this system. College entrance rate is also high. Public high school education cost is affordable. Annual tuition and admission fee of public high school in the most expensive region, Seoul, are only KRW 1,464,900 (USD 1,300) in 2020. Children in a household where its income is less than 50% of the median household income[8] are able to receive 100% subsidy from the government for the tuition, admission fee, and expenses for purchasing textbooks[9]. In 2008 academic year which is the sample year of this study, 82.8% of high school graduates entered university or college. These facts are summarized in Figure 1.1.

Colleges may admit freshmen through two policies, regular (*Jeongsi*) and rolling (*Susi*).

---

[7]ELEMENTARY AND SECONDARY EDUCATION ACT (Act No. 16672, Dec. 3, 2019) CHAPTER II COMPULSORY EDUCATION Article 13 (Obligation of School Enrollment) (1) Every citizen shall send sons and daughters or children under his or her care to an elementary school from March 1 of the year following the year in which the date they reach six years of age falls and have them attend the elementary school until they graduate therefrom. (2) Notwithstanding paragraph (1), every citizen may send sons and daughters or children under his or her care to an elementary school in the year following the year in which the date they reach five years or seven years of age falls. In such cases, he or she shall have his or her sons and daughters or children attend an elementary school from March 1 of the year they are admitted to the elementary school until they graduate therefrom. (3) Every citizen shall send sons and daughters or children under his or her care to a middle school at the beginning of the school year following the school year in which they graduate from an elementary school and have them attend the middle school until they graduate therefrom.

[8]This is an annual income of KRW 23,223,468 which is equivalent to USD 20,609 for a household of three family members.

[9]Ministry of Education implemented the free high school education policy from 2020 as an intermediate step for the compulsory high school.

The main difference between *jeongsi* and *susi* is the significance and impact of SAT scores on admission. Rolling admissions open in Spring and Autumn before *jeongsi* starts. Spring *susi* only requires GPA and resume while *susi* in Autumn may set minimum cutoff SAT grades which are typically very lower than the cutoffs in *jeongsi* given the same college[10]. Typically, *susi* applicants do better in GPA than in SAT. Thus, *susi* applicants exert less efforts on SAT to focus on other admission criteria. Indeed, some *susi* admitted students fail to satisfy the muinimum SAT requirements[11]. The admission timeline for the 2008 academic year is illustrated in Figure 1.2.

Once a student is admitted from a college in the *susi* admission, she must register for the admitted school and cannot apply to other colleges in the same academic year. This prevents the admitted *Susi* students from competing in *jeongsi*. Regular admission typically starts after SAT tested on November and requires both high school GPA and SAT scores together with resume, essay, interview, or any combination of them. Nevertheless, SAT is the primary source of evaluating *jeongsi* applicants for colleges as it is considered to be the most standardized test effectively testing students' ability to study in college. In 2005 academic year, 55.7% of college freshmen are admitted through the regular admission process while 46.9% of 2008 college freshmen achieved regular admission.

As in many countries, growing admission competition among college applicants is also pervasive in Korea. Bound et al. (2009) find data which shows that the US high school students in 2004 spent significantly more time on behavior associated with college preparation than did their counterparts from ten and twenty years before. The issue seems to be more

---

[10]For example, Seoul National University requires Autumn *susi* applicants to achieve just any two second-highest grades out of Korean, math, English, social science, and natural science SAT subjects as the minimum admission requirement. In *jeongsi* admission, the cutoff student's average SAT grade for social studies was 1.0575. This implies that if an applicant fails to get Grade 1 in all SAT subjects, she might not get a *jeongsi* admission.

[11]For example, Korea University required the second-highest grades in two SAT subjects as the minimum admission criteria in 2008 and the 35.5% of admitted students did not achieve them to lose their admissions.

serious in Korea. Korean college applicants do care about their relative ranking in SAT and GPA to get a better college admission. With the limited number of schools relative to the number of students who want to be admitted, Korean students start taking extra lessons in their early ages and spend 6.4 years on the lessons to prepare college admission which is the longest among the twenty two high-income countries (OECD, 2017).

In *jeongsi* admissions, most of Korean colleges require students to take SAT to submit the scores as one of the crucial admission criteria. Annual SAT and quarterly mock exams are hosted by Korea Institute for Curriculum and Evaluation (KICE). SAT can be taken only once a year at the end of the academic year, normally on the second or the third Thursday of November. As the only chance determines the quality of college admission for an applicant and the opportunity cost of failing the first take is another year of investment of financial and labor resources which cannot be underestimated by the twenty-year-old young people, high school seniors have huge incentives to exert their best efforts given the resource constraints set by their caregivers.

1.2.2 Policy Description and Timeline

"Cram schools like the one I taught in — known as "hagwon" in Korean — are a mainstay of the South Korean education system and a symbol of parental yearning to see their children succeed at all costs. "Hagwons" are soulless facilities, with room after room divided by thin walls, lit by long fluorescent bulbs, and stuffed with students memorizing English vocabulary, Korean grammar rules and math formulas. Students typically stay after regular school hours until 10 p.m. or later." — The New York Times, "How South Korea Enslaves Its Students." August 3, 2014.

Korean Ministry of Education decided to launch a new SAT and GPA grading policy.

The government aimed to strengthen public education, reduce unnecessary competition in SAT, and promote diversity in college admissions[12]. The policy was implemented twofold. First, high school GPA grading regime changed from *absolute basis* to *relative basis* which emphasizes student's relative performance at school. Now, the GPA becomes more informative signal as it provides additional information such as percentile, mean, and standard deviation of each subject in the student's quarterly exams at school. This ranking information was not provided to students and colleges before the policy change. Furthermore, before the policy change, *easy-peasy* midterm and final exams at school were *pervasive*[13] as teachers have an incentive to boost their student's GPA to enhance their students' admission performance. Under the new policy, however, when there are many ties due to the easy test their percentile is assigned by taking average of their position. That is, if there are twenty students with maximum score out of one hundred students, now the grades for them should be *Grade 2* for all twenty ties which is the second-highest grade instead of *Grade 1*.

Second, the SAT results will be delivered only in the form of each subject's grade. No further information such as standardized score and exact percentile was available as before. For example, a student who took 2008 SAT and achieved 134 points in math which fell into 97 percentile of population and is assigned to be the first grade which is only given to the top four-percent students in the subject. Overall, the new grading policies give more weight on high school GPA and less weight on College Scholastic Ability Test (SAT) in college admission process by adjusting informativeness of them. Table 1.1 shows the grade distributions of GPA and SAT. Since a SAT raw score is standardized to assign a grade over normal distribution, the median student's grade (50 percentile) in each subject will be Grade 5 which covers $40 \sim 60$ percentiles.

---

[12]Ministry of Education, *Press Release*, August 26, 2004.

[13]Ibid.

Figure 1.3 presents the policy summary and the implementation sequence. It was announced in 2002 and first applied to the 2005 high school freshmen. Before the policy, GPA was given by five grades in an *absolute basis*. For example, when a student gets 90 out of 100 which is the cutoff of the highest grade she would get grade 1. If she gets 75 out of 100 the grade 3 will be given regardless of difficulty of the test and the distribution of the test score. Further, there was no information about the distribution of test scores. Under previous regime, there existed an incentive for school teachers to make easier exams to give more generous GPA to their students. The result was inflation of GPA and colleges who know this situation distrusted high school GPA and more relied on SAT which is the standardized nationwide exam.

The motivation of the policy reform can be found in a survey by the Organisation for Economic Co-operation and Development (OECD, 2017). Among OECD countries Korea ranked at the bottom in descending order of the percentage of 15-year-old students who reported being very satisfied with their life. Furthermore, the survey describes that more than 20% of student in Korea reported that they are not satisfied with their life while less than 4% of students in the Netherlands reported dissatisfaction. Another interesting result of the survey is that only 10% of Korean students cited the pleasure of studying as a motive for taking additional classes. Kahneman and Krueger (2006) show that those who report less satisfaction with their lives as a whole also spend a greater fraction of their time in an unpleasant state. Taking the fact that the 15-year-old Korean respondents of the survey have already taken 6.4 years, which is the longest among OECD countries, of extra instruction after school hours in average into account, Korean students would have been spending time on unpleasant learning activities throughout their childhood and adolescent.

The main purpose of those additional lessons in Korea is to prepare a specialized high school admission and a good college admission by learning English, math, and science in advance of the regular curriculum[14]. Meanwhile, percentage of Korean students who reported that they exercise or practice sports before or after school is only 46.3% which is the lowest not only among OECD countries but also among twenty one partner countries including China, Brazil, Russia, and Singapore (OECD, 2017). That is, Korean high school students spend their time at cram schools rather than at play fields. In my sample year of 2009, 62.8% of total high school students and 69.2% of GPA top 10% students have taken private lessons (Kostat, 2009). Also, Korean private tutoring market accounted for 1.2% of Korean GDP. The result is adverse welfare effect on teenagers' physical and mental health. An average Korean high school student slept only 5.4 hours a day while studied 11 hours a day. Their biggest concern was academic performance (69.1%). A half of high school students have a discriminatory experience due to their low academic performance. In addition, the main cause of parental conflict is academic achievement (50.1%) (Kostat, 2010).

1.3 Theoretical Framework

In this section, I model a welfare maximizing student's problem to show that, first, coarse grading increases incentive to invest more on admission preparation when agents are loss averters. Second, the model shows that the behavioral reactions to the admission weight change is theoretically ambiguous. Third, Coarse grading SAT increases uncertainty in student's status to induce higher efforts. Forth, I show that the relative grading scheme for high school GPA raise the degree of competitiveness among students compared to the original absolute grading policy. Fifth, I show the policy can lead to inefficient learning. Sixth, testable implications are discussed.

[14] All types of prior learning has been strictly prohibited by the special law for normalization of public education from 2014.

1.3.1 Baseline Model: Students' Time Allocation

I model a student's two-period lifetime utility maximization problem to illustrate time allocation behavior to college admission related study and leisure activities. A high school student lives two periods; high school days ($t = 0$) and the rest of her life ($t = 1$). She wants to maximize expected total lifetime utility U. A representative student's optimization problem is modeled as follows:

$$\max_{l_0, l_q, k_q, x} U = u_0(l_0, x) + \beta u_1(Q) \tag{1.1}$$

$$s.t. \quad l_0, l_q, k_q, x > 0 \tag{1.2}$$

$$l_g + l_s + l_r + l_0 = l \tag{1.3}$$

$$p_g k_g + p_s k_s + p_r k_r + px \leq I \tag{1.4}$$

and

$$Q = w_g \cdot g(l_g, k_g) + w_s \cdot s(l_s, k_s) + w_r \cdot r(l_r, k_r) + \epsilon. \tag{1.5}$$

where u_0 is a current utility ($t = 0$) from current leisure l_0 in high school days and u_1 is a discounted future utility ($t = 1$) depending on a quality of college admission Q which is a function of high school GPA $g(l_g, k_g)$, SAT score or grade $s(l_s, k_s)$, resume $r(l_r, k_r)$, and unobservable variables ϵ. I consider myopic (or present bias) students who are assumed to enjoy future utility u_1 from their college admission directly, but the delayed benefit u_1 is discounted by $0 < \beta < 1$. I further assume that the utility function $u(\cdot)$ is increasing, concave, and continuously differentiable in its components.

The admission quality function Q is a non-decreasing function of each component which implies that the partial derivatives of Q with respect to g, s, and r are non-negative. Accordingly, getting a more competitive college admission Q brings higher lifetime utility after high school graduation as it potentially guarantees a better signal on labor market and a more productive social network. The determinants of admission quality g, s, and r are all non-decreasing functions of time and money. That is, the production functions of high school GPA (g), SAT scores (s), and quality of resume (r) are the increasing functions of time (or labor) and money (or capital) similar to a competitive firm in a goods market. Marginal productivity of time l_q and money k_q to each academic production $q = g, s, r$ are assumed to be diminishing in their production factors. For example, achieving higher grade in a SAT subject is getting more difficult and requires more effort and financial investment.

I introduce weights in college admission w_q where

$$\sum_{q=g,s,r} w_q = 1 \tag{1.6}$$

to reflect the policy change interested. I assume that the admission quality Q is determined by the weighted average of three admission components g, s, and r. As an extreme case, when only SAT score determines the admission result without considering high school GPA or resume at all, the weight on SAT w_s in the function $Q(g, s, r)$ would only be positive and equal to one. Before 2007, most Korean colleges relied heavily on SAT scores to screen their applicants. Further, some students from the specialized high schools such as natural science or foreign language high schools could apply to the SAT-only admission channel of the major universities as they have adverse high school GPA due to the highly competitive school cohort. After the policy aiming to promote public education and reduce student's stress from preparation of SAT, colleges had to enhance the weight on school records such as GPA and resume in their admission process following government policy. Ministry of Education was able to enforce this policy by controlling information of SAT results to students and colleges and by implicit financial support threat to colleges as many colleges rely heavily on the central government subsidy.

Equation (1.5) implies that a student's admission quality is determined by her weighted average of SAT score, high school GPA, resume, and unobservable factors. Colleges use Equation (1.5) in their admission process with different weights depending on their preference. By the definition of admission quality Q above, a partial effect of a marginal increase in GPA, SAT, or resume on Q equals its weight. That is,

$$\frac{\partial Q}{\partial q} = w_q. \tag{1.7}$$

Equation (1.1) and (1.5) imply that student behavior and welfare would vary with the college admission policy. Noticing that most colleges in Korea assign weights on each component g, s, and r to look over students' applications, we can state the admission quality $Q(\cdot)$ as a function of the weighted average of three criteria and unobservables ϵ.

The representative intertemporal utility maximizing student faces a time constraint in Equation (1.2) and a household budget constraint in Equation (1.3) where l_q with $q \in \{g, s, r\}$ denotes input of time on each academic activity q, l_0 is time on non-academic activities, k_q is input of money on each academic activity q, p_q is the price of each academic activity q, and px denotes total money on other goods and services. This model assumes that students only use their time in studying or leisure activity. Current leisure l_0 provides direct utility at $t = 0$ while time spending on academic activities l_g, l_s, and l_r combined with capital investment k_q produces admission factors such as GPA, SAT score, and resume which ultimately contribute to the admission quality $Q(g, s, r)$.

As almost all public and private high schools in Korea follow the common curriculum designed by the Ministry of Education, SAT subjects including Korean, English, math, and social and natural science are almost identical to those being taught and examined quarterly in each high school. In fact, questions of SAT should be set based on textbook written by the Ministry of Education or authorized publishers and workbook published by Korea Education Broadcasting System (EBS) which is the official SAT preparation material designated by the Ministry of Education. In this context, the effect of one additional hour of study on SAT math would also have positive impact on GPA. That is, it is natural to expect that for every level of l_q and k_q,

$$\frac{\partial q}{\partial l_q} > \frac{\partial q'}{\partial l_q} \geq 0 \tag{1.8}$$

$$\frac{\partial q}{\partial k_q} > \frac{\partial q'}{\partial k_q} \geq 0 \tag{1.9}$$

where $q, q' \in \{g, s, r\}$ and $q \neq q'$. For simplicity of analysis, instead, I only consider the net effect of l_q and k_q on q after normalizing the indirect effect to be zero. In other words, the complementary effect of allocating resource to one activity on other activity is considered null. This assumption yields

$$\frac{\partial q}{\partial l_q} > \frac{\partial q'}{\partial l_q} = 0 \tag{1.10}$$

$$\frac{\partial q}{\partial k_q} > \frac{\partial q'}{\partial k_q} = 0. \tag{1.11}$$

Equation (1.10) and (1.11) imply that even though the basic concepts and formula tested by SAT and school exams for GPA are the same, when a student invests one additional hour to preparing SAT math her marginal increase in math GPA would be zero since her math midterm exam tests the very specific chapter in the workbook or teacher-made test bank. To capture disparate students' inherent abilities to study or total factor productivity in producing an admission factor $q(\cdot)$, I introduce individual heterogeneity in the model. Each student has different production technology of GPA, SAT, and resume denoted by i.

Then, the individual admission quality Q_i can be written as

$$Q^i = w_g \cdot g^i(l_g^i, k_g^i) + w_s \cdot s^i(l_s^i, k_s^i) + w_r \cdot r^i(l_r^i, k_r^i) + \epsilon^i. \quad (1.12)$$

In other words, given the same amount of leisure l_q and money k_q invested in a certain admission activity q, the marginal productivity of time $\frac{\partial q^i}{\partial l_q}$ and capital $\frac{\partial q^i}{\partial k_q}$ are heterogeneous for every student. The individual ability or productivity for each activity g, s, and r is unobservable private information but assumed to be normally distributed over the population cohort. The equilibrium conditions imply that the lifetime utility maximizing student will balance her time so that the marginal utility gain from the last hour spent on each activity is identical. See A.1 for the derivation of equilibria.

1.3.2 Endogenous SAT Taking

In the model, a representative student makes SAT taking decision by comparing the expected lifetime utilities with and without college degree. When she predicts that her college degree may not dramatically encourage her future welfare, then she may choose to invest more time on current leisure (l_0) and consumption (x) and not to invest to SAT scores ($l_s = 0$ and $k_s = 0$). She, however, still invest some time on GPA and resume to obtain better referral letters from school teachers which will be used on the job market. SAT taking is the first stage decision and it is closely related to the degree of grading policy exposure. This stage can be structurally modeled (Grau, 2018), but in this chapter I assume the SAT taking decision is made significantly based on the exogenous variables such as family background, individual characteristics, and inherent productivity.

1.3.3 Coarse Grading and Uncertain Status

Welfare maximizing college applicant optimizes her time given the college admissions grading schemes. Coarsening SAT grades may encourage or discourage a student's incentive to work hard depending on her expected *status*. More details will be discussed in 1.3.4. This subsection investigates the structural change in student's status with respect to uncertainty of ranking under the new SAT grade system. Harbaugh and Rasmusen (2018) shows that coarse grading scheme such as pass-fail grading reduces public uncertainty by minimizing expected and actual quality of sender because the extra participation outweighs the coarser reporting. As almost all college applicants take SAT, however, college admission game ensures full participation already regardless of grading scheme. Thus, the relationship between the coarse grading and student's ranking uncertainty in the admission game might be different from the case in Harbaugh and Rasmusen (2018).

I admit two sources of uncertainty in a risk-averse[15] student's SAT status given her *score* of a SAT subject. This yields the fact that the uncertainty eliminates *safe zone* within a SAT grade where a student is *ex-ante* confident about her SAT grade at the moment of resource allocation decision. Assume that a student's SAT status s^i in Equation (1.12) is a increasing function of her realized *ex-post* score s_i^*

$$s^i = \psi_i(s_i^*, s_{-i}^*, \Theta) \tag{1.13}$$

where s_{-i}^* is realized rivals SAT scores with $\partial\psi_i/\partial s_{-i}^* \leq 0$ and Θ is the grading scheme which contains the information of degree of coarse grading.

First, one source of uncertainty of SAT status is coming from the individual level uncertainty in SAT score. Suppose when a test taker decides her investment level on SAT the

[15]The same behavioral conclusion can be derived with a loss aversion setup. See A.2.

agent's realized score s_i^* is a random variable consisting of a fixed potential part $s_i^p(l_s^i, k_s^i)$ and an error term υ_i.

$$s_i^* = s_i^p + \nu_i. \tag{1.14}$$

The error term ν_i follows a distribution

$$\nu_i \sim (0, \sigma_i^2). \tag{1.15}$$

I assume that the variance of the realized score is a decreasing function of individual test taking ability and invested resources l_s^i and k_s^i, but never goes to zero. the score variation generates uncertainty in $s*_i$ around s_i^p because σ_i is always positive as there exist uncontrollable factors such as physical and mental condition, testing taking environment, luck on guess, and trembling hands.

Second, another source is cohort level uncertainty. Students face *ex-ante* uncertainty in their SAT grades because the score-grade charts are finalized *after* the SAT tests. To build the charts, the test authority needs information of all students' row scores. Once every cutoff of grade from 1 to 9 is determined, each student's score is assigned to one of nine grades depending on its status over the whole score distribution. Thus, given a student i's realized score s_i^*, her status (grade) s^i is still *ex-ante* random which depends on the rivals' performance s_{-1}^*, difficulty of the test, and the degree of coarse grading.

A risk-averse college applicant needs to make a resource allocation decision under the individual and cohort level uncertainty. When the degree of coarse grading is zero just like the grading scheme before the policy change, only the students with the same score will be assigned to the same status. The size of each bin is *one* point of SAT score. That is, given a SAT score we can expect that the number of students in a bin would be less than any other bin from the coarse grading scheme. Thus, an increase or a decrease in SAT score s_i^* from

s^*_{i0} to $s^*_{i1}(= s^*_{i0} + 1)$ generates the smallest change in status s^i than the status change in any other coarse grading scheme. Now, consider the coarse grading scheme with only nine grades as shown in Table 1.1. If *ex-ante* uncertainty is zero than a student can perfectly predict her SAT status s^i because s^i is only determined by the level of investment which is student's choice variable. Under uncertainty, however, the student should consider the possibility of a dramatic change of her status across grades which will be critical to the admission results in any direction. Suppose a student's anchor score s^p_i is expected to be the raw score of 85 in Korean. Table A.1 shows that the raw score of 85 in Korean in 2009 SAT is assigned to Grade 3. With the same raw score she will get Grade 2 in 2008. If she took 2007 SAT she would have obtained Grade 4. Even though we ignore individual level uncertainty around a student's potential score, she would never be sure about her status s^i because of the rivals performance s^*_{-i} is also ex-ante random.

The above argument is also valid when I fix a student's potential status as her percentile (not her absolute raw score). For example, suppose a student's potential Korean percentile s^p_i is 5% from the top. If she took 2009, 2008, and 2005 SAT Korean her would have earned Grade 2 while she could get Grade 1 if she took the test in 2007 or 2006. Recognizing that there exists more than twenty thousands of Grade 1 students in each cohort, losing Grade 1 just by chance at SAT Korean could result in a dramatic drop of quality of admission. Since the coarse grading policy in Korea assigns all ties to the upper grade even though the total number of students in the bin exceeds the guideline cutoff, a student's realized status is highly uncertain depending on the individual and cohort level uncertainties.

1.3.4 Ambiguous Behavioral Reactions

How does the new grading policies affect students' behavior? This is a key question of this paper. In short, the theory gives ambiguous prediction depending on the degrees of individual and cohort uncertainty, coarse grading, and heterogeneity in students' productivity. Coarse grading of SAT restricts the information quality of SAT transcript. Thus, the policy forces colleges to adjust their weights on admission criteria. Depending on the individual productivity q^i and the predicted SAT grades[16], the directions and magnitudes of students' resource allocation responses ∂l_q and ∂k_q to the policy change would vary. As Han et al. (2016) suggest, coarsening SAT grades may have two offsetting effects. There can be an increase in time investment on SAT when a student thinks her score is around the cutoff of a certain grade. This is because the marginal change of grades in any direction would make a critical change of her admission quality. Contrary to this, another student with a score far from upper or lower bound of a class or who does not need to enhance her grade would like to reduce resources allocated to SAT. The equilibrium first-order conditions (A.2) - (A.5) also capture this ambiguous direction of effect. Reduced leisure can be consumed as l_0 and x to obtain more u_0 while keeping the level of Q constant or can be invested as another academic activities $l_{q \neq s}$ and $k_{q \neq s}$ to get better future rewards. Furthermore, as shown in the previous section, more uncertainty in SAT status under the coarse grading eliminates the chance for students to predict their grades before hand. The empirical exercise would be able to identify the average policy effect on the whole samples and also on the various sub-groups.

1.3.5 Relative Grading GPA and Harsh Competition

Another main change in the admission policy is the relative grading high school GPA which was implemented against the pervasive high school GPA inflation due to the *easy* midterm and final exams administered by each high school authority under the absolute grading

[16] Korean students predict their SAT rankings by referring to their scores of monthly private and quarterly public SAT mock tests. A typical college applicant takes ten to thirty mock tests throughout the three years of high school period. It usually takes about eight hours to complete a set of SAT mock test.

scheme[17]. Under the new grading policy, students face more competitive environment at school as only top 4% of the school cohort can obtain the highest grade. When there are more than 4% of ties at the top, all students should get a grade corresponding to their average percentile. For example, if 20% of students in a school got full credits due to an easy exam, those 20% students' percentile becomes 10% from the top. This implies that no student can get the first grade as no one marked top 4% in this exam. Thus, teachers have now no incentive to make exams easy as low difficulty will harm all students in ties. As a result, it is natural to observe grade deflation under the new relative grading scheme. That is, now high school GPA provides more informative signal about individual student's status on the standardized normal distribution in her school cohort while students have strong incentive to invest more resources and compete with their friends than before to maintain their desired GPA given the limited portion of good grades.

1.3.6 Inefficient Learning

Efficient in learning is achieved when a student is able to complete high school courses at the minimized costs and obtain desired college admission. Contrary to efficient learning, when a student could get a high school degree and the admission only by sacrificing almost all of her household income or at least by investing more than the optimal level of resources, the learning is *inefficient* and over-invested. Suppose n applicants are applying to college A and only q qualified applicants per academic year will be admitted where $n > q$. Applicants are sorted by their admission quality function $Q^i(g^i, s^i, r^i, w_q, \varepsilon)$ where $i = 1, ..., n$. The q students from the top Q^1 to Q^q will be admitted and $(q+1)$th applicant cannot achieve the

[17]For example, Yonsei University reports in 2004 that among their applicants, 7 high schools gave grade A to *all* students in a subject. Also, 37 schools gave grade A to more than one hundred students in a subject by making very easy midterm and final exams.

admission A. Admission A is a binary outcome

$$A = \begin{cases} 1 & \text{if} \quad Q^i \geq Q^q \\ 0 & \text{if} \quad Q^i < Q^q \end{cases}$$

where Q^q is the qth highest student's quality. For simplicity, I assume that there is no identical quality of applicant among any of two or more applicants $i \neq j$. That is, for all $i \neq j$,

$$Q^i \neq Q^j.$$

Assume that a student i updates her expected admission quality $E(Q^i|\Omega)$ and this year's expected admission threshold $E(Q^q|\Omega)$ by using information set Ω which includes previous SAT mock test grades and score distribution, current HGPA, and honors and awards history. When she expects that her admission quality $E(Q^i|\Omega)$ is less than $E(Q^q|\Omega)$, she may want to invest more on studying SAT, preparing midterm and final exam, and polishing resume to achieve higher Q^i than the threshold level of admission quality Q^q. Other applicants who know and observe their rivals' efforts may also have incentives to invest more on admission preparation. This implies that other student's performance would be reflected in Ω and then student i makes investment decision depending on her rival's effort level. Given individual productivity, students will compare an additional lifetime benefit of resource investment, which would be a better admission, and an additional loss of resources. With the limited number of admission, it is inevitable to compete with other students. This results marginal cost of producing one more unit of admission quality to be getting higher as more rivals join this competition and invest more. At some point of this investment war, students face extremely high marginal cost of Q due to diminishing marginal product of leisure and money invested to g, s, r and also due to the increased level of the expected threshold $E(Q^q|\Omega)$ by rivals' investment. That is, as the degree of admission competition increase, students should spend more resources on the admission competition to obtain the same level of admission A which would distort household decision and other markets leading to the over-investment in the college admission game.

1.4 Data

The analysis in this paper is based on Korean Education and Employment Panel Survey (KEEP) data surveyed by Korea Research Institute for Vocational Education & Training (KRIVET) which contain information on student and household demographics, income and wealth, school activities and test scores, and employment status. The survey is a longitudinal research study of two cohorts – young and old – that extracts nationally representative samples from the population group and traces them for over twelve years from 2004 to 2015.

A total of 6,052 respondents were selected to participate in this survey, comprising the younger cohort consisting of 2,294 middle school seniors (3rd graders) and the older cohort consisting of 1,887 high school seniors (3rd graders) and 1,015 vocational and technical high school seniors (3rd graders) at the year of 2004. Supplemental data were provided by teachers and school administrators as well. The data tracks two cohorts. before and after the policy change at the period to investigate their behavior in high school. Thus, my sample from the different two cohorts can be regarded as repeated-cross sectional data.

I exclude the samples who went to college by the rolling system (*Susi 1* and *Susi 2*) from both intervention and non-intervention groups[18]. I only choose non-*susi* SAT takers as my intervention group samples. This is because at the time of the surveys (March for 2005 cohort and July for 2008 cohort) *susi* applicants were still preparing SAT without knowing the application results which were scheduled to be announced around August. This implies

[18]In the alternative setting with *Susi* admitted students as a control group, the empirical results are similar in directions for most of the outcome variables. See Appendix.

that the outcome data in hand was collected at the moment when both regular and rolling applicants were studying SAT and the two types of students' leisure allocation were affected by the grading policy reform. Also The *Susi* admitted students before SAT should enroll the admitted college and cannot participate in regular admission process. Once they are admitted, therefore, they have much less study pressure and no (*Susi 1*) or little (*Susi 2*) incentive to invest to SAT preparation activities.

1.4.1 SAT Score Distribution

Table 1.3 describes the available information of SAT results before and after the policy change in the whole sample of the data[19]. Some students took SAT more than once over multiple years after their high school graduation to get a better college admission. In this paper, only the first-taken SAT results of high school students in each cohort are summarized. 2005 admission SAT (taken in 2004 Autumn) takers received all types of result available on their transcripts including standardized scores, grades, and percentile from the authority. College applicants provided all score information to the applying colleges, maximum three colleges for each student, combining with their high school GPA and resume. Students of 2008 admission SAT (taken in 2007 Autumn) group were available only to obtain their grades for each SAT subject. This implies that students do not know their exact score and the percentile on the population distribution. For example, the grade 1 which is the highest among nine grades is assigned to those who ranked within the highest four percentage in the population of the

[19]Table A.1 shows the population statistics of SAT grade and score distribution for four subjects, Korean, Math A, Math B, and English, from 2005 to 2009 including my two sample cohorts. Every year, the cutoff raw score for each grade in a subject varies depending on the exam difficulty. For the comparability across different cohorts, each SAT subject's result is standardized and the grade cutoffs are assigned over the standard normal distribution. The around-median (40 $\sim$ 60 percentile) students in each SAT subject receive grade 5. The top and the bottom 4% students will be assigned to grade 1 and grade 9, respectively. An average student in 2008 cohort has slightly higher grade ranging from 5.06 to 5.14 for three major subjects than the one in 2005 cohort whose grades ranging from 5.48 to 5.59, but those are located around the median grade of the population which is exactly 5.

subject. That is, the top scorer and a student slightly above the four percentage cutoff would have the same grade. Colleges then are not able to distinguish among those two students by their SAT scores. To screen proper applicants colleges need to more closely review other documents such as high school GPA and resume than the inexact SAT grades.

1.5 Estimation Strategy

Following the causal inference framework organized by Imbens and Rubin (2015), I use difference-in-differences estimation to investigate causal effects of the grading policy change. This can be said to be the differences between potential outcomes of each students' time allocation behavior with and without policy change. Due to the nature of nationwide policy I interested, all students who prepare SAT should be affected by this policy. Given this policy environment, My identification strategy is to observe outcomes of SAT takers and SAT non-takers in two cohorts before and after policy change in 2005. As we do not observe counterfactual potential outcomes such as SAT taking student's leisure time that she would have allocated when there was no policy change, I investigate appropriate comparison group to see the effect of policy on SAT takers.

1.5.1 Difference-in-Differences Estimation

In this policy analysis, I use difference-in-differences estimation with repeated cross-sectional survey data of two different high school cohorts before and after the policy to identify causal effects on student's resource allocation behavior and welfare. As far as I know, no randomized controlled trial study for this education policy at the time of implementation was conducted. Thus, the policy effects can only be investigated by using observational data. Clearly, prior to the implementation of the policy in 2005 no student in my sample was exposed to the policy change. Also, the high school GPA policy change was applied to the 2008 SAT cohort for the first time in 2005 when they were the high school freshmen. New SAT grading policy was also applied to the SAT taken in 2007 for 2008 academic year. Thus, only the younger cohort in my sample was affected as they studied at high school under the new grading regime[20]. Since the policy was announced in 2004 to apply to the 2008 SAT cohort who are 2005 high school freshmen, there could be anticipation effects on them. Endogenous choice of SAT take would be a possible threat to identifying the non-intervention group. This is discussed in Section 1.5.4.

Among those in the younger cohort, only students who want to go to college took SAT. That is, SAT taker group is more affected by the policy change. Interacting variables are *post* and *SAT taker* dummy variables. Under the parallel trend assumption, the difference-in-differences (DD) estimator suggests the time changes over the policy in the average outcomes for the SAT takers and non-takers. The validity of the assumption is discussed in Section 1.5.6.

1.5.2 Primary Outcome Variables: Weekday Leisure

The primary outcome variable of this study are student's *weekday leisure hours per day* during their *last year of high school*. As all Korean high school students are subject to the almost same daily class schedule planned by the central government, I expect that the quantity of time use is the key input for student's welfare and quality of college admission. Thus, in accordance with my model in the previous section, I take weekday leisure time as the primary outcome variable. Also, secondary outcome variables of interest are selected to evaluate students welfare. I consider time spending on private education and time of self study as the secondary outcome variables to measure the policy effect on students'

---

[20]In fact, the 2008 SAT cohort was the only one cohort who could not obtain their SAT results in forms of percentile and standardized scores. Korean government regressed to the original SAT grading policy from the 2009 SAT cohort.

welfare. Unfortunately, amount of financial resources spent to cram schools is not included in this study due to the inflation during the period and large numbers of missing data and inconsistency of survey questions across cohorts.

Figure 1.4 shows mean outcomes by group. I find that average weekday leisure is enjoyed more in the SAT non-takers group than the SAT takers group. Hours spent for self study and private lesson also differ between the groups at the baseline. Proportion of students who participate in academic club activity tends to be higher in SAT takers group. SAT takers are elected more as class or school leaders. These differences in baseline outcomes implies that there exists baseline level difference of outcome variables across groups. One way to interpret this difference is that SAT takers typically spend more time to study than the non-takers who are relatively under less time pressure during their last year of high school.

1.5.3 Balance in Characteristics

The intervention which is the exogenous source of variation in this analysis is the time-specific grading policy changes on SAT and GPA applied only to the younger cohort in my sample. To examine the impact of the policy, I focus on two high school graduate cohorts before (2005) and after (2008) the policy change. Total 435,538 college applicants out of 569,272 high school graduates in 2005 (76.51%) and 446,597 applicants out of 581,921 high school seniors in 2008 (76.75%) took SAT for 2005 and 2008 college admissions, respectively. In this study, *SAT non-takers groups* in each cohort consist of students who did not take SAT ($SATtaker = 0$) in their third-year of high school and were not affected by the policy changes. *SAT takers groups* in each cohort are SAT takers ($SATtaker = 1$) who were planning to apply to college.

Table 1.4 and Figure 1.5 provide mean characteristics for all sample students by group and cohort. The female ratio of each group is more or less 50% which is not so far different from the population female ratio of 50%. Parents education level is suggested in row 2-3. The percentage of college graduate fathers is slightly higher in the SAT takers groups for both cohorts. Except the mothers of non-takers in 2005 cohort, the ratio of college graduate mother is similar across groups. Household satisfaction is measured by a Likert scale from 1 to 5 to measure how students feel about their lives at home. Interesting fact is that SAT taking students are more likely to have higher household satisfaction in both cohorts. Considering the fact that most of Korean parents want their children to go to college, this correlation seems to be plausible as parents and their child consented to the student's career. In summary, the four groups are balanced in terms of observable pre-intervention characteristics except some characteristics such as vocational school ratio, household income, and parents education. I present regression results with and without covariates and interaction terms to show that the inclusion of those covariates does not significantly change the main results.

1.5.4 Time-varying Selection into SAT Taking

An issue in the composition of the SAT takers and non-takers groups is selection on SAT taking due to the policy changes. High school students decide to take SAT based on the consideration of their characteristics, preference on future plan, and marginal changes in benefit and cost of college education due to the policy reform. A student's SAT taking decision may have been related to the policy change. In practice, it is hard to tell whether the SAT taking decision by a student was made before or after the policy announcement or was made based on the consideration of the policy change. Willis and Rosen (1979) suggest that expected gains in life earnings, family-background, financial constraints, and tastes mainly influence the decision to attend college. Adopting this framework, it would be rare to find a student whose preference is almost indifferent between going to the college or working after high school graduation and the only critical factor for her to make college decision is the policy change.

Table 1.5 shows that there was no remarkable change of the SAT taking trend among the high school graduates measured by the proportion of SAT applicants before and after the policy change in 2008 out of population high school graduates. The proportion of SAT takers maintained mid-seventy percent out of population graduates during 2005 to 2009 which covers my two sample cohorts of 2005 and 2008. There, however, is a possibility of compositional changes in the two groups.

The test for time-varying selection into the policy is whether observable characteristics predict SAT taking and whether the coefficients change pre and post reform. When there exists a strong selection into SAT taking (thus into policy exposure) based on gender or family income my empirical model needs to account for selection on observables. I run a regression in which SAT taking decision is predicted by the interaction of time period and baseline characteristics such as gender, parents' college degree, and household income:

$$SATtakers = \beta_0 + \delta_0 post + \beta_1 \mathbf{X} + \delta_1 (post \times \mathbf{X}) + u. \tag{1.16}$$

Figure 1.6 shows that female students, students having college graduate mothers, and relatively poor students are more likely to take SAT after the policy reform. That is, there is time-varying selection into SAT taking related to gender, mother's college degree, and household income. Increase in female labor market participation rate around the global financial crisis may be one reason for the compositional changes in SAT taking behavior. According to the evidence of time-varying selection, I include interaction terms of ($\mathbf{X} \times post$ in the regressions that estimate the policy effects.

1.5.5 Selection into SAT Taking on Unobservables

Another identification concern is the possibility of selection into SAT taking decision on unobservable factors which affect both outcome values and SAT taking decision. This selection might bias my main results in some ways. My model predicts potential selection into SAT taking where students with higher ability to study have less opportunity costs of study and students with higher preference to college obtain higher marginal benefit. Thus, the overall direction of bias is ambiguous. For example, female students do better in GPA due to some unobservable characteristics different from male students. In Korean co-ed high schools, the proportion of male students is 30.6% (Humanities Program) and 53.9% (Natural Science Program) but only 29.6% and 50.8% are male at the top GPA grades, respectively[21]. If female students are better at GPA than male then the female students who were used to be underdogs in SAT race might have been motivated to start college admissions game when there is an increase of admission weight on GPA.

1.5.6 Balance in Potential Outcome Trends

My DD design requires that the trend of outcome variables observed in the non-takers group is the same trend that would have been observed in the SAT takers group in the absence of the grading policy change. This parallel trends assumption may be convincing if in Equation (1.17) the period fixed effect δ_0 and the average group fixed effect β_1 are invariant across the groups and time respectively, so that the difference in outcomes between groups is time invariant. In this study, parallel trends assumption implies that the changes or trends in resource allocation behavior of the SAT taking group in the 2008 SAT cohort would have been the same as that of the non-takers group in the 2008 SAT cohort if there was no grading policy shock. Then, δ_1 in Equation (1.17) would be the valid DD estimator.

[21] Jinhaksa, 2011-2014 high school graduates, number of observations = 252,316

I provide evidence to support the hypothesis of parallel outcome trends between groups. There are several methods to support the assumption used in the empirical studies by investigating the data. First, the balance in descriptive pre-intervention characteristics between groups can be shown to check the similarity of two groups respect to the various baseline features in average as discussed in Subsection 1.5.3. Second, the similarity in the pre-intervention outcome trends between the groups can make the assumption more plausible even though the assumption is about the trends in post-intervention counterfactual outcomes. It is useful to check the pre-intervention outcome data to predict the counterfactual post-intervention outcome trends. The more similar the pre-intervention outcome trends, we might be more easily convinced that the counterfactual outcome trend of SAT takers would be similar to the post outcome trend of the non-takers group. Unfortunately, visual inspection of multiple periods of pre-intervention outcomes is not available in my observational data. It includes only two cohorts data and the first observed period of the post-intervention cohort is the year of the policy implementation. An alternative inspection is done by comparing the eight years of post-intervention outcome trends after the high school graduation between the groups. The key assumption of this investigation is that the policy only affect the students' high school leisure allocation behavior and does not have any effects after graduation. I take up this issue in Subsection 1.7.1.

1.5.7 Empirical Model

The econometric approach for analyzing the impact of the policy change on student's behavior is given by a difference-in-differences model where I compare mean outcome for SAT takers group to that of non-takers group. I estimate

$$y = \beta_0 + \delta_0 post + \beta_1 SATtakers + \delta_1(post \times SATtakers) + \mathbf{X}\delta + \gamma(\mathbf{X} \times post) + u \quad (1.17)$$

where y is the outcome of interest such as time allocation. The DD estimator δ_1 captures the intervention effect of the policy on the outcomes of SAT takers group compared to the unexposed outcomes. δ_1 is considered to be zero in the absence of the policy under the proper assumptions. The interaction terms $\mathbf{X} \times post$ account for time-varying selection on observables. I cluster standard errors by the sample students' high school. The main estimation includes $\mathbf{X}$ which is a vector of controlled variables including baseline characteristics such as gender, parents' education level, household income which have been fixed at the time of intervention, and their interaction terms with post.

1.6 Results

1.6.1 Average Policy Effect

Table 1.6 Column (1) shows that the policy reduces SAT takers' weekday leisure by 0.5 hour a day. Considering that the mean outcome of leisure for all observations is 3.3 hours a day, this is a large negative effect on current leisure. This suggest that the level of current period - high school days - utility u_0 would have been worsen off after the policy. An interesting result in Column (2) indicates that SAT takers slept 12 minutes more at night after the reform. It is the only evidence that the reform would help ameliorate current utility. Column (3) indicates that the policy encouraged students to take more private lessons. An average SAT taker spent almost one hour more per week at cram schools after the policy while Column (4) shows a 1.2 hour increase in self study time. More private lesson and self study time imply that college applicants are worse off in terms of their current utility as a result of a decrease in current leisure l_0. They have invested more on college prep activities to enhance their GPA and SAT. Column (5) and Column (6) presents no significant impact on leader role take-up rate and academic club sign-up rate. Estimation without interaction terms in Appendix Table A.2 also shows the robust results. It is not sure how much the increased self study time was invested in GPA (g) and SAT (s). A student who has higher marginal productivity in SAT than GPA would have invested more on SAT study, and vice versa. This time assignment also depends on the exogenous changes in admission weights between SAT and GPA after the policy reform.

Recalling the fact that the main policy goal is to reduce welfare-harming admission competition, the primary analysis above implies that there is no evidence of increase in students' leisure which is a key ingredient of current welfare u_0[22]. A slight increase in sleep hours cannot fully explain the huge decrease in leisure. Thus, despite of the small increase in sleep, it is hard to say that the policy had successfully achieved the goal. Indeed, students had to study more during their high school days.

1.6.2 Subgroup Analyses

I investigate the heterogeneous effects of the grading policy by high school GPA by estimating the same models in Section 1.6.1 with interactions. Table 1.7 Panel A Column (1) and (2) shows no significant changes in leisure and sleep hours for the top GPA students (Grade 1 and 2). These can be reasonable results as they had already been enjoying less leisure and sleeping less than the average. My theoretical model expects that the saved time from the shorter leisure would have been invested to college admission preparation activities to optimize their lifetime welfare. Column (4) finds that the top GPA students invested more than an hour per day to self study while there are no significant changes in private lesson, leadership and academic club activities. The estimation result confirms that the current leisure was substituted to additional investment in self studying. Table 1.7 Panel B provides the regression results on the low GPA students (Grade 7, 8, and 9). Contrary to the top students, I do not find any significant changes in all outcomes of the students at the bottom

[22]I find the similar results in the alternative setting which compares SAT takers (regular) and rolling (susi) admitted students. See Appendix A Figure A.5 to A.8

three GPA grades (23 percentile from the bottom) with an increase in proportion of leadership role take. Table 1.8 shows that mid-GPA students reduced leisure about 0.3 hour and slept more by 0.3 hour per day. They also had an hour increase of self study similar to the top GPA students.

Overall, I find the heterogeneous effects depending on students' GPA. GPA contains information about student's unobservable inherent productivity and motivation and would be an increasing function of the two. Students may respond differently to the incentive changes according to their degrees of productivity and motivation. The empirical results show that the top and mid GPA students increased study hours more than the low GPA students after the reform. They might want to minimize the possibility of losing high-ranked college admissions to avoid the worst case scenario by studying harder under the uncertainties discussed in Section 1.3.3.

1.7 Robustness

1.7.1 Balance in Non-Intervention Outcome Trends

Unlike other DD studies where intervention effects last through following periods, the grading policy is only applied to the college applicants. After the admission competition, all students become adults and are free from the policy except some students who decide to prepare their second or third SAT. Since those retakers are not going back to the regular high school curriculum, their time consumption patterns are also considered to be one variation of early twenty's lifestyle. Thus, I compare the weekday leisure trends between the two groups in the two cohorts for eight years (from year $T+1$ to $T+8$) after the year of their first SAT which were taken in the last year of high school (T).

Figure 1.7 displays box plots which show that all four groups have the similar trends of weekday leisure after the year of SAT ($t = 0$). The median weekday leisure hours per day in the first year after the SAT ($t+1$) is about five hours in all four groups. This could possibly show that without the admission competition, The two groups have the similar time allocation patterns in average. From the second year after graduation, the trends are slightly going down for all groups as they are getting older to reach their late-twenty. Supplement time use trends of daily sleep hours in Figure 1.8 also shows similar patterns to Figure 1.7 across groups. Lack of sleep in adolescence may cause developmental problems such as learning, memory, attention, cognition, and emotion processing (Tarokh et al., 2016). Furthermore, insufficient sleep is significantly positively correlated with unsafe behaviors among youths such as risky driving, risky sexual activity, and drug-tobacco-alcohol use (Weaver et al., 2018).

1.7.2 Propensity Score Matching for Difference-in-Differences Model

An alternative approach to this observational study is to use matching method for balancing observable characteristics across groups. To deal with the selection problem in 1.5.4 and the threat to the balance in pre-treatment characteristics discussed in 1.5.3, I perform propensity score matching (PSM) to be incorporated in the DD estimation. The purpose of PSM is to reduce baseline heterogeneity and construct a valid counterfactual to address the selection bias. By following the definition of Rubin (2001), The propensity score of a student i in this research is defined as the probability of being treated or exposed to the college admission policy as a college applicant ($SAT_i = 1$ vs. $SAT_i = 0$) given the observed value of a vector of observed characteristics X_i for the student i. That is, the propensity score e_i is defined by

$$e_i \equiv e(X_i) \equiv Pr(SAT_i = 1 | X_i). \tag{1.18}$$

The biggest advantage of using propensity score is that treatment and control groups would have the same distribution of all observed characteristics after matching as long as those two groups show the same distribution of propensity scores. Regression models and matching methods aim to adjust confounding covariates which are associated with treatment and outcome. In this sense, I expect propensity score matching would help balancing pretreatment characteristics of control and treatment groups in my study.

Implementing propensity score matching in difference-in-difference strategy is widely used in health and medical research. Su et al. (2019) use kernel matching and radius matching methods to match the original panel data between treatment and control group in the baseline. Sari and Osman (2015) use radius matching to construct a control group after restricting the potential control group using key characteristics of the intervention group. Stuart et al. (2014) use a propensity score weighting strategy that weights the four groups define by time and intervention status to be balanced on a set of characteristics. After matching, DD estimation may avoid or minimize selection bias across group and time in an observational study by balancing pre-treatment observable characteristics between two the groups. That is, exactly matched students in treatment and control groups would have the balanced distributions of covariates. A benefit of the method by Stuart et al. (2014) is that it can be implemented with the repeated cross-sections data. Thus, I use the approach by Stuart et al. (2014) in this section. I define group 1 to 4 as follows:

$$Group = \begin{cases} 1 & \text{if 2005 SAT Takers} \\ 2 & \text{if 2008 SAT Takers} \\ 3 & \text{if 2005 Non-Takers} \\ 4 & \text{if 2008 Non-Takers} \end{cases}$$

In using matching methods, an identifying assumption of selection on observables should be satisfied to guarantee that observable covariates can capture all relevant differences between treatment and control groups. Under this ignorability assumption (Rosenbaum and Rubin, 1983), the difference in outcome variables between the matched treatment and control groups would be only by the intervention and the differences in baseline characteristics would be random. In addition, the propensity score for all samples should be strictly greater than zero and less than one. That is, for any given observed covariates we have a chance to observe either SAT taker or non-taker.

I construct multiple group propensity score weighting for all individuals in four groups categorized by time and policy exposure. Parents' college education, log household income, and gender are selected to calculate the propensity scores and balance the four groups. I use a multinomial logistic regression to predict the propensity of being in one of the groups as a function of covariates above. This gives four propensity scores to each student. The estimated four scores sum to one for each individual. Stuart et al. (2014) calculate the weights for individual i as

$$w_i = e_1(X_i)/e_g(X_i) \tag{1.19}$$

where g is group indicator defined above. Figure 1.9 shows the distribution of weights by sample groups. 2005 SAT takers all receive a weight of 1. Excluding outliers, the median weights are less than 2 for 2005 non-takers and 2008 SAT takers while the median weight of 2008 non-takers is less than 4. Figure 1.10 to 1.12 illustrate the kernel density of household income across groups before and after the matching. Note that the reference group is 2005 SAT takers shown by solid red line. As shown in Figure 1.13 I find that propensity score matching significantly reduces group differences. I run weighted DD regression models with covariates by using the weights calculated above to ensure that the four groups are weighted to be similar to the SAT takers group in the pre-period. Table 1.9 presents the robust results to the main outcomes in the Table 1.6.

1.8 Concluding Discussion and Policy Implications

I considered a college admission game in which high school students compete on the limited college admissions which are assigned in order of students' relative ranking. My model under the game of status suggested that the new policy of Coarse grading SAT and relative grading GPA introduced uncertainty in expected return to SAT investment and more competitiveness in GPA race which result in more investment in college prep activities.

This study did not find empirical evidence that coarsening SAT grades and changing GPA grading from absolute to relative resulted in welfare improvement. I found the unintended and negative impacts of the reform on student's leisure allocation together with increases in self study hours. The results suggest that education policy makers need to understand the students' behavior under status uncertainty and highly competitive environment in designing college admission policies.

Enjoying life is important for all age groups. Especially, children who are in their physical and emotional development stages need to consume enough time with their family and friends to build up their healthy life style. When a market fails to bring a decent childhood for them, Government policy can change people's incentives to move away from an outcome failing to achieve the maximized social welfare. This study, however, shows that without comprehensive understanding of what makes people do what they do the government intervention may not work well.

FIGURES AND TABLES

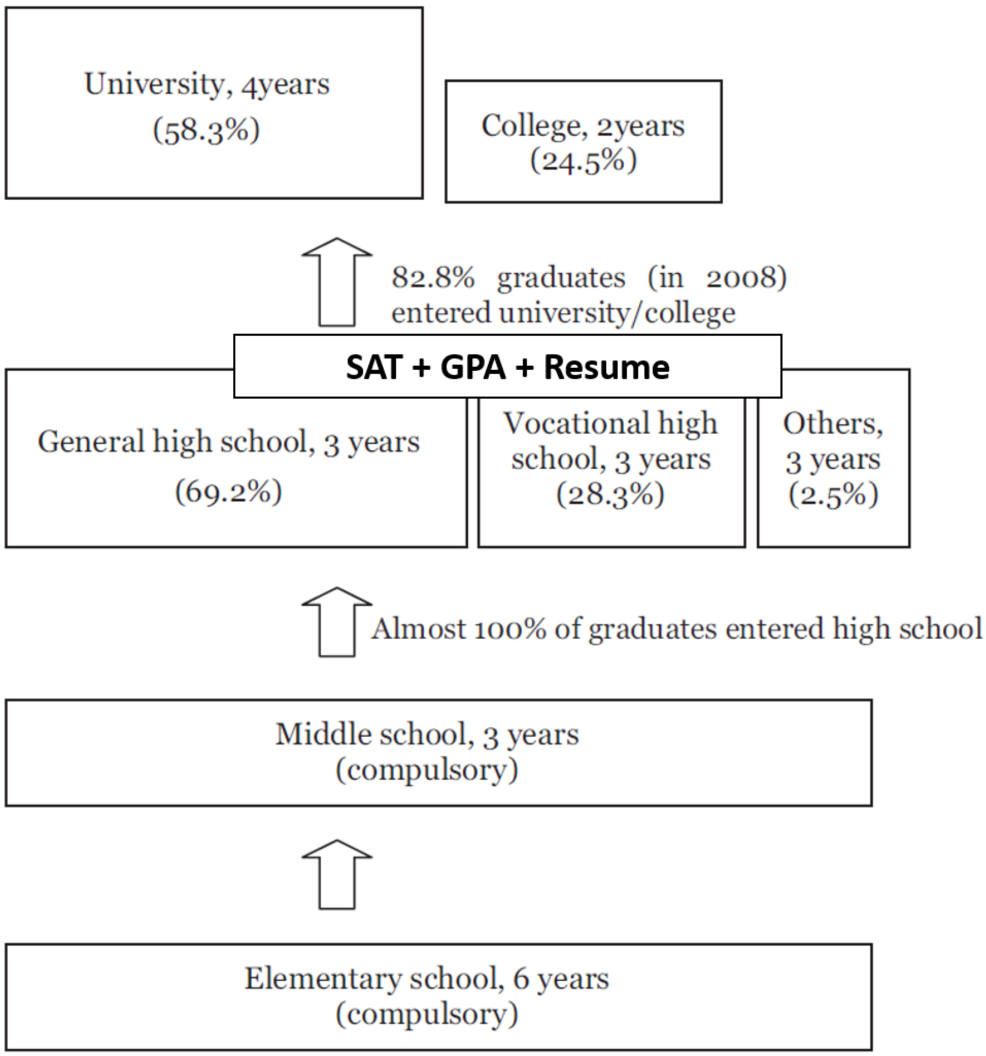


Figure 1.1: Korean Education System (Nam, 2014)

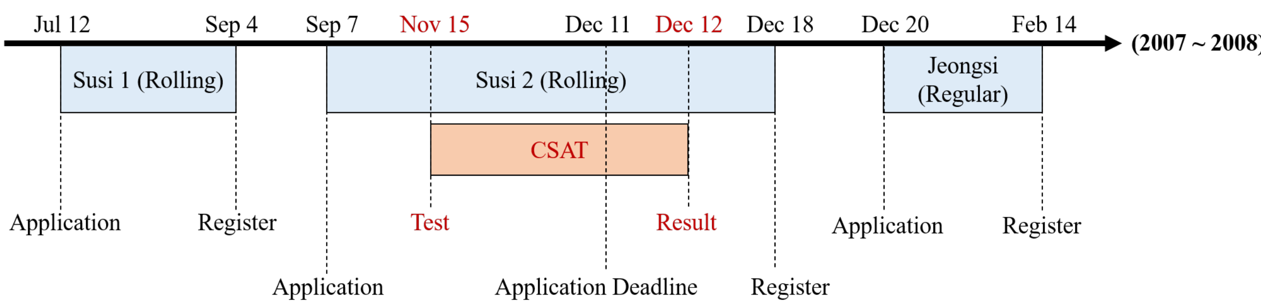


Figure 1.2: 2008 Academic Year College Admission Timeline

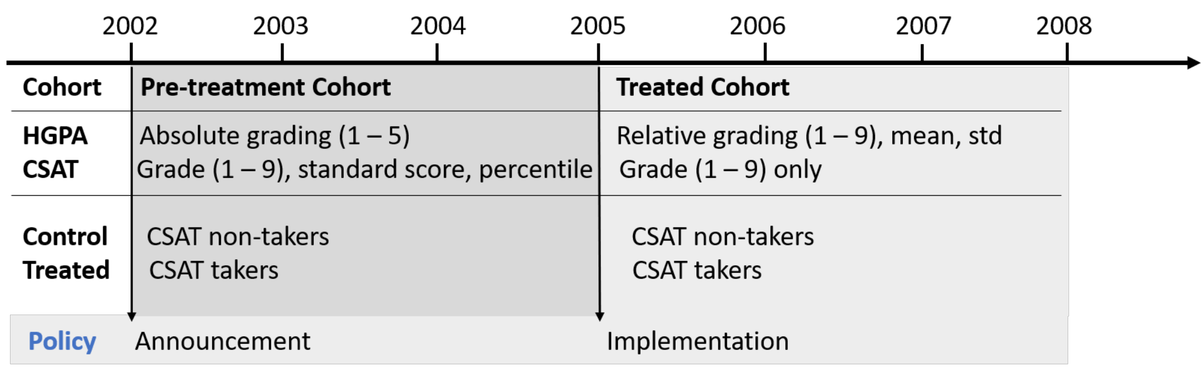


Figure 1.3: Policy Implementation Timeline

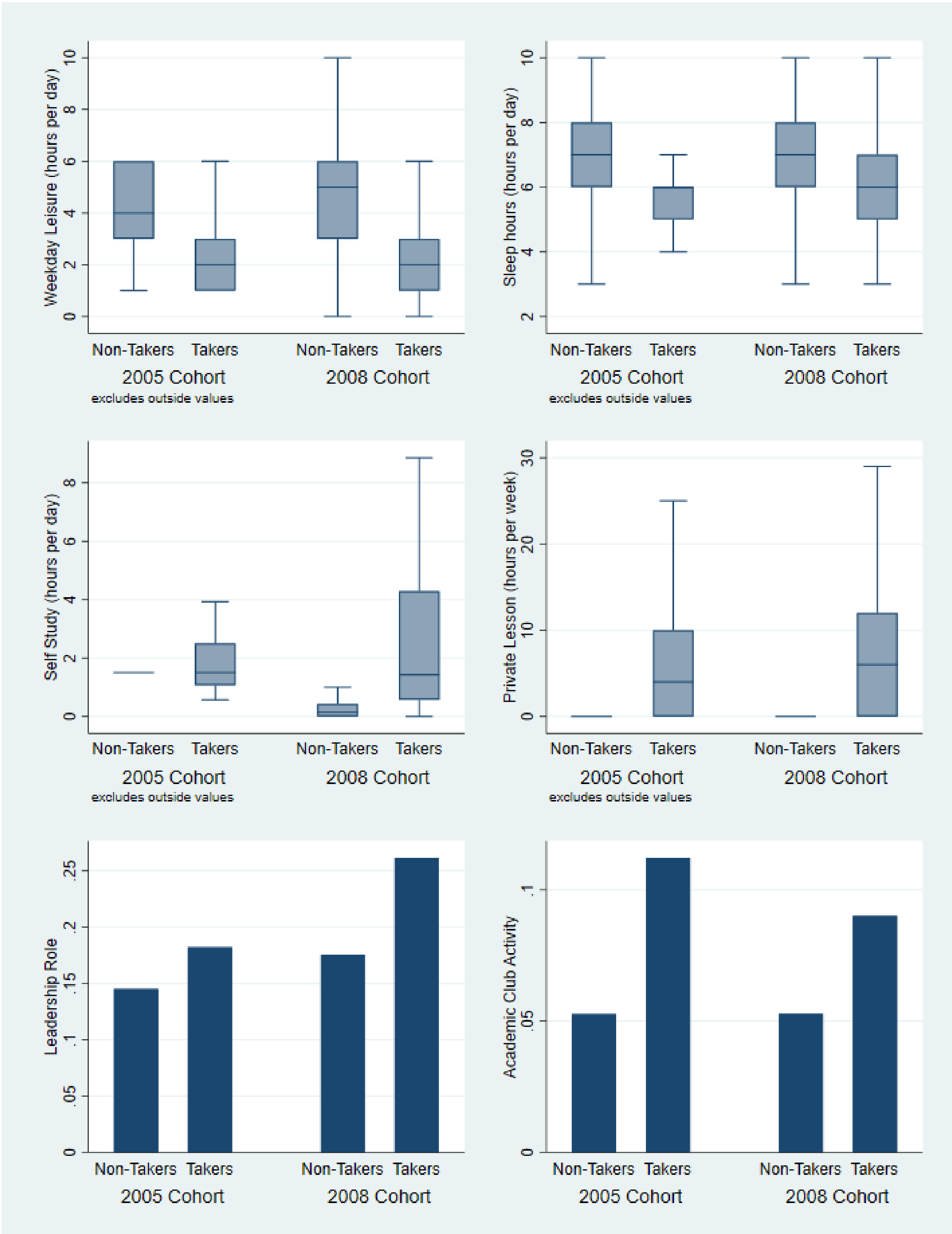


Figure 1.4: Mean Outcomes

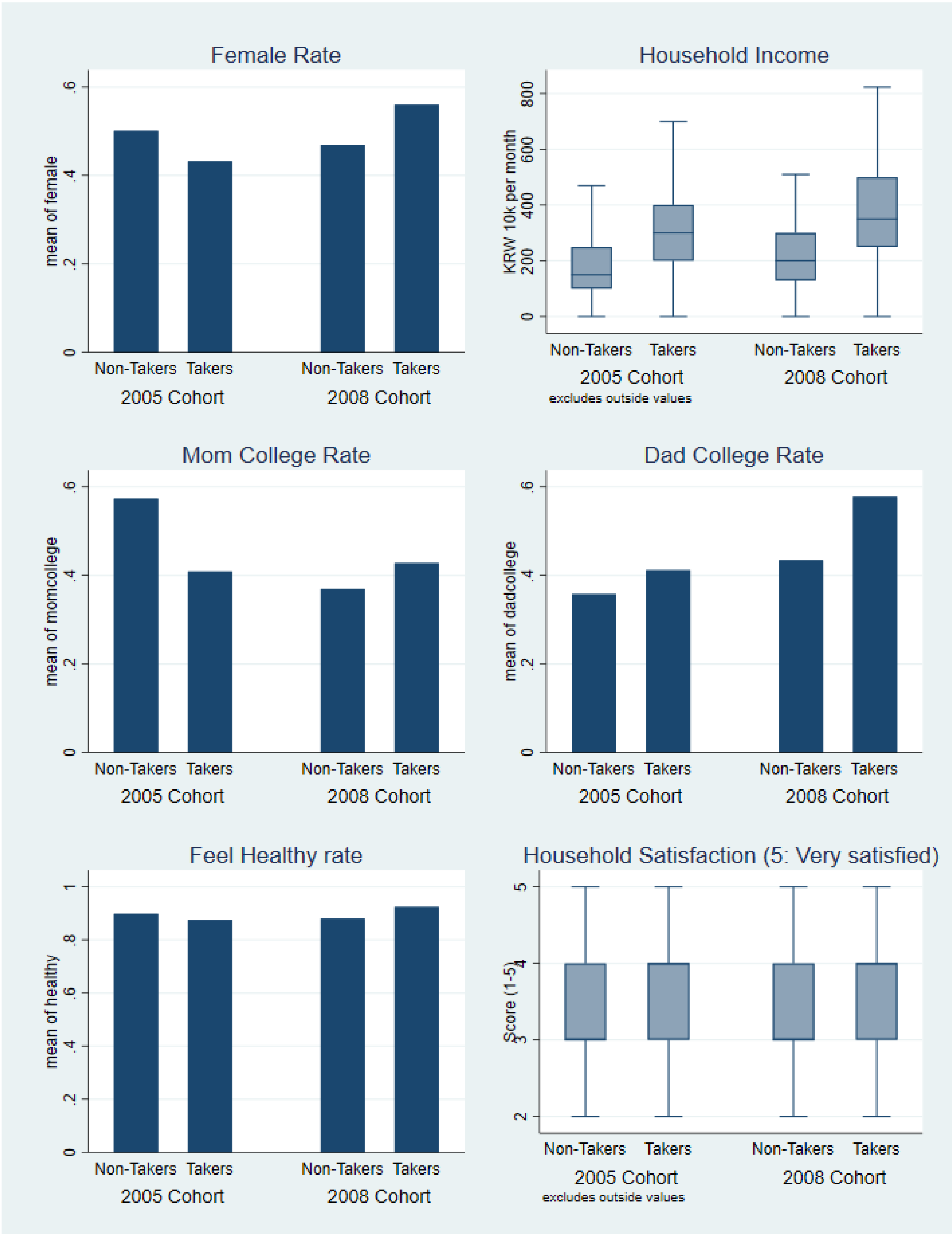
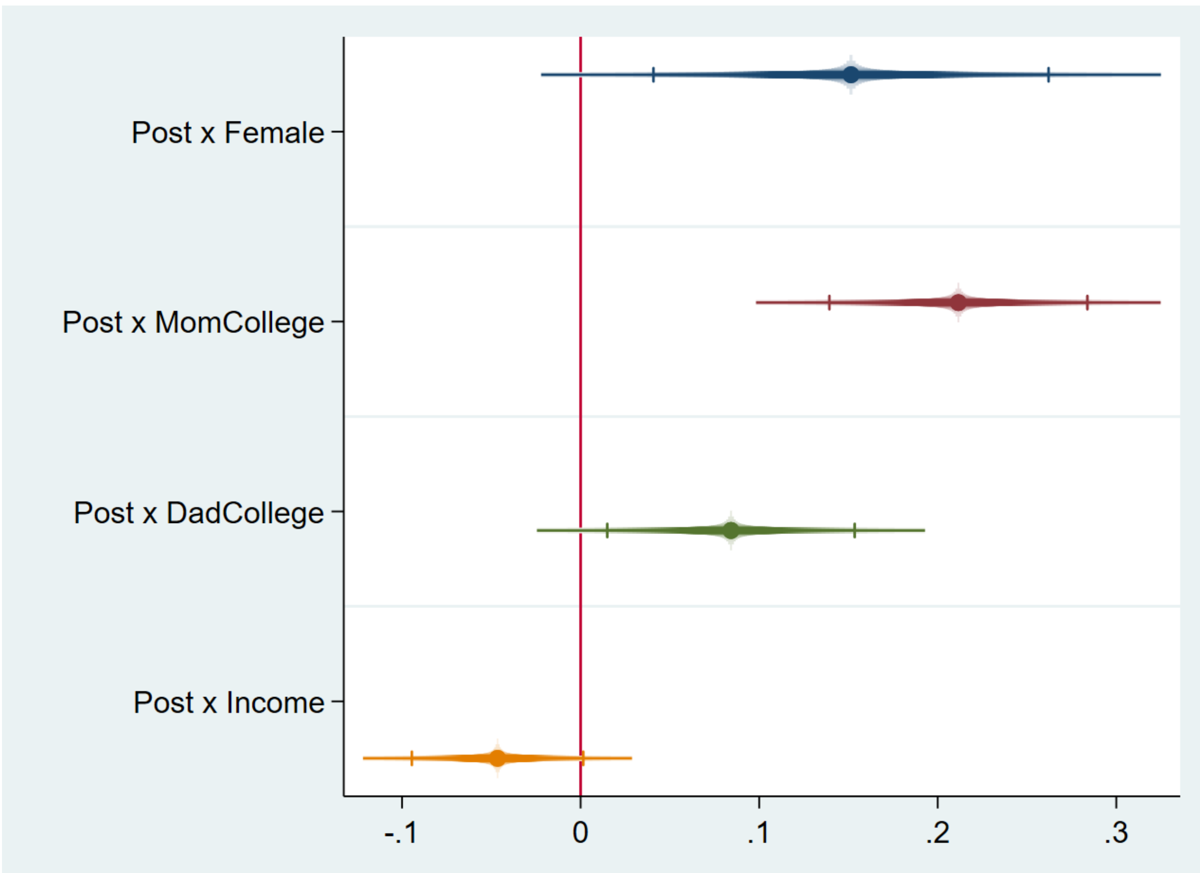


Figure 1.5: Mean Characteristics



Cluster standard errors by school in parentheses.
$p < 0.1$ for outside of inner tick, $p < 0.01$ for outside of the end of the bar.

Figure 1.6: Test for Pre-Post Selection Variation into SAT Taking

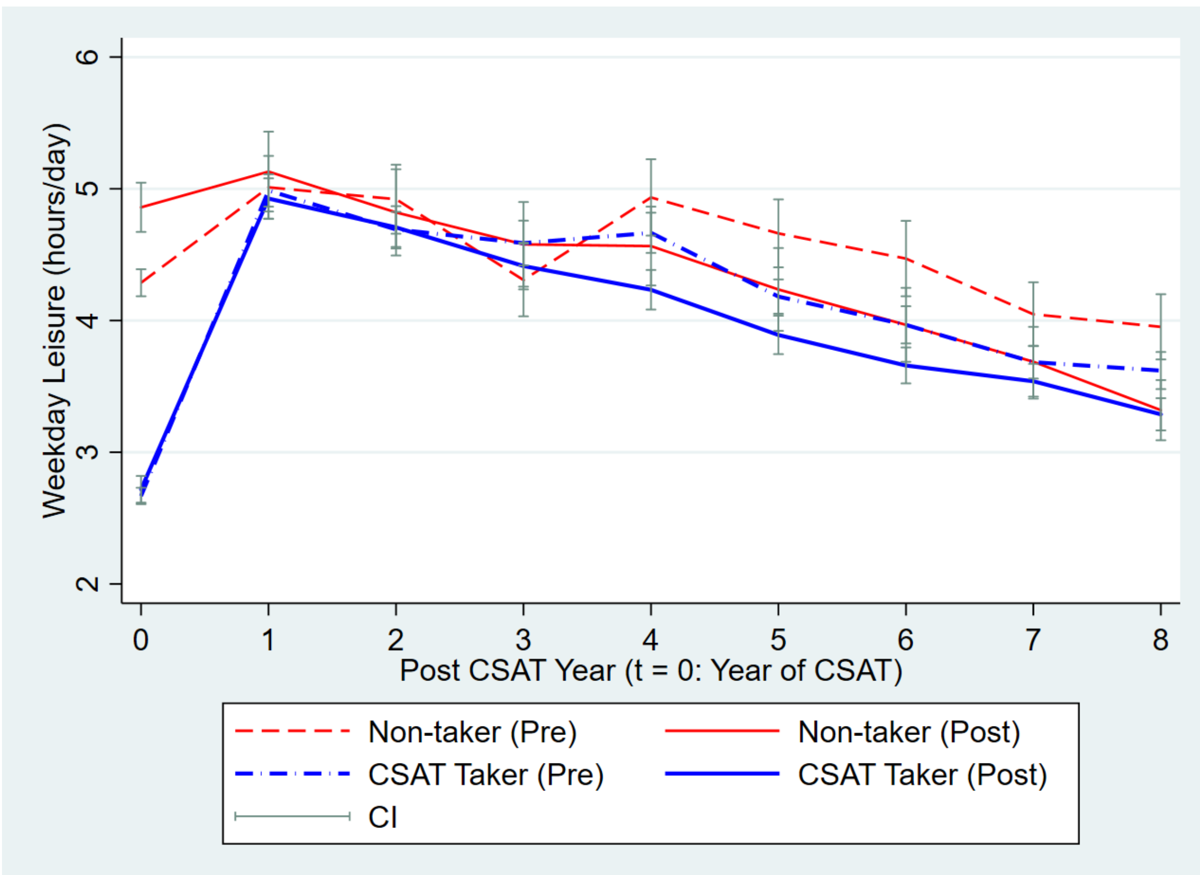


Figure 1.7: Non-Intervention Outcome Trends after SAT (hours/day)

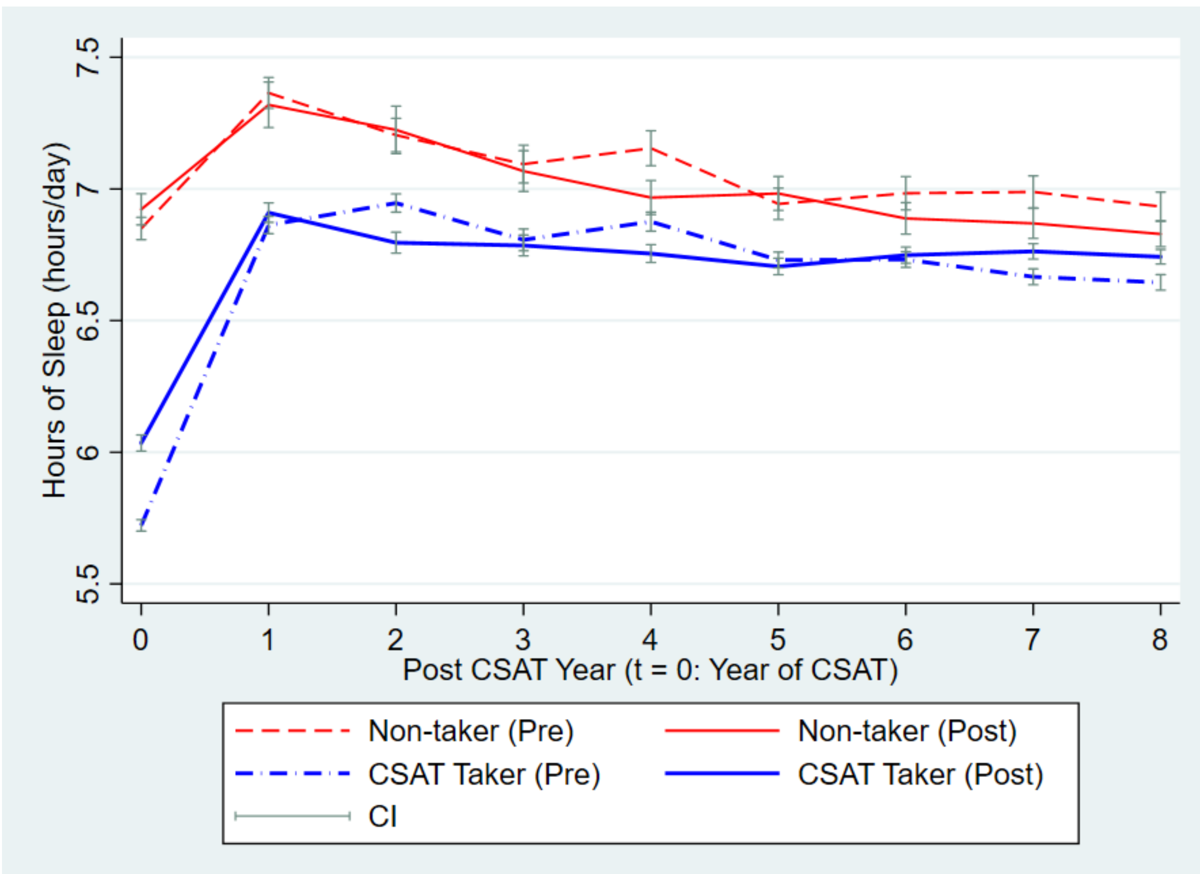
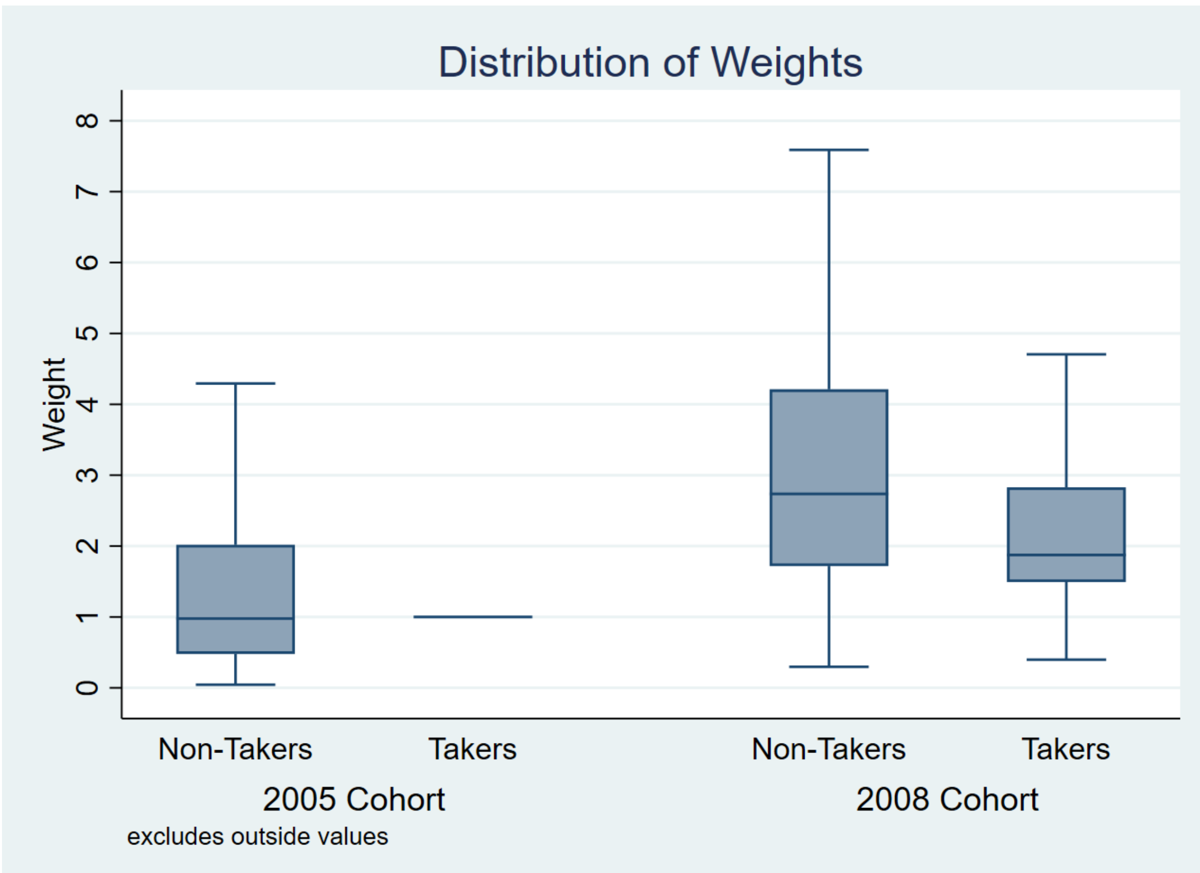


Figure 1.8: Non-Intervention Sleep Hours Trends after SAT (hours/day)



* Weights are calculated by $w_i = e_1(X_i)/e_g(X_i)$ (Stuart et al., 2014). By definition, 2005 SAT takers' weights are all 1.

Figure 1.9: Distribution of Weights

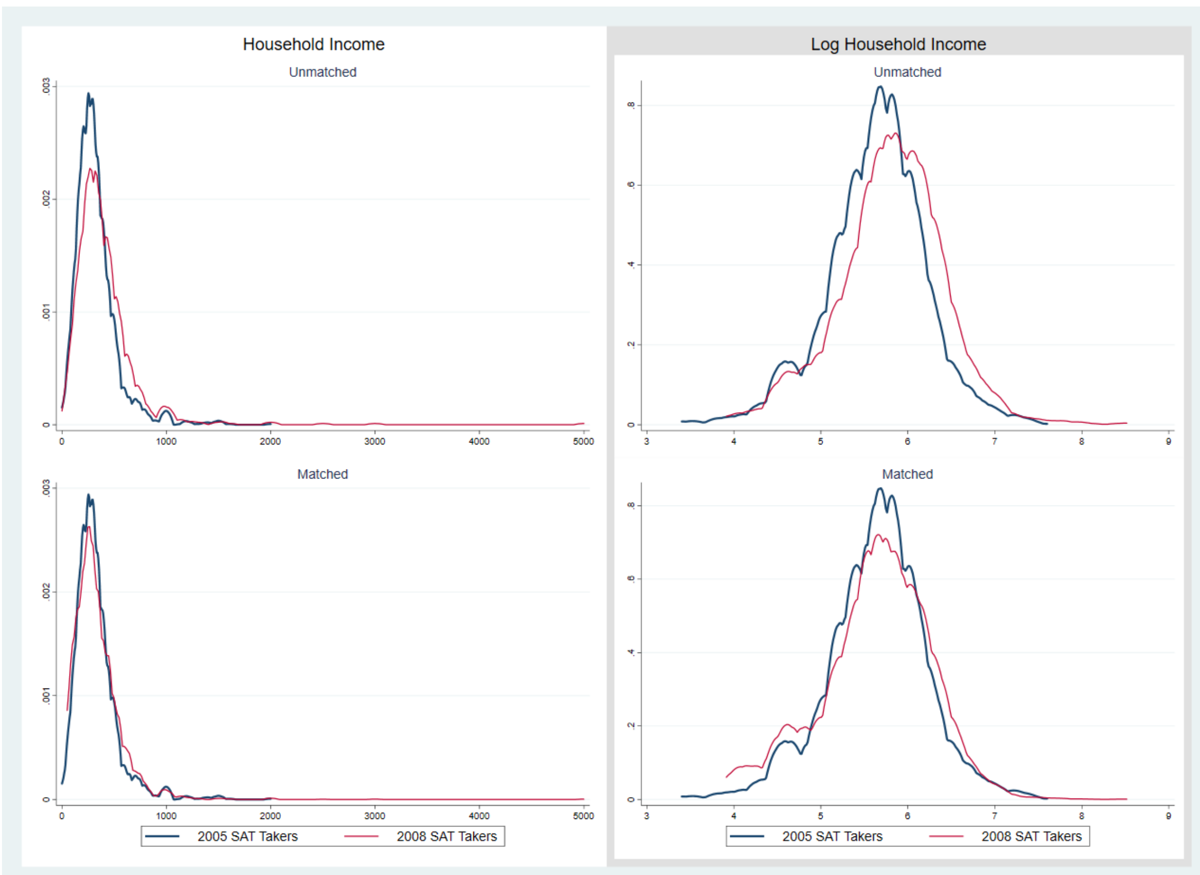


Figure 1.10: Density Plots of Household Income (Group 1 vs 2)

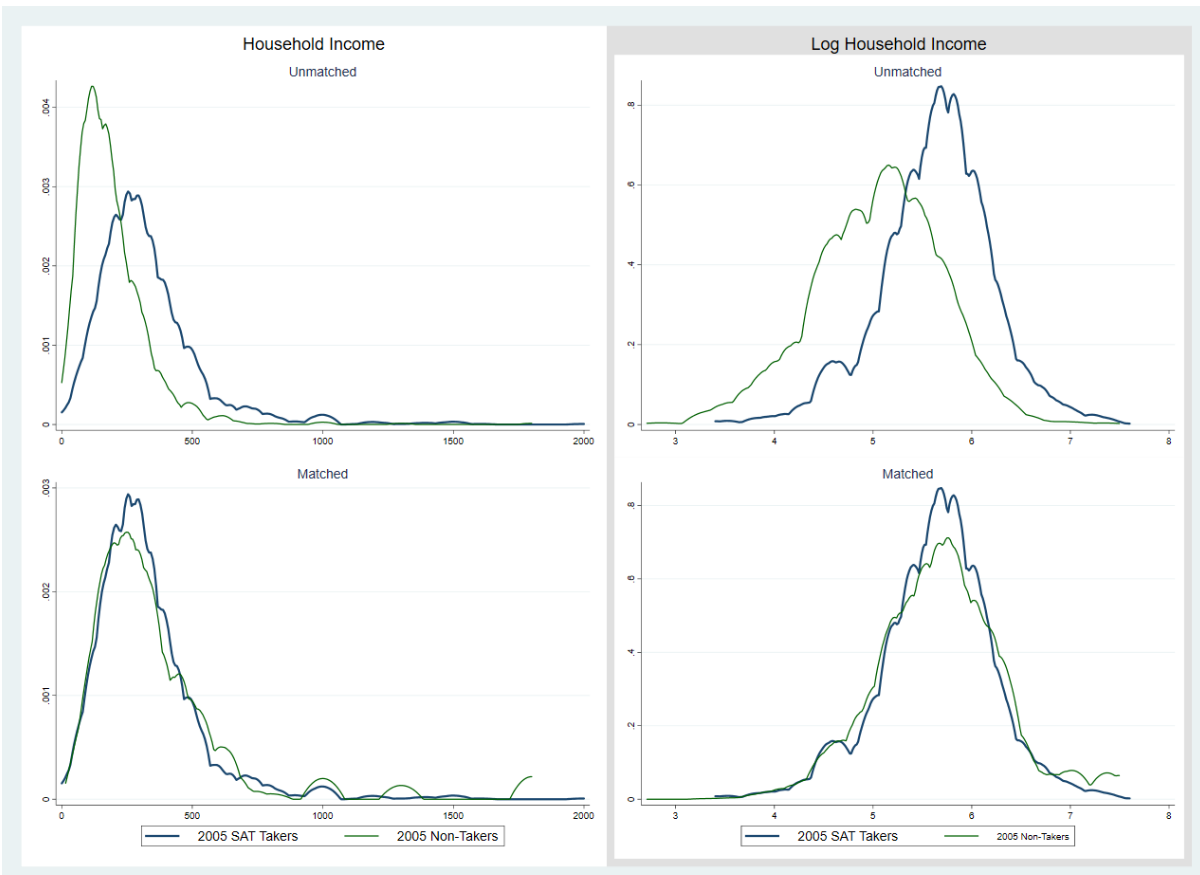


Figure 1.11: Density Plots of Household Income (Group 1 vs 3)

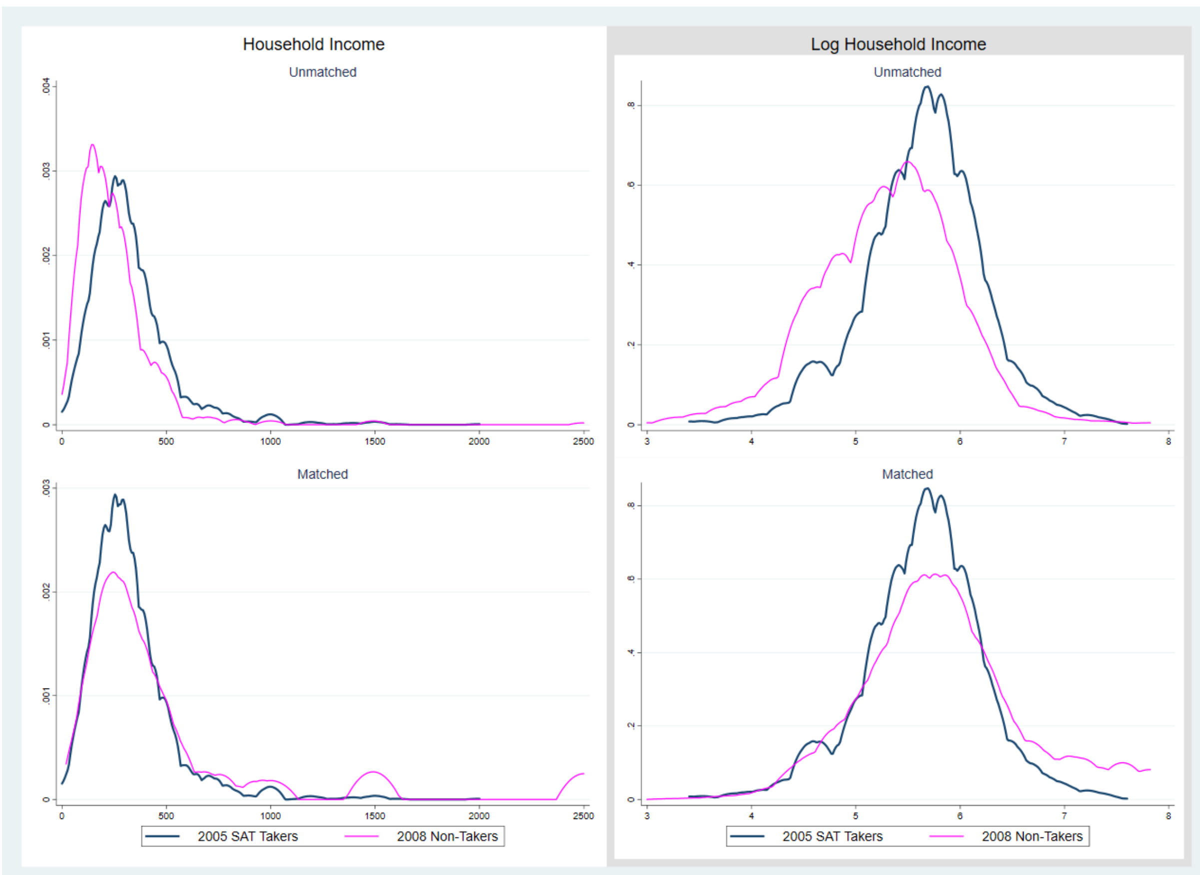
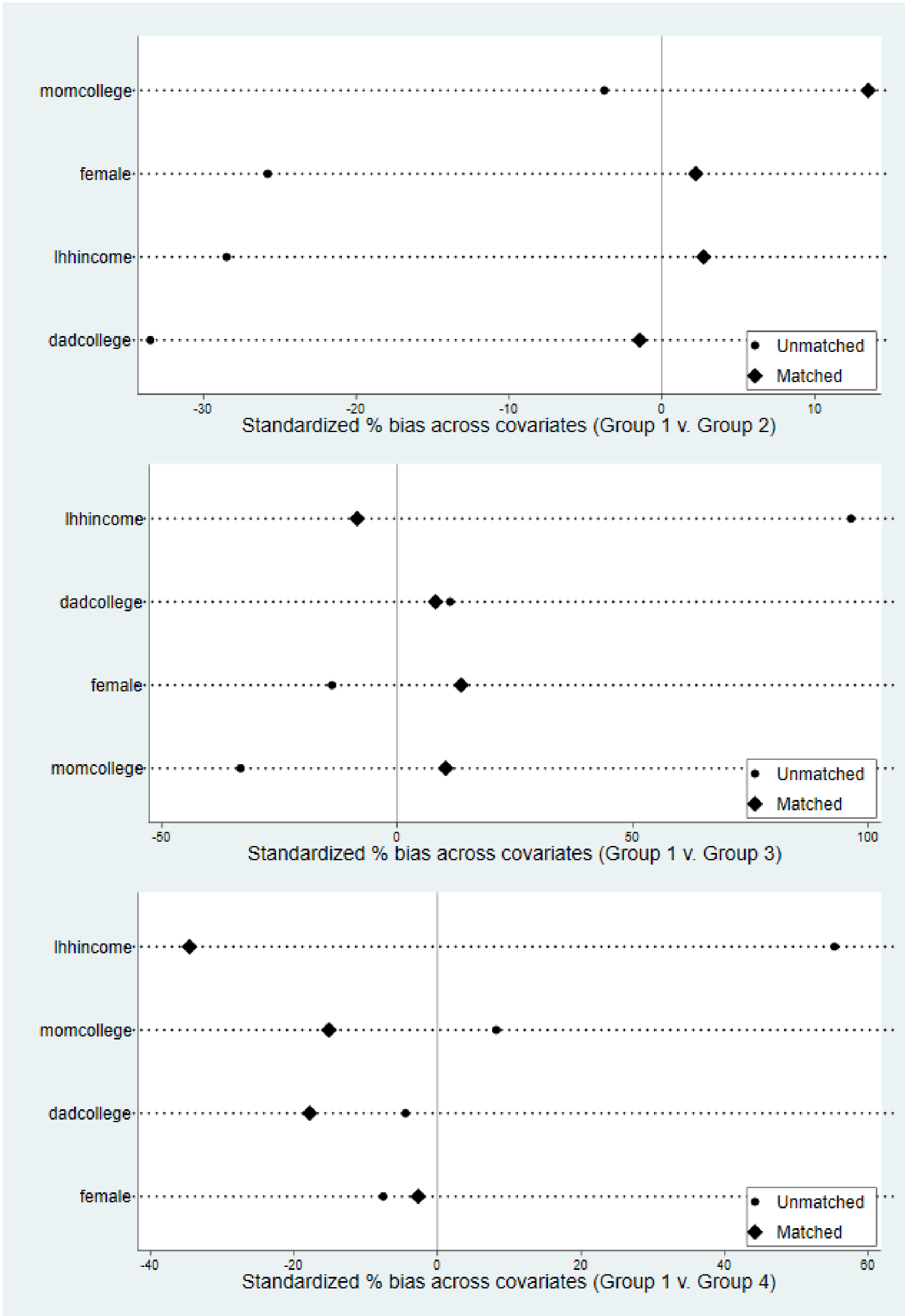


Figure 1.12: Density Plots of Household Income (Group 1 vs 4)



* The standardized % bias is defined by $100 \times (\bar{x}_1 - \bar{x}_g)/((s_1^2 + s_g^2)/2)^{1/2}$ where $\bar{x}_1$ is the sample mean of group 1 (2005 SAT Takers) and $\bar{x}_g$ is the sample mean of group g where $g = 2, 3, 4$ (group 2: 2008 SAT Takers, group 3: 2005 Non-Takers, group 4: 2008 Non-Takers). The bias above 20 in magnitude is considered to be significant difference (Rosenbaum and Rubin, 1985).

Figure 1.13: Propensity Score Matching Balance

Grade	Cumulative Percentage
1	4% or Under
2	4% ∼ 11%
3	11% ∼ 23%
4	23% ∼ 40%
5	40% ∼ 60%
6	60% ∼ 77%
7	77% ∼ 89%
8	89% ∼ 96%
9	Over 96%

[a] Source: Ministry of Education.
[b] Grade 1 is the highest. Students raw SAT scores are normalized. Each student's percentile is matched to one of nine preset grades.

Table 1.1: GPA and SAT Preset Grade Scale

Sample Groups	Pre	Post	Total
SAT takers	1,631 (63%)	856 (60%)	2,487
Non-Takers	951 (37%)	569 (40%)	1,520
Total	2,582 (100%)	1,425 (100%)	4,007

Table 1.2: SAT Takers (Regular) v. Non-Takers

		2005 SAT Takers				2008 SAT Takers			
Result	Subject	Obs	Mean	SD	Min / Max	Obs	Mean	SD	Min / Max
Grade	Korean	2,184	5.58	1.91	9 / 1	1,262	5.12	2.26	9 / 1
	Math	1,943	5.48	1.81	9 / 1	1,181	5.05	2.15	9 / 1
	English	2,172	5.59	1.91	9 / 1	1,246	5.11	2.30	9 / 1
Percentile	Korean	2,199	40.67	27.77	0 / 100				
	Math	2,199	37.20	28.81	0 / 100				
	English	2,199	39.93	28.24	0 / 100				
Score	Korean	2,199	92.95	22.00	0 / 135				
	Math	2,199	83.42	34.56	0 / 150				
	English	2,199	92.11	22.15	0 / 139				

[a] No percentile and standardized score data of 2008 SAT is available due to the grading policy. The population median grade is 5 for all subjects. Difference in observations between categories in 2005 SAT is due to missing data. In 2008 SAT, some students selectively took exams depending on the admission criteria of applying colleges.

Table 1.3: Sample statistics of SAT results (All Samples in Data)

	2005 Non-takers (N = 951)	2005 SAT Takers (N = 1,631)	2008 Non-takers (N = 569)	2008 SAT Takers (N = 856)
Female	50%	43%	47%	56%
Mother College	57%	41%	37%	43%
Father College	36%	38%	43%	58%
Household Income	189	326	237	403
Health Condition	3.54	3.52	3.66	3.77
HH Satisfactory	3.25	3.69	3.43	3.81
Vocational School	91%	14%	87%	20%
Punishment	17%	7%	13%	5%
Prize	36%	38%	22%	45%
Smoking	27%	8%	23%	6%
Drinking	63%	40%	69%	47%
Daily Gaming	28%	8%	29%	9%

[a] Health condition (5: Very healthy) and Household satisfactory level (5: Very satisfied) range from 1 to 5. Punishment at school, prize, smoking, and drinking are binary variables where they take 1 with any experience.
[b] The unit of household income is KRW 10,000 which is equivalent to USD 8.96 as of April 2021.

Table 1.4: Mean Characteristics by Group and Cohort

Year	Number of Graduates	Number of SAT Applicants
2001	736,171	603,238 (81.94%)
2002	670,713	541,662 (80.76%)
2003	590,413	482,089 (81.65%)
2004	588,550	476,129 (80.90%)
2005	569,272	435,538 (76.51%)
2006	568,055	422,310 (74.34%)
2007	571,357	425,396 (74.45%)
2008	581,921	446,597 (76.75%)
2009	576,298	448,472 (77.82%)
2010	633,539	532,436 (84.04%)

[a] Data source: KOSIS, KICE

[b] There was no propotional changes in SAT applicants out of population high school graduates around the policy changes from 2005 to 2009 implying that the average SAT taking decision was not affected by the policy.

Table 1.5: High School Graduates and SAT Applicants

	(1) Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study	(5) Leader Role	(6) Academic Club
Post	0.791***	0.142	-0.592	-1.320***	0.009	-0.013
	(0.164)	(0.111)	(0.400)	(0.116)	(0.029)	(0.022)
SAT Taker	-1.721***	-1.213***	5.514***	0.576***	0.036**	0.063***
	(0.079)	(0.069)	(0.311)	(0.051)	(0.017)	(0.012)
Post × SAT	-0.516***	0.197*	0.784	1.248***	0.043	-0.029
	(0.172)	(0.118)	(0.490)	(0.152)	(0.031)	(0.020)
N	4007	3998	4007	4007	4007	4007
Mean	3.280	6.166	4.588	1.719	0.189	0.085

[a] Cluster standard errors by school in parentheses.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Units are hours per day for leisure, sleep, and self study, hours per week for private lesson, and binary for leader role and academic club activities.
[d] Mean outcome is the average of total observations including both cohorts and groups.

Table 1.6: Average Policy Effects on Time Allocation (Interactions)

A. TOP GPA STUDENTS

	(1) Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study	(5) Leader Role	(6) Academic Club
Post	1.642***	0.386	0.623	-0.975***	0.005	-0.063
	(0.625)	(0.394)	(1.402)	(0.351)	(0.132)	(0.086)
SAT Taker	-1.721***	-1.351***	7.120***	0.904***	0.092	0.095*
	(0.284)	(0.268)	(0.971)	(0.175)	(0.089)	(0.054)
Post × SAT	-0.731	0.050	-1.470	1.107***	-0.032	-0.099
	(0.488)	(0.420)	(1.267)	(0.317)	(0.121)	(0.073)
N	332	332	332	332	332	332
Mean	3.145	6.059	5.669	1.971	0.370	0.111

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B. BOTTOM GPA STUDENTS

	(1) Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study	(5) Leader Role	(6) Academic Club
Post	1.283***	0.034	0.467	-1.487***	-0.071	-0.048
	(0.489)	(0.311)	(0.864)	(0.174)	(0.055)	(0.036)
SAT Taker	-1.270***	-0.937***	4.618***	0.219**	0.017	-0.001
	(0.183)	(0.140)	(0.609)	(0.096)	(0.034)	(0.023)
Post × SAT	-0.598	0.159	-0.329	0.637	0.104	0.055
	(0.597)	(0.390)	(1.427)	(0.420)	(0.086)	(0.057)
N	540	537	540	540	540	540
Mean	3.921	6.402	2.693	1.227	0.087	0.039

[a] Cluster standard errors by school in parentheses.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Panel A is for the top 11 percentile GPA students (Grade 1 and 2) and Panel B is for the bottom 23 percentile GPA students (Grade 7 – 9).
[d] Units are hours per day for weekday leisure and self study, hours per week for private lesson, and binary for leader role and academic club activities.
[e] Mean outcome is the average of total observations including both cohorts and groups.

Table 1.7: Policy Effects on Time Allocation by High School GPA (Interactions)

MID GPA STUDENTS (Interactions)

	(1) Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study	(5) Leader Role	(6) Academic Club
Post	0.501***	0.111	-0.921*	-1.312***	0.011	-0.003
	(0.168)	(0.116)	(0.491)	(0.136)	(0.034)	(0.026)
SAT Taker	-1.805***	-1.259***	5.505***	0.597***	0.017	0.067***
	(0.084)	(0.070)	(0.349)	(0.055)	(0.019)	(0.013)
Post × SAT	-0.269	0.290**	0.908	1.196***	0.029	-0.037*
	(0.180)	(0.120)	(0.563)	(0.161)	(0.035)	(0.022)
N	3135	3129	3135	3135	3135	3135
Mean	3.184	6.136	4.800	1.777	0.188	0.090

[a] Cluster standard errors by school in parentheses.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Regression results with interactions for the mid-class GPA students (Grade 3 and 6).
[d] Units are hours per day for weekday leisure and self study, hours per week for private lesson, and binary for leader role and academic club activities.
[e] Mean outcome is the average of total observations including both cohorts and groups.

Table 1.8: Policy Effects on Time Allocation (Mid GPA Students)

	(1) Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study	(5) Leader	(6) Academic Club
Post	0.364**	0.012	0.510	-0.732***	0.039	0.010
	(0.158)	(0.112)	(0.471)	(0.234)	(0.032)	(0.023)
SAT Taker	-1.756***	-1.209***	5.547***	0.520***	0.028	0.054***
	(0.085)	(0.072)	(0.325)	(0.050)	(0.018)	(0.015)
Post × SAT	-0.394**	0.260*	0.423	1.061***	0.041	-0.029
	(0.180)	(0.138)	(0.649)	(0.259)	(0.035)	(0.027)
N	3546	3538	3546	3546	3546	3546

[a] Cluster robust standard errors by school in parentheses.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Units are hours per day for leisure, sleep, and self study, hours per week for private lesson, and binary for leader role and academic club activities.

Table 1.9: Propensity Score Matching DD Estimates

Chapter 2

NONPARAMETRIC APPROACH TO THE ADMISSION POLICY EFFECTS ON STUDENTS' TIME ALLOCATION

This chapter extends the college admission policy analysis by employing various nonparametric methods such as changes-in-changes and quantile difference-in-differences estimators introduced by Athey and Imbens (2006). The result from each extension model is consistent with the result from the standard difference-in-differences model. The average policy effect on leisure is 5 hours decrease per week and the result is consistent with the result from the standard difference-in-differences model. The policy treatment effect differs across groups categorized by outcome levels. The CIC estimates show that hard working students reduced leisure by 7 hours a week which is larger impact than the average working students.

2.1 Introduction

The average effect of the nationwide grading policy reform in Chapter 1 can be estimated by the standard difference-in differences (DID) model. Policy makers sometimes would like to predict the intervention effects on the whole distribution of recipient. The average effect is sensitive to the changes in extreme values and hard to answer which person is affected more. In an extreme case, there might be a case that the mean value does not represent anyone of her group. Also, the DID results may be functional form sensitive. In this context, this chapter employs various nonlinear models such as changes-in-changes (CIC) and quantile difference-in-differences (QDID) estimators introduced in (Athey and Imbens, 2006) to extends the college admission policy implication analyzed in Chapter 1. The college admission policy is not only important to the mean characteristic students, but also crucial to the every single student on the whole distribution. The main benefit of the CIC method is the fact that it can recover the entire counterfactual distribution of outcomes that would have been experienced by the SAT takers group in the absence of the policy reform. The result from each extension model is consistent with the result from the standard DID model. Furthermore, the intervention effect on students' time allocation behavior differs across groups categorized by their outcome levels.

2.2 Nonparametric Difference-in-Differences Estimation

In a non-experimental empirical work, identifying the counterfactual outcome values for the treatment group in the post-treatment period is the key step to estimate the treatment effect of an economic shock. The standard Difference-in-Difference (DID) model aims to estimate the average treatment effect on the treated group by employing the common trend assumption across groups. Many researchers, however, are interested in heterogeneous effects of intervention to acquire information about the distributional changes in outcome values given some observable characteristics. That is, we want to know how responses to a policy change varies across recipients. This information helps us to understand how those economic programs work. I identify following two empirical estimators and apply them to an empirical work which aims to estimate the college admission policy effects in Korea. The Quantile Difference-in-Difference (QDID) model is a type of DID approach to estimate the intervention effect on each quantile rather than on the average. Another approach to estimate the whole unobserved counterfactual outcome distribution or the distribution of treatment effects is the Changes-in-Changes (CIC) model developed by Athey and Imbens (2006).

2.2.1 Theoretical Framework

I use the potential outcome framework by Rubin (1974) and the notation in Athey and Imbens (2006) for the rest of analysis. The main idea of the approaches discussed here is to estimate the intervention effect on the intervention group by identifying the hypothetical counterfactual difference between the unobservable non-exposed outcome and the observed intervened outcome in the intervention group in the post-intervention period. There exist four subgroups classified by group (G_i) and time (T_i) where $G_i \in \{0, 1\}$ and $T_i \in \{0, 1\}$. $G = 0$ is the SAT non takers group and $G = 1$ is the SAT takers group. The data points are observed in time period $T_i \in \{0, 1\}$ where the period 0 is pre-intervention period (2005 SAT cohort) and the period 1 is post-intervention period (2008 SAT cohort). Each individual i's outcome in one of the groups is denoted by $Y_{gt,i}$. Let Y_i^N be the outcome without intervention and Y_i^I be the exposed outcome. Thus, what we observe is the realized outcome

$$Y = I \cdot Y^I + (1 - I) \cdot Y^N \tag{2.1}$$

where I is an indicator for the intervention. In the two-group-two-period model, the indicator $I = G \cdot T$ and the observed data are (Y, G, T, X). Thus, we have four subgroups in population and samples and only the intervention group in the period 1 is treated ($I = 1$). Individual characteristics $X_{gt,i}$ can be added to the regression equation to reduce selection bias. The distribution of $Y_{gt,i}$ can be summarized by its cumulative distribution function (CDF) $F_{Y,gt}$. Econometrician observes four CDFs $F_{Y,00}$, $F_{Y,01}$, $F_{Y,10}$, and $F_{Y,11}$ from the four repeated cross-sectional groups (pre vs. post; SAT takers vs. non-takers). In this study the Y_{11} is the post policy SAT takers who actually were exposed to the intervention.

The standard DID model estimates the difference between the average changes for the intervention group and the unexposed group overtime. The issue in a observational study is that for the intervention group, the untreated outcome of the intervention group in the period 1 Y_{11}^N is unobservable as all subjects in the intervention group are exposed to the policy reform in $T = 1$. The standard DID approach identifies the counterfactual outcome Y_{11}^N by assuming that the trend of outcome values across the groups after the intervention would have been the same if there was no intervention. This assumption allows us to identify Y_{11}^N using the observable values in the data.

Contrary to the standard DID model, CIC model allows the changes in the outcome distribution of SAT takers and non-takers group over time. This change is assumed to be attributed to the changes in the effect of unobservable characteristics on outcomes over time. As long as the data preserve several identification assumptions to be introduced in next section, CIC model estimates the intervention effect without parallel trends assumption.

2.2.2 The Changes-in-Changes Model

The CIC estimator is derived by across group comparison according to the outcomes and across time comparision according to the quantiles. That is, for a given outcome value of pre-intervention student y_{10}, the corresponding quantile in the SAT takers group $q = F_{Y,10}(y_{10})$ would be the reference quantile for recovering the counterfactual outcome value $y_{11}^N(q)$ while y_{10} will be the reference outcome for indicating corresponding quantile q' in the non-takers group distributions which will be used to find the comparison outcome values $y_{00}(q')$ and $y_{01}(q')$ to be used to construct the difference over time in the control group at the quantile q' which can be written by

$$\Delta^{CIC} = F_{Y,01}^{-1}(q') - F_{Y,00}^{-1}(q') = F_{Y,01}^{-1}(F_{Y,00}(y_{10})) - y_{10}. \tag{2.2}$$

Note that Δ^{CIC} can be identified only by using the observable data points. Equation 2.2 yields the counterfactual outcomes of intervention group in the post period without intervention for a given quantile q

$$y_{11}^{N}(q) = y_{10} + \Delta^{CIC} = F_{Y,01}^{-1}(F_{Y,00}(y_{10})). \tag{2.3}$$

Figure 2.1 graphically shows how to find the counterfactual outcome value of intervention group in the post period. This transformation only requires observable outcome values and quantiles. Then the average policy effect is simply expressed by the post average outcome difference between the intervention group and the transformed counterfactual outcome value. In a continuous CIC model the effect of the policy on a quantile q of the distribution of $F_{Y,10}(y)$ is calculated as

$$\tau_q^{CIC} \equiv F_{Y^I,11}^{-1}(q) - F_{Y^N,11}^{-1}(q) = F_{Y^I,11}^{-1}(q) - F_{Y,01}^{-1}(F_{Y,00}(F_{Y,10}^{-1}(q))). \tag{2.4}$$

This illustrates the transformation of outcome y_{10} to the corresponding counterfactual outcome y_{11}^{N} by only using the observable outcome values. The upper panel shows the control group distributions before and after the policy intervention. Starting from an observed outcome y_{10}, the treatment effect on quantile q of the distribution of $F_{Y,10}$ can be point identified under the conditional independence assumption. The solid white line in the bottom panel is the restored CIC counterfactual CDF of Y_{11}^{N}. The white dash-dot line in the bottom panel shows the counterfactual distribution of Y_{11}^{N} calculated by the QDID approach. QDID model simply shows DD estimation for each quantile. I implemented CIC and QDID estimation in STATA 15.1 using the {cic} package written by Kranker (2019).

Change-in-changes estimator also relies on several assumptions. I briefly review the identification assumptions for the counterfactual distribution of outcome applied to the empirical analysis in this chapter.

ASSUMPTION 1. (Assumption 3.1 in Athey and Imbens (2006)) *The outcome of an individual in the absence of intervention satisfies the relationship* $Y^N = h(U, T)$.

Assumption 1 establishes a relationship between the outcome Y and the unobservable random variable U and time T. The outcome production function $h(\cdot)$ implies that outcomes are determined by a single index U which summarizes all relevant unobservables and not directly affected by group assignment. I interpret the index U as the maximum daily leisure hours a student want to enjoy given the individual ability, motivation to study, and opportunity cost of time. This assumption implies that regardless of the treatment assignment individuals with the same unobservable characteristics in the same period have the same outcome values. Whether this leisure production function is the same across SAT takers and non-takers should be tested. I assume for now that the assumption is valid.

ASSUMPTION 2. (4.2) *The production function* $h(u, t)$*, where* $h : \mathbb{U} \times \{0, 1\} \to \mathbb{R}$*, is nondecreasing in* u *for* $t \in \{0, 1\}$.

In this study, high school students' unobservable characteristics which may affect their economic behavior can be ability, mental and physical health, and motivation. We may think that for example a higher ability increases productivity of time use on study and thus leads to a higher leisure other things being equal.

ASSUMPTION 3. (3.3 and 4.3) U_0 *and* U_1 *are continuous and we have time invariance within groups:* $U \perp T \mid G$.

Any changes in the variance of outcomes over time within a group are due to changes over time in the production function $h(\cdot)$ and any differences between the groups are stable over time. This implies that the relationship between willingness to enjoy leisure and motivation, ability, and mental and physical health is stable over time.

Under the assumptions above, the effect of the treatment on quantile q of the distribution of $F_{Y,10}(y)$ in a continuous CIC model[1] is calculated as

$$\tau_q^{CIC} \equiv F^{-1}_{Y^I,11}(q) - F^{-1}_{Y^N,11}(q) = F^{-1}_{Y^I,11}(q) - F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q))). \tag{2.5}$$

Also, Athey and Imbens (2006)[2] shows that the tight bounds of the counterfactual outcome distribution $F_{Y^N_{11}}(y)$ in the discrete CIC model can be identified by

$$F^{LB}_{Y^N,11}(y) \leq F_{Y^N_{11}}(y) \leq F^{UB}_{Y^N,11}(y) \tag{2.6}$$

where

$$\begin{cases} \text{for } y < inf\mathbb{Y}_{01}, & F^{LB}_{Y^N,11}(y) = F^{UB}_{Y^N,11}(y) = 0, \\ \text{for } y > sup\mathbb{Y}_{01}, & F^{LB}_{Y^N,11}(y) = F^{UB}_{Y^N,11}(y) = 1, \\ \text{for } y \in \mathbb{Y}_{01}, & \begin{array}{l} F^{LB}_{Y^N,11}(y) = F_{Y,10}(F^{(-1)}_{Y,00}(F_{Y,01}(y))) \text{ and} \\ F^{UB}_{Y^N,11}(y) = F_{Y,10}(F^{-1}_{Y,00}(F_{Y,01}(y))). \end{array} \end{cases} \tag{2.7}$$

Note that an inverse distribution function is defined by $F_Y^{-1}(q) = inf\{y \in \mathbb{Y} : F_Y(y) \geq q\}$ while $F_Y^{(-1)}(q)$ is an alternative inverse distribution function defined by $F_Y^{(-1)}(q) = sup\{y \in \mathbb{Y} \cup \{\infty\} : F_Y(y) \leq q\}$. The last equality of Equation 2.7 is exactly the same distribution function as the distribution of counterfactual outcome in a continuous outcome setup. That is,

$$F^{UB}_{Y^N,11}(y) = F_{Y^N,11}(y) = F_{Y,10}(F^{-1}_{Y,00}(F_{Y,01}(y))). \tag{2.8}$$

Athey and Imbens (2006) also show that the bounds can be tightened to the point identification in the discrete CIC model by (i) adding

---

[1]See Theorem 3.1 of Athey and Imbens (2006).

[2]See Theorem 4.1 of Athey and Imbens (2006).

ASSUMPTION 4. (4.4) *Conditional Independence: We have* $U \perp G \mid Y, T$.

or by (ii) introducing observable covariates (X) with

ASSUMPTION 5. (4.5) *Covariate Independence: We have* $U \perp X \mid G$.

In this paper, my main models estimate the CIC estimators in a continuous outcome framework. I also estimate the policy effects in the models with discrete outcomes and the models with alternative control group setting and continuous outcomes.

2.2.3 Quantile Difference-in-Difference Model

Another model to restore the counterfactual outcomes for estimating intervention effects is the quantile difference-in-difference model (QDID). The idea of the QDID model captures the outcome difference over time in the control group at an interested quantile q by using the formula

$$\Delta^{QDID} = F^{-1}_{Y,01}(q) - F^{-1}_{Y,00}(q). \tag{2.9}$$

Recall that Equation 2.2 uses the same outcome value $y_{10} = F^{-1}_{Y,10}(q)$ to find the quantile q' in the control groups to identify $\Delta^{CIC} = F^{-1}_{Y,01}(q') - F^{-1}_{Y,00}(q')$. The QDID model does not use y_{10} to find the corresponding quantile in control group, but uses the same quantile q as that in the treatment group. Equation 2.9 is also considered to be the change in the treatment group over time when there would be no intervention. That is, for each $y_{10}(q)$ at quantile q we have the counterfactual outcomes

$$y^N_{11}(q) = F^{-1}_{Y^N,11}(q) = F^{-1}_{Y,10}(q) + \Delta^{QDID}. \tag{2.10}$$

Then we obtain the treatment effect at a quantile q as

$$\tau_q^{QDID} \equiv F^{-1}_{Y^I,11}(q) - F^{-1}_{Y^N,11}(q) \tag{2.11}$$

$$= F^{-1}_{Y^I,11}(q) - (F^{-1}_{Y,10}(q) + F^{-1}_{Y,01}(q) - F^{-1}_{Y,00}(q)) \tag{2.12}$$

$$= (F^{-1}_{Y^I,11}(q) - F^{-1}_{Y,10}(q)) - (F^{-1}_{Y,01}(q) - F^{-1}_{Y,00}(q)). \tag{2.13}$$

Also recall that the CIC treatment effect is $\tau_q^{CIC} \equiv F^{-1}_{Y^I,11}(q) - F^{-1}_{Y^N,11}(q) = F^{-1}_{Y^I,11}(q) - F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q)))$. In this QDID model, we only consider a quantile q in both groups to capture the four outcome values to calculate the treatment effect. Thus, we see that the treatment effects in the two models are defined in different ways depending on their definition of reference points across time and group.

2.3 Data

I use the same data with the study in Chapter 1 which is sampled from the Korean Education and Employment Panel Survey (KEEP) data surveyed by Korea Research Institute for Vocational Education & Training (KRIVET). In this study, I choose two cohorts before and after the policy change to investigate their time use in high school. Thus, my four group samples from the different two cohorts across pre and post policy reform are regarded as repeated-cross sectional data.

Table 2.1 shows the summary statistics of weekday leisure at each percentile in SAT takers groups and non-takers groups of two cohorts. I find that there exists a weakly increasing tendency of leisure in time for their mean and each quantile outcome values. That is, the mean outcome of 2008 SAT takers is 2.71 hours per day which is slightly higher than 2005 SAT takers mean (2.67 hours per week). Non-takers means were also increased from 4.29 to 4.84. The distributions of weekday leisure for non-takers group in 2005 and SAT takers group in 2008 are skewed to the right while that for non-takers group in 2008 and SAT takers group in 2005 are skewed to the left.

2.4 Results

Table 2.2 presents the continuous CIC regression results[3] of the policy effect on SAT taking students' leisure by outcome percentiles at 10th, 30th, 50th (median), 70th, and 90th with covariates of gender, parents' education, and household income. Column (1) shows the robust results to the DD estimation in the Section 1.6 which is significant reduction of weekday leisure especially at the 30th and 50th quantiles by more than a half hour per day. This indicates that hard working students below the median tend to reduce more leisure. On the contrary, the students at the 90th quantile who used to enjoy much leisure increased leisure by almost an hour after the reform. This result may be due to the heterogeneous preference and productivity among students. Column (2) contains estimates of regressions of the daily ours of sleep by quantile. It also confirms increases in sleep hours for students at 10th, 40th, 60th, and 90th quantiles. Column (3) shows no effect on the hours of private lesson which is consistent with the results in Table 1.6 and 1.7. Column (4) presents the significant effects of increasing self study hours for the 10th to 60th quantile students. This is also consistent with the standard DD result in Table 1.6.

2.5 Concluding Remarks

This chapter estimated the nationwide college admission grading policy effect on students' time allocation using the non-parametric estimation methods such as CIC and QDID models. Assumptions for consistent CIC estimates are introduced, but more rigorous justification of the assumptions should be implemented for the valid application. The results confirm that

[3]The direction of coefficients in the results in a discrete setting shown in Table B.1, B.2, and B.3 and in a QDID setting in Table B.4 are unchanged. See Appendix B.

it is hard to find an evidence of increase in leisure for most of quantiles except the 90th. Also, the reason of non-significant effects on self study at the upper quantile students (from 70th to 90th) might be because of the time constraints of available time and effort. That is, they had already been spending sufficiently huge time on self study and were unable to significantly increase time for self study further.

FIGURES AND TABLES

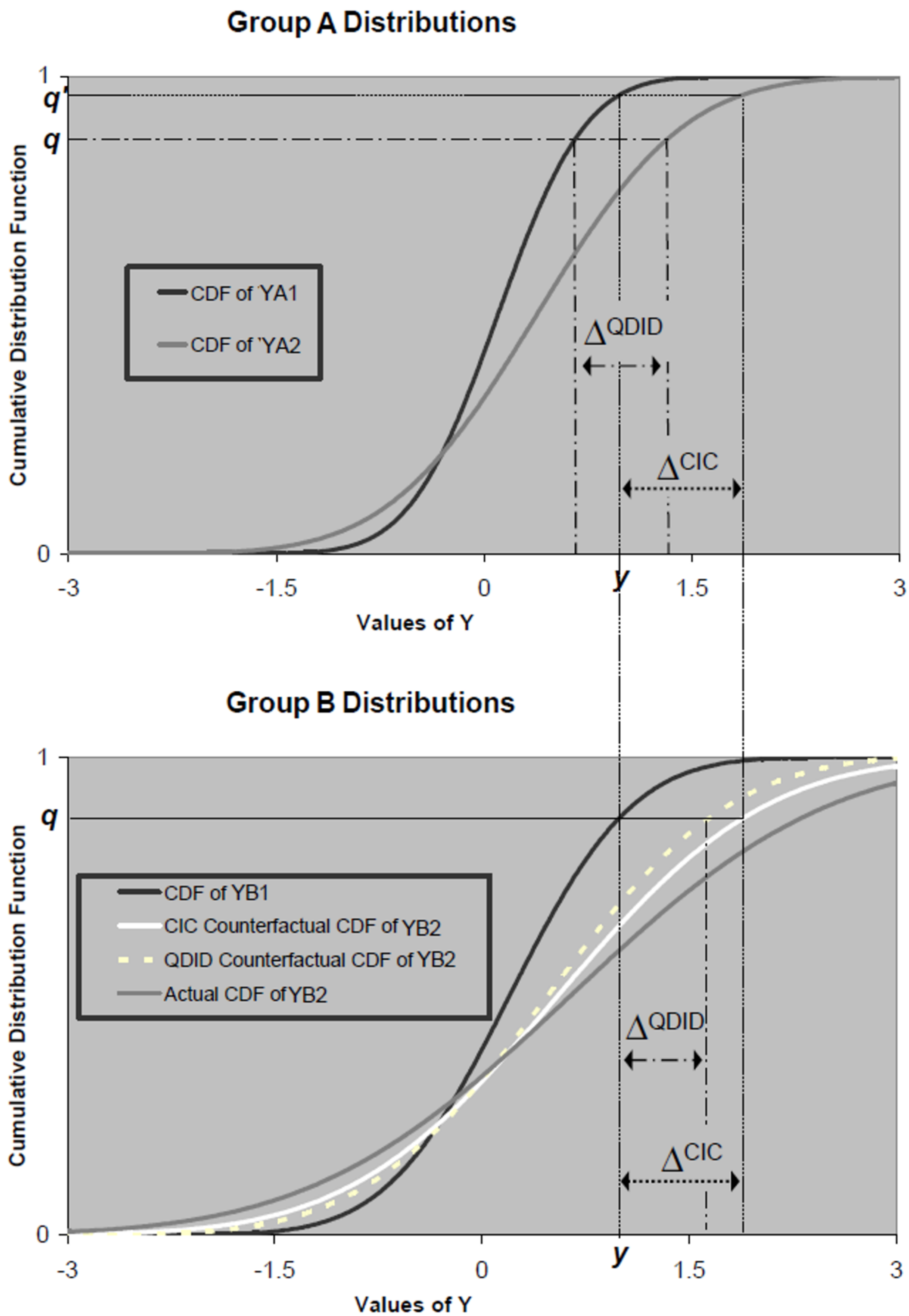


Figure 2.1: Illustration of Transformations (Athey and Imbens, 2006)

Group	Mean (s.d.)	Percentile				
		10th	30th	50th	70th	90th
Non-Takers						
2005 (N = 951)	4.29 (1.61)	2.00	3.00	4.00	6.00	6.00
2008 (N = 569)	4.84 (2.51)	2.00	4.00	5.00	6.00	8.00
SAT Takers						
2005 (N = 2199)	2.67 (1.49)	1.00	2.00	3.00	3.00	5.00
2008 (N = 1266)	2.71 (2.12)	1.00	1.00	2.00	3.00	6.00

Table 2.1: Weekday Leisure by percentile (hours/day)

	(1) Weekday Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study
Mean	-0.364**	0.309**	0.942	0.393
	(0.183)	(0.121)	(1.021)	(0.386)
10th	-0.068	0.945**	-0.263	0.307***
	(0.157)	(0.399)	(0.231)	(0.064)
20th	-0.137	0.047	-0.230*	0.368***
	(0.300)	(0.109)	(0.132)	(0.046)
30th	-0.757***	0.112	-0.007	0.452***
	(0.144)	(0.169)	(0.171)	(0.062)
40th	-0.141	0.509	-0.463	0.635***
	(0.278)	(0.396)	(1.600)	(0.057)
50th	-0.657*	0.132	-0.475	0.831***
	(0.338)	(0.312)	(1.696)	(0.160)
60th	-0.681**	0.066	0.667	1.443***
	(0.268)	(0.044)	(1.720)	(0.236)
70th	-0.103	0.129*	1.074	0.505
	(0.348)	(0.071)	(1.763)	(0.651)
80th	0.060	0.659*	0.478	-0.045
	(0.193)	(0.356)	(2.545)	(1.452)
90th	0.371	0.915***	1.445	-0.705
	(0.304)	(0.214)	(2.344)	(1.619)
N	3573	3564	3573	3573

[a] The variance is estimated by bootstraping the results 100 times.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Units are hours per day.

Table 2.2: Changes-in-changes Estimates (Continuous)

Chapter 3

THE BEHAVIORAL EFFECTS OF A SOCIAL MEDIA FRUIT DRINK COUNTERMARKETING MESSAGES: EVIDENCE FROM A RANDOMIZED EXPERIMENT AMONG US LATINX PARENTS

Can Facebook countermarketing messages reduce Latinx parents' choice of sugary fruit drinks for their children age 0-5? In this chapter[1] We estimate the treatment impacts of fruit drink countermarketing messages delivered via Facebook groups on parents' drink choice. A six-week randomized controlled trial among 1,628 US Latinx parents with children age 0-5 in 2019 found that social media countermarketing messages significantly reduced percent of parents choosing fruit drink from online store in the two intervention groups, one with fruit drink countermarketing messages and the other one with countermarketing and water promotion messages combined, compared to the control group with car seat messages (-13.7 percentage points vs -29.7 points vs -5.4 points) after the interventions. This implies that social media may be a effective and cost-efficient method for disseminating unhealthy beverage countermarketing and water promotion messages which can reduce unhealthy consumption.

3.1 Motivation

Sugar sweetened beverages (SSBs) provide instant gratification. They are sweet, tasty, and widespread. Many kinds of beverages on the supermarket shelves are SSBs including sugar-added fruit drinks, sports beverages, and soda. At the same time they are major health

[1]This chapter is based on Krieger et al. (2021) and under embargo until the journal publication decision is made and the submitted manuscript is published.

threats to preschool children as early consumption of SSBs is positively associated with obesity, diabetes, and dental caries (Bleich (2018), Valenzuela et al. (2020)). Banerjee and Mullainathan (2010) describe sugary products as *temptation goods* which provide visceral and primitive satisfaction in the moment but are not utility for our future selves. People are tend to show unhealthy behaviors such as over-consuming sugar, fat, alcohol, and cigarettes (Kremer et al., 2019). Furthermore, the negative impacts from SSB intake are more common among certain ethnic or cultural groups. For example, Latino children consume more SSBs than non-Hispanic white children (Lee et al. (2021); Beck et al. (2017)).

Which intervention will effectively reduce SSB consumption of children, especially more vulnerable groups? We may think of two approaches. One approach on the demand side is to lower marginal utility for every quantity consumed. The other one is to increase marginal cost of consumption. The latter one can be easily implemented by imposing a tax on SSBs (but it may not be easy in a political context). SSB taxes have already been adopted in the several US cities such as Boulder, CO, Philadelphia, PA, San Francisco, CA, and Seattle, WA. Those "Sin taxes" on SSB are studied as an effective instrument for reducing sugary drinks consumption (Allcott et al., 2019a) and nutritional inequality (Allcott et al., 2019b). Countermarketing unhealthy drinks tailored for Latinx parents could be considered to be another effective public health intervention for children's healthy drinking habit combined with the sin tax which imposes economic disincentives to consumers (Palmedo et al., 2017). However, the intervention effects on consumers' behavior are hard to be solely identified and estimated without clear evidence in an experimental environment. Study of countermarketing in economics and health has long been focused on tobacco industry (Dave et al., 2019) and there is limited evidence of the effects of SSBs countermarketing interventions.

This study provides experimental evidence of marginal effects of culturally-tailored countermarketing messages via social media on Latinx parents' beverage choice behavior for their children. Section 3.2 describe study methods and Section 3.3 shows empirical results. Section 3.4 concludes.

3.2 Methods

3.2.1 Study Design and Procedure

We conducted a three-arm randomized controlled trial (RCT) during October and November 2019 enrolling 1,628 Latinx parents of children age 0-5 in the United States to deliver fruit drink countermarketing messages and water promotion messages via Facebook group for six weeks. Participants were recruited by a survey research firm using its nationwide proprietary database and targeted social media messages. They were self-identified as daily social media users and Latinx parents of age 18 years or older and of a child aged 0-5. At the beginning of the study, 1,628 eligible parents completed an online baseline survey and received a $15 incentive. Figure 3.1 illustrates the study procedures. A total number of 1,628 screened parents completed the baseline survey. After the initial survey, they were randomized and assigned to one of three study groups.

3.2.2 Per-Protocol Analysis

After six weeks of interventions, 1,628 randomized parents were informed to complete an online exit survey to receive $20 incentive. A total number of 1,284 participants ran through the interventions and completed the exit survey. Table 3.2 shows the baseline summary statistics of 1,284 parents who completed interventions and the exit survey. The remaining participants are no longer randomly assigned. Attrition rates (one minus the ratio of the completed participants to the randomized participants), however, were ranging from 18% to 23% for three groups and attrition patterns and the characteristics of groups were similar across the groups as seen in Table 3.1 and Table 3.2. Note that no statistically significant mean difference between groups is found in the observed characteristics of the per protocol groups[2]. In this chapter, I focus on these 1,284 per protocol participants to estimate average treatment effects (ATE) on those strict compliers.

3.2.3 Interventions

During the trial, three messages out of total nine messages are posted each week on the Facebook groups for six weeks. Each intervention message, therefore, is shown to all parents twice (two weeks) during the period. A message consist of a text header and an image. Participants can click *like* and comment on the posted messages. The parents in the first intervention group (fruit drink-only group) receive fruit drink countermarketing messages posted on the assigned Facebook group. These messages are designed to provide correct information about the ingredients and adverse health effects of sugar-added fruit drinks. The second intervention group (combination intervention group) receive the same countermarketing messages as well as water promotion messages. Control group parents are exposed to car seat safety messages which are assumed to be independent of beverage choice behavior. Figure 3.2 and Figure 3.3 are the example messages for each group.

3.2.4 Outcome of Interest

Primary outcome of interest in this study is the proportion of fruit drink choice for children at the online virtual store before and after the interventions in each group. In the baseline and exit surveys, participants are asked to imagine a trip to a supermarket to purchase a beverage for their oldest child age 0-5 years. The virtual online store shelf offers a typical variety of drinks usually found at a store including six fruit drinks, two waters, two 100% juices, one soda, and one milk. Figure 3.4 illustrates the beverages participants faced. The

[2] Comparison of control and two intervention groups is done by ANOVA for continuous variables and Chi-square test for proportions.

beverage choices were made twice, once at baseline and once at exit surveys. Parents were offered two $2 coupons that can be redeemed at the actual store to purchase the beverages they chose.

3.2.5 Statistical Methods

The per-protocol analysis includes only those who completed both interventions and exit surveys. We use linear regression models with robust standard errors to estimate intervention effects of the proportion of fruit drink choice for each group. We regress individual post fruit drink choice on the dummy of each intervention with a control of their baseline fruit drink choice. Also, we estimate the mean differences of the effect size in the intervention groups compared to the control group and the mean differences between intervention groups. P-values of between group comparisons for fruit drinks and for other beverages are separately corrected by using Bonferroni-Holm correction (Holm, 1979) to test multiple family-wise hypotheses.

3.3 Results

Table 3.1 shows the baseline characteristics of 1,628 randomized participants. The mean age of participants is 31 for all three groups. 90% of them are female and the average age of child is 3.3 years old. All three groups show similar distributions of education level and household income. This section answers the questions: Did interventions change parents' beverage choice for their children? Were the changes significant? Which group was affected more?

3.3.1 Primary Outcome: Proportion of Fruit Drink Choice

Table 3.3 shows the proportions of virtual store choice of each beverage in each group before and after the interventions. At the beginning of the study, all three groups have similar proportions of fruit drink choice ranging from 43.6% (fruit drink-only group) to 48.7% (car seat control group). 42.2% of the control group chose fruit drink after the car seat safety campaign (-6.5 percentage points) while the fruit drink-only group (-19.2 percentage points) and the combination group (28.1 percentage points) have larger drops in fruit drink choice after seeing the messages.

The regression results are shown in Table 3.4. Parents who received only the fruit drink countermarketing messages reduced fruit drink choice significantly more than the control group (Column 1). The effect is larger in the combination group who received both the countermarketing and water promotion messages (Column 2). An interesting result is that the fruit drink reduction effect in the combination group is even significantly greater than that in the fruit-drink only group (Column 3).

3.3.2 Secondary Outcomes: Proportion of Other Beverage Choice

Water choice in three groups were less than a quarter at the baseline (Table 3.3). After the interventions, the proportions of participants both in the two intervention groups increased substantially (Table 3.4). The fruit drink-only group chose water 21.8 percentage points more than the control group. A significantly greater portion of the combination group chose water compared to the control group (+38.0 percentage points) as well as to the fruit drink-only group (+16.3 percentage points) after receiving countermarketing and water promotion messages. Choices of other beverages were not affected significantly by the interventions except for a significant reduction of 100% fruit juice in the combination group relative to the control group (-10.0 percentage points). There was, however, no significant difference across intervention groups in 100% fruit juice choice.

3.4 Concluding Remarks

Findings in the previous section indicate that at least in the short-run, countermarketing messages could bring a large behavioral effect helping people to avoid unhealthy drink choice. This effect may be coming from an information update about adverse health effects of SSBs intake and negative emotional appeals which might have affected parents' preference on SSBs. In addition to the fruit drink reduction effect, the countermarketing messages contributed to an increase in water choice at the online store. Moreover, the combined water promotion campaign may not only boost water choice but also discourage fruit drink choice relative to the countermarketing-only intervention. One reason of the decrease in 100% fruit in the combination intervention group could be the fact that participants tend to confuse (sugar-added) fruit drink and 100% fruit juice (with no added-sugar). Thus, helping participants to know clearly which is the healthy alternative to the sugar added fruit drinks for their children would effectively strengthen the intervention impacts.

FIGURES AND TABLES

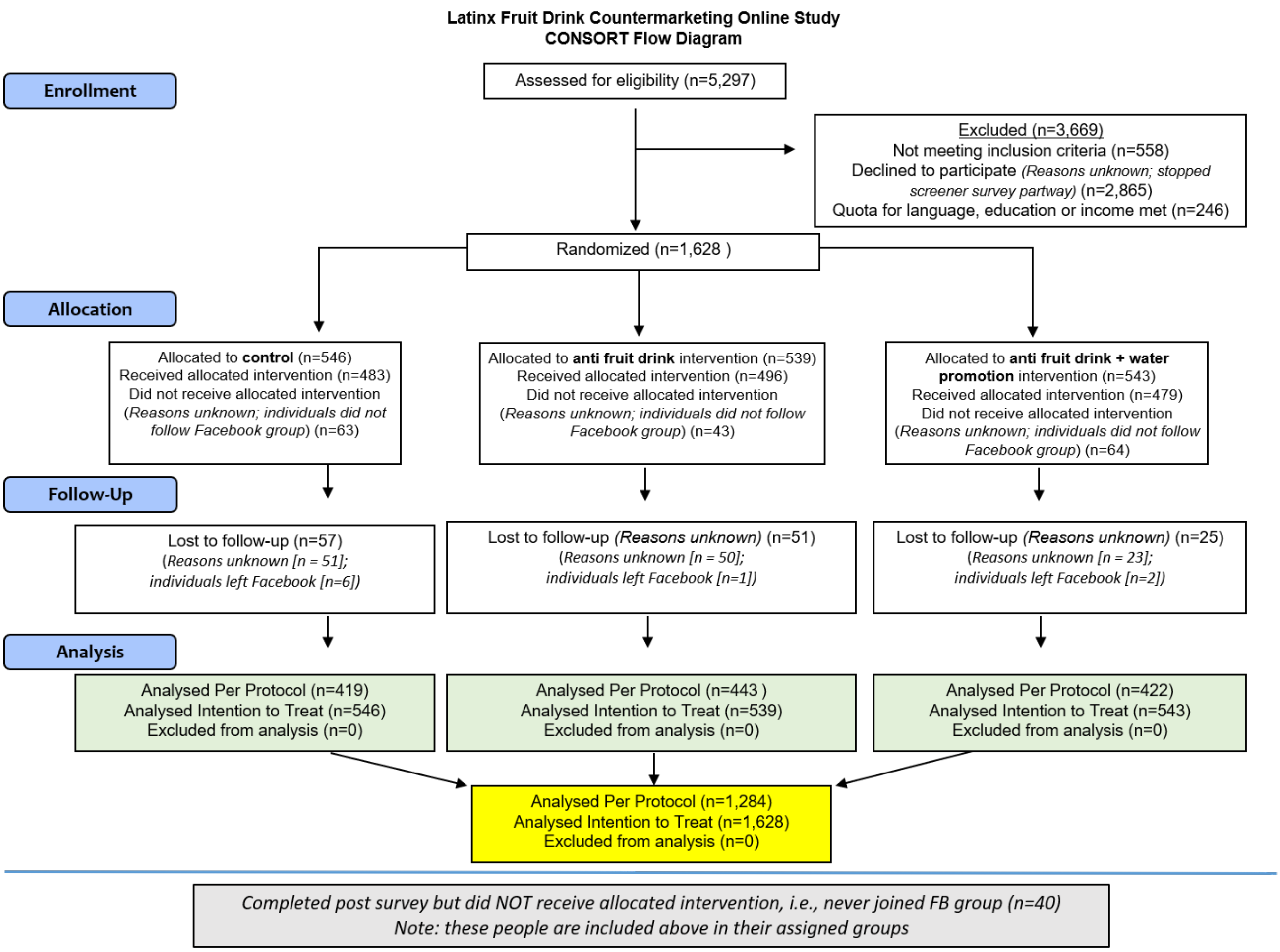


Figure 3.1: CONSORT Flow Chart

Sugary fruit drinks rot your kid's teeth. Water naturally protects them.

Parents say NO to fruit drinks and YES to water!

* Fruit drink-only group is exposed to only the countermarketing messages. Combination group received both countermarketing and water promotion messages.

Figure 3.2: Countermarketing and Water Promotion Messages

Figure 3.3: Car Seat Safety Message (Control Group)

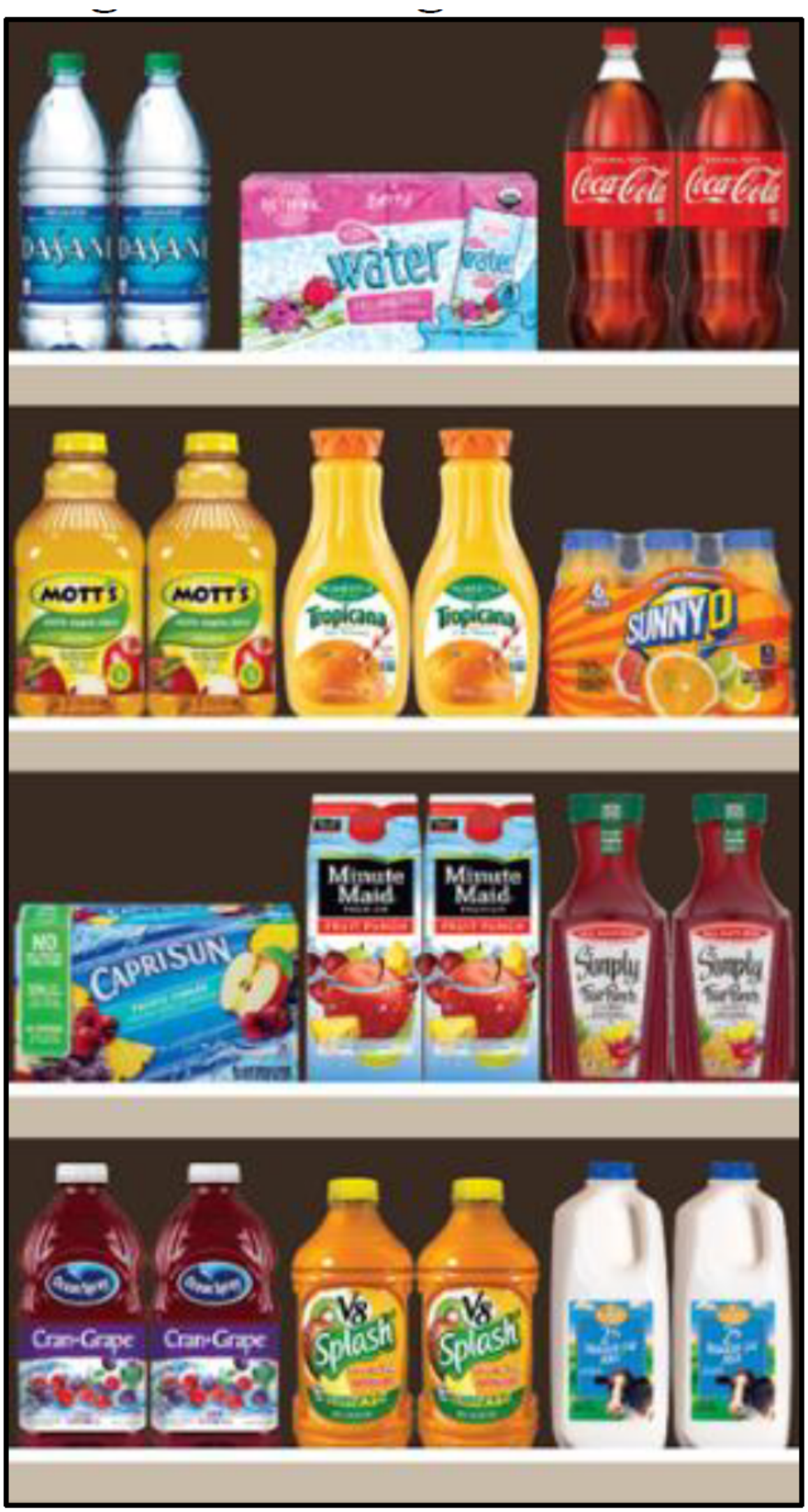


Figure 3.4: Image of Simulated Virtual Store

	Fruit Drink (N=539)	Fruit Drink + Water (N=543)	Control (N=546)
PARENT			
Age*, mean (SD)	30.76 (5.13)	31.03 (5.26)	31.03 (5.09)
Sex*, n (%)			
Female	482 (89.8)	498 (92.1)	503 (92.3)
Male	55 (10.2)	43 (7.9)	42 (7.7)
White race, n (%)	277 (51.4)	280 (51.6)	271 (49.6)
Mexican origin or descent, n (%)	374 (69.4)	385 (70.9)	370 (67.8)
Education*, n (%)			
High school or less	199 (37.5)	183 (34.3)	203 (37.7)
Some college	182 (34.3)	172 (32.2)	172 (31.9)
College or more	150 (28.2)	179 (33.5)	164 (30.4)
Preferred speaking Language, n (%)			
English more than Spanish	253 (46.9)	232 (42.7)	233 (42.7)
English and Spanish equally	189 (35.1)	201 (37.0)	205 (37.5)
Spanish more than English	97 (18.0)	110 (20.3)	108 (19.8)
English when using SNS, n (%)	432 (80.1)	411 (75.7)	432 (79.1)
Use Facebook >once a day, n (%)	497 (92.2)	508 (93.6)	507 (92.9)
Use Instagram >once a day, n (%)	368 (68.3)	340 (62.6)	358 (65.6)
CHILD			
Age, mean (SD)	3.34 (1.41)	3.34 (1.44)	3.30 (1.41)
Sex*, n (%)			
Female	282 (52.7)	261 (48.4)	272 (50.1)
Male	253 (47.3)	278 (51.6)	271 (49.9)
HOUSEHOLD			
Income, n (%)			
100% of FPL or less	232 (43.0)	215 (39.6)	235 (43.0)
100% to 199% of FPL	174 (32.3)	170 (31.3)	161 (29.5)
200% to 399% of FPL	96 (17.8)	110 (20.3)	107 (19.6)
400% of FPL or more	37 (6.9)	48 (8.8)	43 (7.9)
Number of adults, n (%)			
1	62 (11.5)	67 (12.3)	60 (11.0)
2	367 (68.1)	363 (66.9)	386 (70.7)
>=3	110 (20.4)	113 (20.8)	100 (18.3)
Number of kids, n (%)			
1	126 (23.4)	154 (28.4)	169 (31.0)
2	192 (35.6)	199 (36.6)	170 (31.1)
3	130 (24.1)	117 (21.5)	125 (22.9)
>=4	91 (16.9)	73 (13.4)	82 (15.0)

[a] * contains refused or unknown responses. Excluded from denominators.

Table 3.1: Baseline Characteristics (All Randomized Participants)

	Fruit Drink (N=443)	Fruit Drink + Water (N=422)	Control (N=419)
PARENT			
Age*, mean (SD)	30.75 (5.18)	30.89 (5.23)	31.26 (5.27)
Sex*, n (%)			
Female	399 (90.1)	393 (93.1)	388 (92.6)
Male	44 (9.9)	29 (6.9)	31 (7.4)
White race, n (%)	232 (52.4)	233 (55.2)	216 (51.6)
Mexican origin or descent, n (%)	317 (71.6)	311 (73.7)	290 (69.2)
Education*, n (%)			
High school or less	155 (35.1)	130 (31.1)	149 (35.9)
Some college	155 (35.1)	144 (34.7)	137 (33.0)
College or more	131 (29.7)	144 (34.7)	129 (31.1)
Preferred speaking Language, n (%)			
English more than Spanish	220 (49.8)	190 (45.0)	190 (45.3)
English and Spanish equally	156 (35.2)	157 (37.2)	158 (37.7)
Spanish more than English	67 (15.1)	75 (17.8)	71 (16.9)
English when using SNS, n (%)	371 (83.8)	332 (78.7)	344 (82.1)
Use Facebook >once a day, n (%)	411 (92.8)	397 (94.1)	393 (93.8)
Use Instagram >once a day, n (%)	307 (69.3)	265 (62.8)	276 (65.9)
CHILD			
Age, mean (SD)	3.37 (1.39)	3.32 (1.44)	3.27 (1.40)
Sex*, n (%)			
Female	235 (53.0)	202 (48.1)	213 (50.1)
Male	208 (47.0)	218 (51.9)	204 (49.9)
HOUSEHOLD			
Income, n (%)			
100% of FPL or less	176 (39.7)	161 (38.2)	167 (39.9)
100% to 199% of FPL	149 (33.6)	134 (31.8)	133 (31.8)
200% to 399% of FPL	86 (19.4)	86 (20.4)	88 (21.0)
400% of FPL or more	32 (7.2)	41 (9.7)	31 (7.4)
Number of adults, n (%)			
1	47 (10.6)	49 (11.6)	45 (10.7)
2	305 (68.9)	290 (68.7)	302 (72.1)
>=3	91 (20.5)	83 (19.7)	72 (17.2)
Number of kids, n (%)			
1	109 (24.6)	122 (28.9)	137 (32.7)
2	146 (33.0)	153 (36.3)	125 (29.8)
3	117 (26.4)	87 (20.6)	97 (23.2)
>=4	71 (16.0)	60 (14.2)	60 (14.3)

[a] * contains refused or unknown responses. Excluded from denominators.

Table 3.2: Baseline Characteristics (Per protocol)

Beverage Choice - Online Store	Baseline % (N)	Exit % (N)
FRUIT DRINK (N=443)		
Fruit Drink	43.6 (193)	24.4 (108)
Soda	1.1 (5)	1.6 (7)
Water	22.6 (100)	43.8 (194)
Milk	9.0 (40)	14.9 (66)
100% fruit juice	23.7 (105)	15.3 (68)
FRUIT DRINK + WATER (N=422)		
Fruit Drink	44.5 (188)	16.4 (69)
Soda	2.1 (9)	2.1 (9)
Water	22.3 (94)	60.0 (253)
Milk	7.3 (31)	9.5 (40)
100% fruit juice	23.7 (100)	12.1 (51)
CAR SEAT CONTROL (N=419)		
Fruit Drink	48.7 (204)	42.2 (177)
Soda	2.1 (9)	4.1 (17)
Water	18.6 (78)	20.5 (86)
Milk	8.8 (37)	11.7 (49)
100% fruit juice	21.7 (91)	21.5 (90)

Notes: Analyses using only complete data.

Table 3.3: Online Store Beverage Choice (Per Protocol)

Beverage Choice Online Store	Adjusted Risk Difference[a] (95% CI)		
	Fruit Drink vs. Control	Fruit Drink + Water vs. Control	Fruit Drink vs. Fruit Drink + Water
Fruit Drink[b]	-16.3*** (-22.1, -10.6)	-24.6*** (-30.2, -19.1)	-8.3** (-13.5, -3.2)
Soda	-2.0 (-4.1, 0.1)	-1.9 (-4.1, 0.3)	0.1 (-1.5, 1.7)
Water	21.8*** (16.1, 27.4)	38.0*** (32.3, 43.7)	16.3*** (10.0, 22.6)
Milk	3.1 (-1.2, 7.4)	-1.6 (-5.5, 2.3)	-4.8 (-8.9, -0.7)
100% fruit juice[c]	-6.7 (-11.5, -1.9)	-10.0*** (-14.7, -5.2)	-3.3 (-7.7, 1.1)

Notes: Analyses using only complete data. *< 0.05, **< 0.01, ***< 0.001.

[a] Linear regression model with robust standard error.

[b] P-values of adjusted risk differences are adjusted using Bonferroni-Holm correction of 3 tests for fruit drinks (primary prespecified outcome).

[c] P-values of adjusted risk differences are adjusted using Bonferroni-Holm correction of 12 tests for other beverages.

Table 3.4: Treatment Effects on Online Store Beverage Choice (Per Protocol)

BIBLIOGRAPHY

Allcott, Hunt, Benjamin B. Lockwood, and Dmitry Taubinsky, "Should We Tax Sugar-Sweetened Beverages? An Overview of Theory and Evidence," *Journal of Economic Perspectives*, August 2019, *33* (3), 202–27.

_ , Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell, "Food Deserts and the Causes of Nutritional Inequality," *The Quarterly journal of economics*, 2019, *134* (4), 1793–1844.

Andrews, Rodney J., Jing Li, and Michael F. Lovenheim, "Quantile Treatment Effects of College Quality on Earnings," *Journal of Human Resources*, 2016, *51* (1), 200–238.

Athey, Susan and Guido W. Imbens, "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 2006, *74* (2), 431–497.

Avery, Christopher, Soohyung Lee, and Alvin E Roth, "College Admissions as Non-Price Competition: The Case of South Korea," Working Paper 20774, National Bureau of Economic Research December 2014.

Banerjee, Abhijit and Sendhil Mullainathan, "The Shape of Temptation: Implications for the Economic Lives of the Poor," Working Paper 15973, National Bureau of Economic Research May 2010.

Beck, Amy, Alicia Fernandez, Jenssy Rojina, and Débora Cabana, "Randomized Controlled Trial of a Clinic-Based Intervention to Promote Healthy Beverage Consumption Among Latino Children," *Clinical pediatrics*, 05 2017, *56*, 9922817709796.

Becker, Gary S., "A Theory of the Allocation of Time," *The Economic Journal*, 1965, *75* (299), 493–517.

Bodoh-Creed, Aaron L. and Brent R. Hickman, "College assignment as a large contest," *Journal of Economic Theory*, 2018, *175* (C), 88–126.

Bond, Timothy N., George Bulman, Xiaoxiao Li, and Jonathan Smith, "Updating Human Capital Decisions: Evidence from SAT Score Shocks and College Applications," *Journal of Labor Economics*, 2018, *36* (3), 807–839.

Bound, John, Brad Hershbein, and Bridget Terry Long, "Playing the Admissions Game: Student Reactions to Increasing College Competition," *Journal of Economic Perspectives*, December 2009, *23* (4), 119–46.

Che, Yeon-Koo and Youngwoo Koh, "Decentralized College Admissions," *Journal of Political Economy*, 2016, *124* (5), 1295–1338.

Dave, Dhaval, Daniel Dench, Michael Grossman, Donald S. Kenkel, and Henry Saffer, "Does e-cigarette advertising encourage adult smokers to quit?," *Journal of Health Economics*, 2019, *68*, 102227.

Dubey, Pradeep and John Geanakoplos, "Grading exams: 100,99,98,... or A,B,C?," *Games and Economic Behavior*, 2010, *69* (1), 72 – 94. Special Issue In Honor of Robert Aumann.

Gibson, Matthew and Jeffrey Shrader, "Time Use and Labor Productivity: The Returns to Sleep," *The Review of Economics and Statistics*, 2018, *100* (5), 783–798.

Grau, Nicolás, "The impact of college admissions policies on the academic effort of high school students," *Economics of Education Review*, 2018, *65*, 58 – 92.

Gronau, Reuben, "Leisure, Home Production, and Work–the Theory of the Allocation of Time Revisited," *Journal of Political Economy*, 1977, *85* (6), 1099–1123.

Hafalir, Isa, Rustamdjan Hakimov, Dorothea Kübler, and Morimitsu Kurino, "College admissions with entrance exams: Centralized versus decentralized," *Journal of Economic Theory*, 05 2018, *176*.

Han, Chirok, Changhui Kang, and Sam-Ho Lee, "Measuring effort incentives in a tournament with many participants: Theory and application," *Economic Inquiry*, 2016, *54* (2), 1240–1250.

Harbaugh, Rick and Eric Rasmusen, "Coarse Grades: Informing the Public by Withholding Information," *American Economic Journal: Microeconomics*, February 2018, *10* (1), 210–35.

Holm, Sture, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 1979, *6* (2), 65–70.

Hopkins, Ed and Tatiana Kornienko, "Running to Keep in the Same Place: Consumer Choice as a Game of Status," *American Economic Review*, September 2004, *94* (4), 1085–1107.

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press, 2015.

Kahneman, Daniel and Alan B. Krueger, "Developments in the Measurement of Subjective Well-Being," *Journal of Economic Perspectives*, March 2006, *20* (1), 3–24.

Kranker, Keith, "CIC: Stata module to implement the Athey and Imbens (2006) Changes-in-Changes model," *Statistical Software Components S458656, Boston College Department of Economics*, 2019.

Kremer, Michael, Gautam Rao, and Frank Schilbach, "Chapter 5: Behavioral development economics in," *Vol. 2 of Handbook of Behavioral Economics: Applications and Foundations 1, North-Holland*, 2019, pp. 345–458.

Krieger, James, Taehoon Kwon, Rudy Ruiz, Lina Pinero Walkinshaw, Jiali Yan, and Christina Roberto, "Don't Be Fooled - A Social Media Fruit Drink Countermarketing Campaign for Latinx Parents of Children Age 0-5: A Randomized Controlled Trial," *Submitted for publication*, 2021.

Köbberling, Veronika and Peter P. Wakker, "An index of loss aversion," *Journal of Economic Theory*, 2005, *122* (1), 119 – 131.

Lee, Matthew M, Emily Altman, and Kristine A Madsen, "Secular Trends in Sugar-Sweetened Beverage Consumption Among Adults, Teens, and Children: The California Health Interview Survey, 2011–2018," *Preventing Chronic Disease*, 2021, *18.*

Moldovanu, Benny, Aner Sela, and Xianwen Shi, "Contests for Status," *Journal of Political Economy*, 2007, *115* (2), 338–363.

N., Vercammen Kelsey A. Bleich Sara, "The negative impact of sugar-sweetened beverages on children's health: an update of the literature," *BMC Obesity*, 2018, *5.*

OECD, *PISA 2015 Results (Volume III)* 2017.

Olszewski, Wojciech and Ron Siegel, "Large Contests," *Econometrica*, 2016, *84* (2), 835–854.

_ **and** _ , "Pareto Improvements in the Contest for College Admissions," *Manuscript, Department of Economics, Northwestern University*, 2019.

Palmedo, P. Christopher, Lori Dorfman, Sarah Garza, Eleni Murphy, and Nicholas Freudenberg, "Countermarketing Alcohol and Unhealthy Food: An Effective Strategy for Preventing Noncommunicable Diseases? Lessons from Tobacco," *Annual Review of Public Health*, 2017, *38* (1), 119–144. PMID: 28384081.

Rosenbaum, Paul R. and Donald B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 1983, *70* (1), 41–55.

_ **and** _ , "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, 1985, *39* (1), 33–38.

Rubin, Donald B., "Estimating causal effects of treatments in randomized and nonrandomized studies.," *Journal of Educational Psychology*, 1974, *66* (5), 688–701.

_ , "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation," *Health Services & Outcomes Research Methodology*, 2001, *2*, 169–188.

Sari, Nazmi and Meric Osman, "The effects of patient education programs on medication use among asthma and COPD patients: a propensity score matching with a difference-in-difference regression approach.," *BMC Health Services Research*, 2015, *15* (1), 1 – 9.

Schultz, Theodore W., "The Value of the Ability to Deal with Disequilibria," *Journal of Economic Literature*, 1975, *13* (3), 827–846.

Stuart, Elizabeth, Haiden A. Huskamp, Kenneth Duckworth, Jeffrey Simmons, Zirui Song, Michael E. Chernew, and Colleen L Barry, "Using propensity scores in difference-in-differences models to estimate the effects of a policy change," *Health Services and Outcomes Research Methodology*, 11 2014, *14* (4), 166–182.

Su, Dai, Ying chun Chen, Hong xia Gao, Hao miao Li, Jing jing Chang, Di Jiang, Xiao mei Hu, Shi han Lei, Min Tan, and Zhi fang Chen, "Effect of integrated urban and rural residents medical insurance on the utilisation of medical services by residents in China: a propensity score matching with difference-in-differences regression approach," *BMJ Open*, 2019, *9* (2).

Tarokh, Leila, Jared M. Saletin, and Mary A. Carskadon, "Sleep in adolescence: Physiology, cognition and mental health," *Neuroscience Biobehavioral Reviews*, 2016, *70*, 182 – 188. The Adolescent Brain.

Thomas, Duncan, John Strauss, and Maria-Helena Henriques, "How Does Mother's Education Affect Child Height?," *The Journal of Human Resources*, 1991, *26* (2), 183–211.

United Nations, *World Youth Report 2003 - Rethinking leisure time: Expanding opportunities for young people and communities* 2006.

Valenzuela, Maria Josefina, Beverley Waterhouse, Vishal R Aggarwal, Karen Bloor, and Tim Doran, "Effect of sugar-sweetened beverages on oral health: a systematic review and meta-analysis," *European Journal of Public Health*, 08 2020, *31* (1), 122–129.

Weaver, Matthew D., Laura K. Barger, Susan Kohl Malone, Lori S. Anderson, and Elizabeth B. Klerman, "Dose-Dependent Associations Between Sleep Duration and Unsafe Behaviors Among US High School Students," *JAMA Pediatrics*, 12 2018, *172* (12), 1187–1189.

Willis, Robert J. and Sherwin Rosen, "Education and Self-Selection," *Journal of Political Economy*, 1979, *87* (5), S7–S36.

Zimmerman, Seth D., "The Returns to College Admission for Academically Marginal Students," *Journal of Labor Economics*, 2014, *32* (4), 711–754.

Appendix A
CHAPTER 1 APPENDIX

A.1 Derivation of Equilibria

Equation (1.1) and (1.5) consist the objective function

$$\max_{l_0,l_q,k_q,x} u_0(l_0,x)+\beta u_1(w_g\cdot g(l_g,k_g)+w_s\cdot s(l_s,k_s)+w_r\cdot r(l_r,k_r)+\epsilon) \tag{A.1}$$

by expressing $Q(\cdot)$ as a function of a linear combination of SAT, GPA, and resume. The equilibrium conditions of interior solution will be:

$$[l_0] \quad \frac{1}{\beta}\frac{\partial u_0}{\partial l_0}+w_g\frac{\partial u_1}{\partial g}\frac{\partial g}{\partial l_g}\frac{\partial l_g}{\partial l_0}+w_s\frac{\partial u_1}{\partial s}\frac{\partial s}{\partial l_s}\frac{\partial l_s}{\partial l_0}+w_r\frac{\partial u_1}{\partial r}\frac{\partial r}{\partial l_r}\frac{\partial l_r}{\partial l_0}=0 \tag{A.2}$$

$$[x] \quad \frac{1}{\beta}\frac{\partial u_0}{\partial x}+w_g\frac{\partial u_1}{\partial g}\frac{\partial g}{\partial k_g}\frac{\partial k_g}{\partial x}+w_s\frac{\partial u_1}{\partial s}\frac{\partial s}{\partial k_s}\frac{\partial k_s}{\partial x}+w_r\frac{\partial u_1}{\partial r}\frac{\partial r}{\partial k_r}\frac{\partial k_r}{\partial x}=0 \tag{A.3}$$

$$[l_q] \quad \frac{1}{\beta}\frac{\partial u_0}{\partial l_0}\frac{\partial l_0}{\partial l_q}+w_q\frac{\partial u_1}{\partial q}\frac{\partial q}{\partial l_q}+w_{q'}\frac{\partial u_1}{\partial q'}\frac{\partial q'}{\partial l_{q'}}\frac{\partial l_{q'}}{\partial l_q}+w_{q''}\frac{\partial u_1}{\partial q''}\frac{\partial q''}{\partial l_{q''}}\frac{\partial l_{q''}}{\partial l_q}=0 \tag{A.4}$$

$$[k_q] \quad \frac{1}{\beta}\frac{\partial u_0}{\partial x}\frac{\partial x}{\partial k_q}+w_q\frac{\partial u_1}{\partial q}\frac{\partial q}{\partial k_q}+w_{q'}\frac{\partial u_1}{\partial q'}\frac{\partial q'}{\partial k_{q'}}\frac{\partial k_{q'}}{\partial k_q}+w_{q''}\frac{\partial u_1}{\partial q''}\frac{\partial q''}{\partial k_{q''}}\frac{\partial k_{q''}}{\partial k_q}=0 \tag{A.5}$$

where $q,q',q''\in\{g,s,r\}$ and $q\neq q'\neq q''$. The time constraint (1.3) implies that a marginal increase (or decrease) in current leisure l_0 sacrifices (or is allocated to) l_g, l_s, and l_r. This yields,

$$\frac{\partial l_g}{\partial l_0}+\frac{\partial l_s}{\partial l_0}+\frac{\partial l_r}{\partial l_0}=-1. \tag{A.6}$$

A.2 Coarse Grading and Uncertain Status of the Loss Averters

I show that the loss averse decision makers also want to avoid the situation of being in lower grades in the SAT subjects than their *reference status* under the coarse grading scheme. First, there exists *ex-ante* uncertainty of a student's *rank* at the time of student's resource allocation decision before the score/grade distribution of SAT is finalized after the actual tests. In my model as described in Section 1.3, a student exerts her best effort by choosing the optimal level of time use l_s^* and the optimal amount of money k_s^* to produce SAT score (or grade) $s(l_s^*, k_s^*)$. Given a student's input level (l_s^*, k_s^*), standardized score method ensures almost continuously increasing rank function in SAT score. That is, student i's SAT score s_i would be distinguishable to student j's score s_j if there is a marginal difference between s_i and s_j. As long as $|s_i - s_j| > 0$, colleges can obtain an informative signal by comparing s_i and s_j. By contrast, coarse grading system may not generate an informative signal if $|s_i - s_j|$ is not sufficiently large enough to separate those two scores into different grades. A critical environment for students who make decisions on their resource allocation is the fact that no students, parents, and teachers can know the population distribution of SAT score or grade prior to the actual test. Suppose the math in a SAT is really easy so that more than 4% of the SAT takers got the full credits, then the highest grade can be assigned to all those students even though the top 4% is the cutoff guideline for the first grade. In the opposite situation, if the math is too hard so only less than 1% of applicants got the full credits, the cutoff of the highest grade will be at around the top 4% and the cutoff score should be way lower. Suppose a student decides her optimal investment level for SAT l_s^* and k_s^*. Ignoring any uncertainty in her SAT score without loss of generality, she will obtain the raw score of $s(l_s^*, k_s^*)$ in a SAT subject. Her *status* or percentile, however, is *ex-ante* uncertain as the difficulty of the exam and the performance of her cohort may not be the same as previous years and previous mock exams. Thus, she does not know her exact status corresponding to her SAT score $s(l_s^*, k_s^*)$ at the time she makes her time allocation decisions. Under the uncertain status, Equation (1.12) can be revisited with a random factor on the SAT part to generate uncertainty of status affecting the admission quality.

$$Q^i = w_g \cdot g^i(l_g^i, k_g^i) + w_s \cdot R(s^i(l_s^i, k_s^i), s^{j \neq i}(l_s^j, k_s^j), d) + w_r \cdot r^i(l_r^i, k_r^i) + \epsilon^i. \tag{A.7}$$

The *status* (or ranking) function $R(\cdot)$ which converts individual SAT score s^i to a rank or a grade over the population distribution is determined *after* the test by the degree of coarse grading d and the SAT scores of all SAT takers. The functional form of d is determined by the government. An increase in d means an increase of the number of students assigned into a grade. For example, the maximum d would be the *pass or fail* system where there exist only two partitions above and below a certain cutoff point. The minimum d would be the system with the finest grading where the raw scores are available for every player in the admission game. Note that this source of ex-ante uncertainty $R(\cdot)$ is different from the overall unobservable factors ϵ^i.

Second, I assume that Korean high school students have *loss aversion* preference on the future utility $u_1(\cdot)$. The key features of loss aversion is asymmetric reaction to gain and loss. That is, loss aversion describes the concavity of utility for gains and the convexity of utility for losses. The intuition is straightforward. A college applicant will be happy if she luckily gets a slightly better admission than she expected. By the way, if she gets a slightly lower admission than her reference college, her parents, teachers, friends, and herself will be very disappointed since everyone thinks that she could have done better if she performed as usual. As in Köbberling and Wakker (2005), I modify the future utility function $u_1(\cdot)$ to a value function

$$v_1(Q(\cdot)) = \begin{cases} \frac{1-e^{-\mu Q}}{\mu} & \text{for all} \quad Q \geq Q^* \\ \lambda(\frac{e^{vQ}-1}{v}) & \text{for all} \quad Q < Q^* \end{cases} \tag{A.8}$$

where Q^* is the reference level of admission quality determined by referring to the previous mock test results. λ is the degree of loss aversion and takes a value of $\lambda \geq 1$. For $0 < v < \mu$, v_1 is convex for losses, concave for gains, and importantly, given the same deviation of admission quality ΔQ from a student's reference admission, the student losses more utility under the worse admission than the amount of utility gain from a higher admission relative to her reference admission. This implies $-v_1(Q^* - \Delta Q) > v_1(Q^* + \Delta Q)$. Adapting the loss aversion value function $v_1(\cdot)$ instead of $u_1(\cdot)$ yields

$$\max_{l_0, l_q, k_q, x} u_0(l_0, x) + \beta v_1(Q) \tag{A.9}$$

where

$$Q = w_g \cdot g^i(l_g^i, k_g^i) + w_s \cdot R(s^i(l_s^i, k_s^i), s^{j \neq i}(l_s^j, k_s^j), d) + w_r \cdot r^i(l_r^i, k_r^i) + \epsilon^i. \tag{A.10}$$

Equation (A.9) summarizes a student's lifetime welfare, consisting of the normal form of current utility u_0 and the loss averse utility v_1 discounted by β.

Finally, I investigate the effect of coarse grading on loss averse student's resource allocation behavior. What will happen if the degree of coarse grading d increases by adopting the nine grade-class system? I compare d_0 and d_1 where d_0 denotes the original finest grading scheme with the standardized SAT scores and percentiles available. d_1 denotes the new coarse grading scheme with the nine grades proportionally assigned over the standardized normal distribution with no available percentile and raw scores. Accordingly R_0 denotes the student's SAT rank under the finest grading scheme d_0 raging from the lowest SAT score which is zero to the highest standardized score s_{max}. R_1 denotes the student's SAT rank under the new coarse grading system.

$$\begin{cases} R_0 \in \{0, 1, 2, ..., 50, 51, ..., s_{max}\} \\ R_1 \in \{1, 2, ..., 9\} \end{cases} \tag{A.11}$$

Table A.1 shows the population statistics of SAT results from 2005 to 2009 school year. My sample cohorts are included in the 2005 and 2008 school year tests. Standardized scores are calculated under the assumption that the population scores are normally distributed[1]. When a SAT subject is easier than the other years, the population mean is high and the standard deviation tends to low as students may get some baseline points from the easy questions. These two effects yield the lower Z-score. As a result, the cutoff standardized score for Grade 1 becomes lower. This implies that a harder exam yields the higher maximum standardized score s_{max}. For example, the cutoff standardized score of Grade 1 of SAT Korean in 2006 school year is 125 which is the lowest among the other years because of its low degree of difficulty (and higher population mean). This means that if a smart student missed just only one single three-point question in the 2006 SAT Korean by her trembling hand and got the raw score of 97 out of 100, then she will get Grade 2 which is *not* the highest grade. Under d_0, the same amount of gain or loss ds of raw score from a student's reference score s^* will bring a minimum change of the student's status dR_0. Under d_1, with the same amount of deviation (gain and loss) ds of raw score from one's reference score s^*, the utility change of the *loss averse* student will be greater in the case of loss rather than the case of gain. This

[1]Standardized score is a linear transformation of Z-score which is calculated by the formula

$$\text{Standardized Score} = (\text{Reference Mean}) + (\text{Z-score}) \times (\text{Reference Standard Deviation})$$

where Z-score = (Student's Raw Score - Population Mean) / Standard Deviation. Reference mean (Reference standard deviation) for Korean and math are 100 (20) and for social and natural sciences are 50 (10).

implies that the new grading scheme d_1 increases variance of status $Var(R)$ given the same resource allocation s:

$$Var(R \mid d_0, s) < Var(R \mid d_1, s). \tag{A.12}$$

I assume that the individual variance of SAT score is symmetric around the reference score. That is, the probability of getting n points higher score than the reference score s^* is the same as the probability of getting n points lower score than s^*. Repeated mock tests help students to update their reference scores and to make the optimal exam day routine. As a result, the only factor that can affect her SAT score other than her resource investment is *luck*. From Equation (A.12) a student's expected loss-averse utility under the new coarse grading scheme d_1 is now smaller than that under d_0 as d_1 gives more drastic changes in status. To avoid the critical loss of quality of admission under the new coarse grading policy, the incentive to work hard and to invest more money on cram schools would increase. If this is true, the policy which aimed to reduce competition in SAT would have failed.

	2009 School Year															
	Korean				Math A				Math B				English			
Grade	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%
1	92	131	23,615	4.23	81	135	4,965	4.08	79	138	16,795	4.22	95	131	23,590	4.27
2	86	125	41,589	7.44	72	125	9,542	7.83	68	128	27,346	6.87	89	126	39,343	7.12
3	79	117	68,183	12.2	63	116	14,010	11.50	55	116	51,255	12.89	79	117	80,173	14.51
4	71	108	91,839	16.43	53	106	21,027	17.26	42	104	66,402	16.69	69	108	83,865	15.18
5	61	96	113,095	20.24	42	95	24,847	20.40	28	90	81,935	20.60	55	95	108,563	19.65
6	51	85	93,689	16.76	32	84	19,947	16.37	19	82	69,325	17.43	41	82	94,631	17.13
7	41	74	65,794	11.77	22	74	14,117	11.59	14	77	42,566	10.70	30	72	64,785	11.73
8	30	62	40,553	7.26	14	65	9,213	7.56	9	73	30,279	7.61	22	65	38,046	6.89
9	Under	Under	20,495	3.67	Under	Under	4,160	3.41	Under	Under	11,869	2.98	Under	Under	19,507	3.53

	2008 School Year															
	Korean				Math A				Math B				English			
Grade	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%
1	90	131	23,693	4.31	99	128	5,103	4.16	93	135	15,952	4.16	96	130	23,502	4.32
2	83	124	43,478	7.91	93	123	12,346	10.08	85	129	26,484	6.90	90	125	41,729	7.67
3	76	116	60,919	11.08	86	118	11,699	9.55	71	118	48,724	12.70	82	118	62,432	11.48
4	67	107	92,557	16.84	74	108	22,447	18.32	54	105	62,811	16.37	69	107	96,995	17.83
5	56	95	116,450	21.18	59	95	23,696	19.34	36	91	79,851	20.81	54	94	103,988	19.12
6	46	85	91,002	16.55	45	84	19,347	15.79	24	81	66,732	17.39	39	81	93,620	17.21
7	37	75	64,820	11.79	31	72	14,642	11.95	17	76	41,623	10.85	30	74	65,381	12.02
8	27	65	35,812	6.51	20	63	8,489	6.93	12	72	26,356	6.87	22	67	36,843	6.77
9	Under	Under	21,000	3.82	Under	Under	4,764	3.89	Under	Under	15,167	3.95	Under	Under	19,455	3.58

	2007 School Year															
	Korean				Math A				Math B				English			
Grade	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%
1	95	127	29,708	5.40	89	134	4,906	4.18	96	137	16,142	4.21	96	130	25,899	4.74
2	91	123	41,408	7.53	81	126	8,000	6.82	84	128	33,812	8.82	91	126	34,834	6.38
3	86	118	57,503	10.46	71	116	15,216	12.97	72	119	39,342	10.27	82	118	69,201	12.67
4	78	109	91,778	16.69	61	106	20,171	17.20	53	104	67,339	17.57	69	107	103,520	18.95
5	67	97	110,757	20.14	50	95	23,627	20.15	34	90	79,649	20.79	56	95	95,258	17.44
6	55	85	94,825	17.24	39	84	19,684	16.78	22	81	71,064	18.55	41	82	97,263	17.80
7	43	72	66,465	12.09	27	73	13,021	11.10	16	76	38,029	9.92	30	72	62,342	11.41
8	32	60	36,756	6.68	18	64	8,224	7.01	11	72	25,103	6.55	22	65	38,311	7.01
9	Under	Under	20,760	3.77	Under	Under	4,424	3.77	Under	Under	12,716	3.32	Under	Under	19,693	3.60

	2006 School Year															
	Korean				Math A				Math B				English			
Grade	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%
1	98	125	30,362	5.50	84	134	5,595	4.25	84	139	14,995	4.09	92	133	25,615	4.66
2	95	121	40,617	7.36	76	125	10,170	7.72	73	128	27,684	7.54	83	127	38,608	7.03
3	90	117	61,835	11.21	67	116	15,963	12.11	60	118	41,758	11.38	74	118	62,724	11.41
4	83	110	87,830	15.92	56	106	22,267	16.90	42	103	64,019	17.44	63	107	92,923	16.91
5	74	99	119,028	21.58	41	94	27,167	20.62	30	89	76,028	20.72	50	93	116,413	21.18
6		86	89,420	16.21		83	21,736	16.50		82	63,095	17.19		82	87,700	15.96
7		72	64,862	11.76		73	15,531	11.79		78	40,793	11.11		74	69,267	12.60
8		57	35,550	6.45		65	8,819	6.69		74	27,795	7.57		68	34,718	6.32
9		Under	22,050	4.00		Under	4,521	3.43		Under	10,849	2.96		Under	21,576	3.93

	2005 School Year															
	Korean				Math A				Math B				English			
Grade	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%	Raw Score	Std. Score	N	%
1	94	128	26,870	4.73	88	131	7,206	4.94	88	140	16,231	4.53	96	132	23,863	4.18
2	88	123	45,647	8.04	81	125	11,776	8.08	77	131	23,528	6.56	89	125	48,316	8.47
3	84	117	59,328	10.45	73	118	15,702	10.77	60	117	43,391	12.11	76	117	66,462	11.65
4	76	108	102,838	18.11	62	108	24,043	16.49	40	101	61,270	17.09	60	108	92,500	16.22
5	66	97	116,112	20.44	45	94	30,487	20.91	27	90	71,218	19.87	40	95	116,385	20.40
6		86	87,880	15.47		82	23,825	16.34		83	63,941	17.84		82	96,374	16.89
7		73	68,507	12.06		72	18,414	12.63		79	42,343	11.81		72	68,452	12.00
8		59	38,601	6.80		67	8,560	5.87		76	23,387	6.52		66	39,509	6.93
9		Under	22,167	3.90		Under	5,810	3.98		Under	13,126	3.66		Under	18,570	3.26

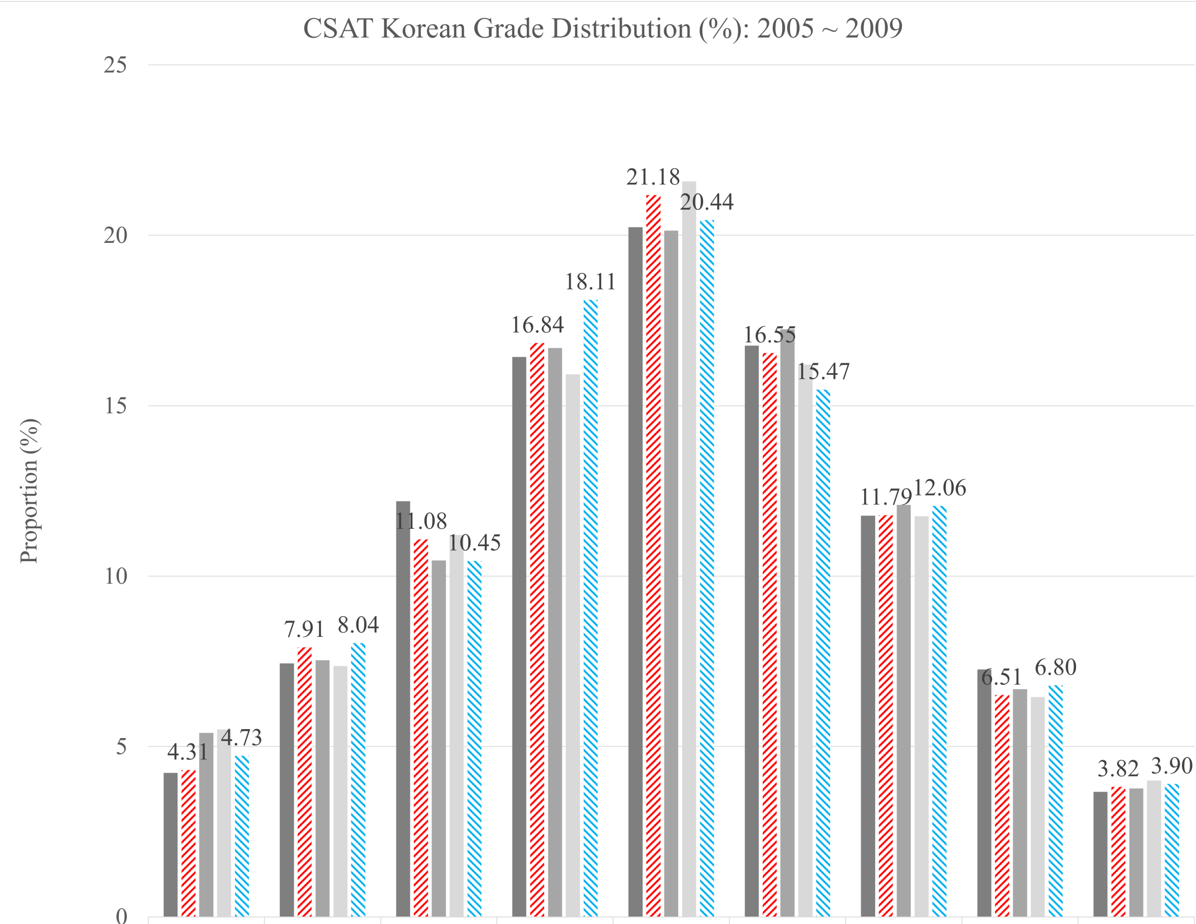
[a] Source: KICE, Megastudy, Kim01 Edu Consulting.
[b] Raw scores are not announced by KICE. They are estimated by private institutes, *Megastudy* and *Kim01 Edu Consulting.* Students get higher raw scores under an easy exam which results in higher Grade 1 cutoff.

Table A.1: Population Statistics of SAT Results (2005 – 2009)

	(1) Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study	(5) Leader Role	(6) Academic Club
Post	0.549***	0.115	0.085	-0.965***	0.031	0.000
	(0.134)	(0.093)	(0.248)	(0.073)	(0.024)	(0.014)
SAT Taker	-1.749***	-1.233***	5.661***	0.571***	0.037**	0.060***
	(0.078)	(0.070)	(0.324)	(0.049)	(0.016)	(0.012)
Post × SAT	-0.548***	0.190	0.907*	1.368***	0.049	-0.022
	(0.170)	(0.118)	(0.506)	(0.173)	(0.030)	(0.021)
N	4007	3998	4007	4007	4007	4007
Mean	3.280	6.166	4.588	1.719	0.189	0.085

[a] Cluster standard errors by school in parentheses.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Units are hours per day for leisure, sleep, and self study, hours per week for private lesson, and binary for leader role and academic club activities.
[d] Mean outcome is the average of total observations including both cohorts and groups.
[e] This table shows the average effects with the control of individual and household characteristics and interaction terms. Significance levels of all outcome variables are the same or similar to the results in Table 1.6.

Table A.2: Average Policy Effects on Time Allocation (No Covariates)



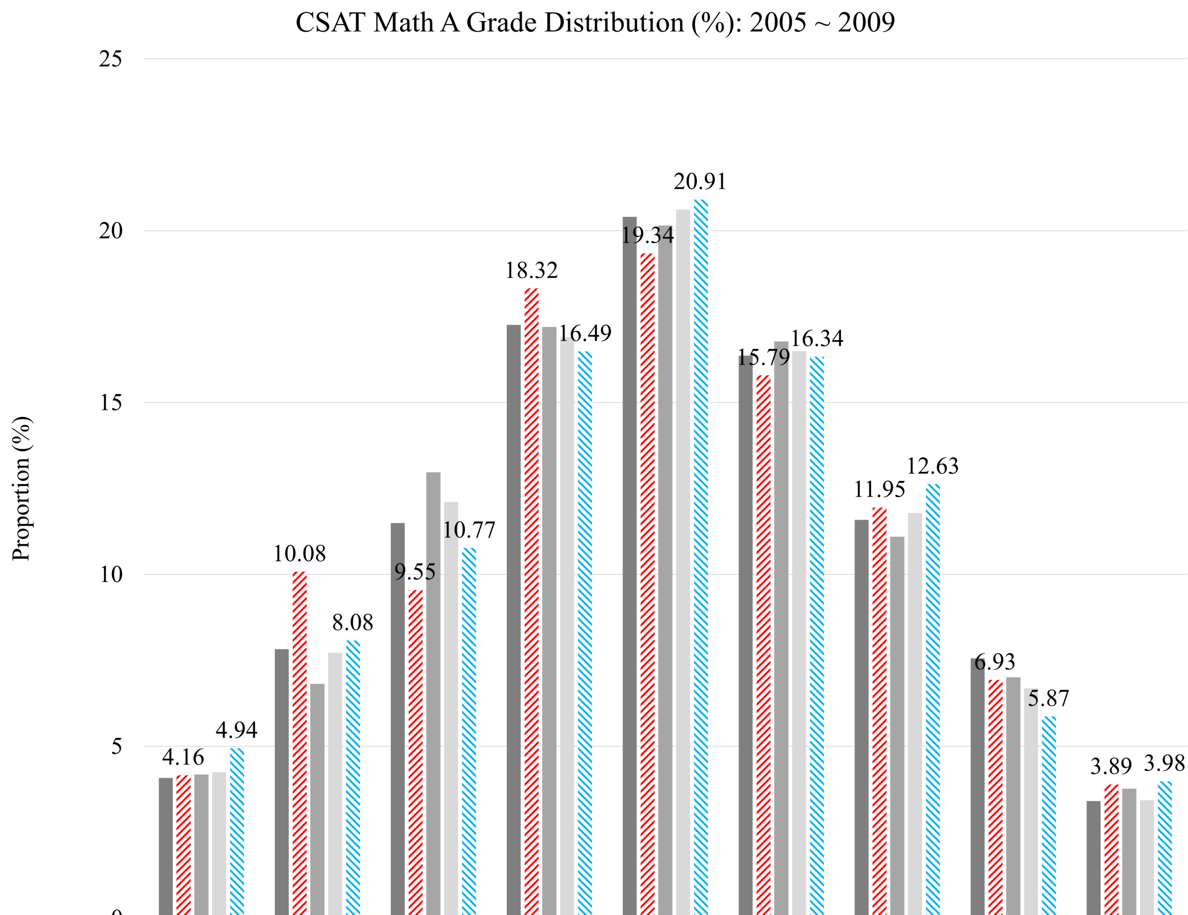
	1	2	3	4	5	6	7	8	9
09 Korean	4.23	7.44	12.2	16.43	20.24	16.76	11.77	7.26	3.67
08 Korean	4.31	7.91	11.08	16.84	21.18	16.55	11.79	6.51	3.82
07 Korean	5.4	7.53	10.46	16.69	20.14	17.24	12.09	6.68	3.77
06 Korean	5.5	7.36	11.21	15.92	21.58	16.21	11.76	6.45	4
05 Korean	4.73	8.04	10.45	18.11	20.44	15.47	12.06	6.80	3.90

CSAT Korean Grade (1: Highest - 9: Lowest)

09 Korean 08 Korean 07 Korean 06 Korean 05 Korean

a. Source: Ministry of Education.
b. Grade 1 is the highest. Students raw SAT scores are normalized. Each student's percentile is matched to one of nine preset grades in Table 1.1. When there are ties around the grade cutoff, the higher grade will be given to all.

Figure A.1: SAT Korean Grade Distribution



	1	2	3	4	5	6	7	8	9
09 Math A	4.08	7.83	11.5	17.26	20.4	16.37	11.59	7.56	3.41
08 Math A	4.16	10.08	9.55	18.32	19.34	15.79	11.95	6.93	3.89
07 Math A	4.18	6.82	12.97	17.2	20.15	16.78	11.1	7.01	3.77
06 Math A	4.25	7.72	12.11	16.9	20.62	16.5	11.79	6.69	3.43
05 Math A	4.94	8.08	10.77	16.49	20.91	16.34	12.63	5.87	3.98

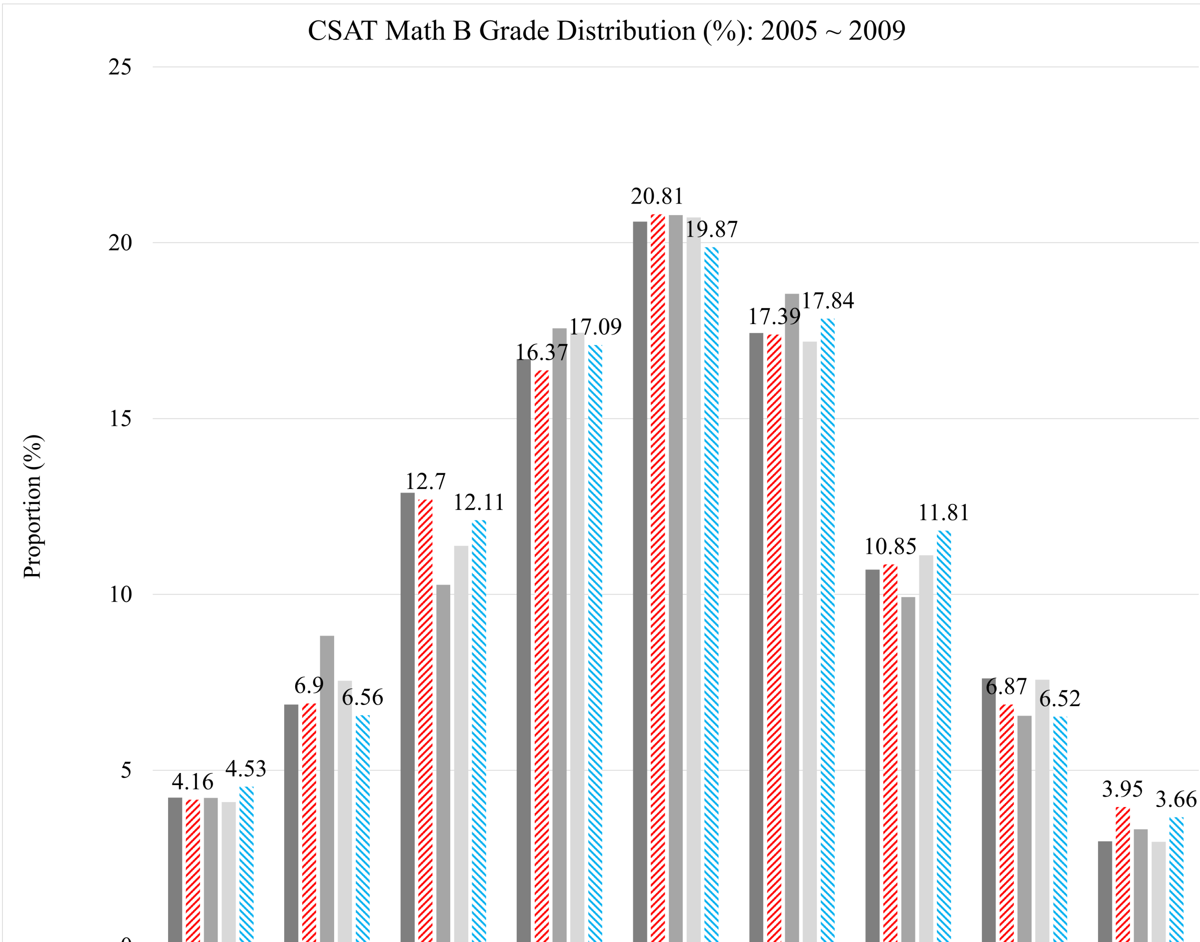
CSAT Math A Grade (1: Highest - 9: Lowest)

09 Math A 08 Math A 07 Math A 06 Math A 05 Math A

a. Source: Ministry of Education.

b. Grade 1 is the highest. Students raw SAT scores are normalized. Each student's percentile is matched to one of nine preset grades in Table 1.1. When there are ties around the grade cutoff, the higher grade will be given to all.

Figure A.2: SAT Math A Grade Distribution

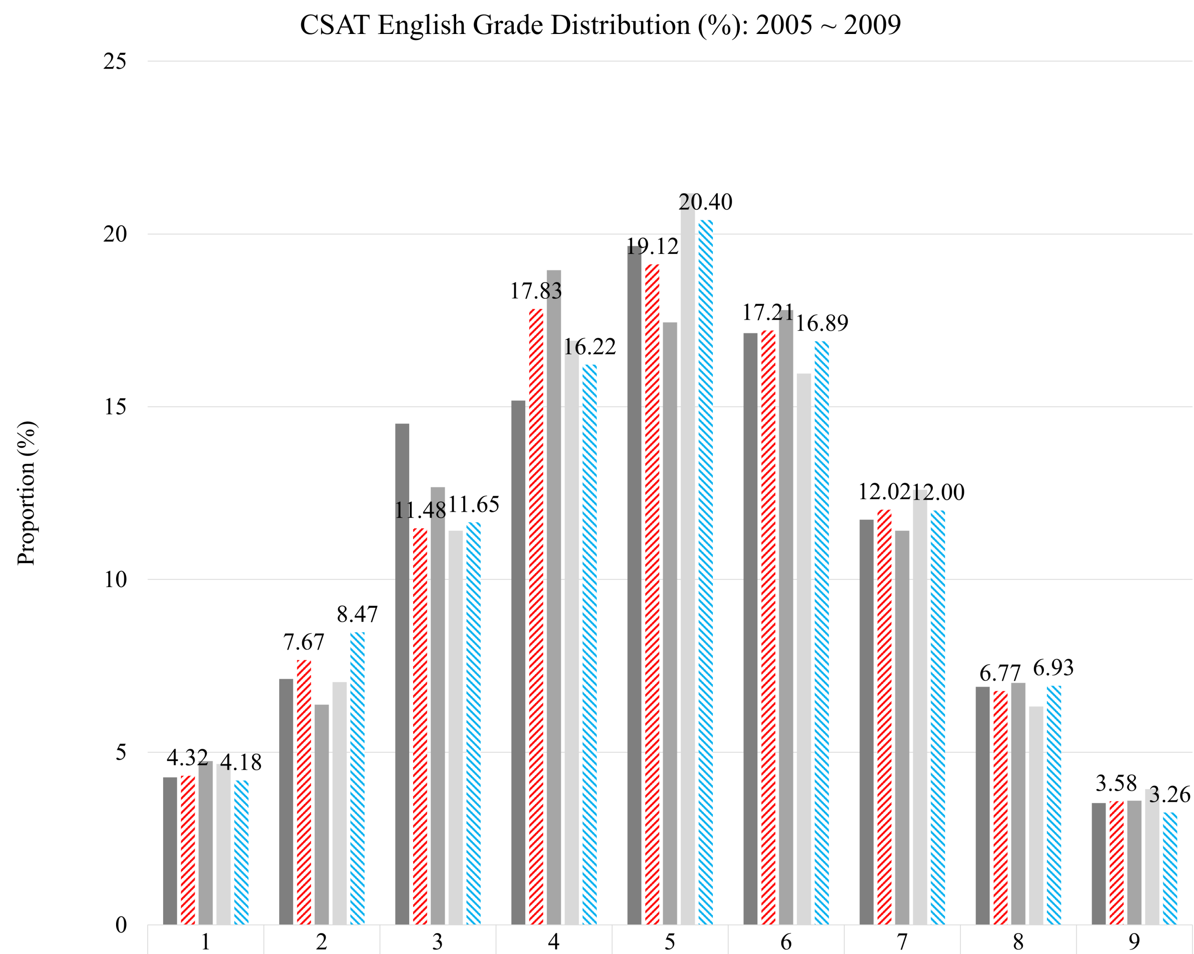


	1	2	3	4	5	6	7	8	9
09 Math B	4.22	6.87	12.89	16.69	20.6	17.43	10.7	7.61	2.98
08 Math B	4.16	6.9	12.7	16.37	20.81	17.39	10.85	6.87	3.95
07 Math B	4.21	8.82	10.27	17.57	20.79	18.55	9.92	6.55	3.32
06 Math B	4.09	7.54	11.38	17.44	20.72	17.19	11.11	7.57	2.96
05 Math B	4.53	6.56	12.11	17.09	19.87	17.84	11.81	6.52	3.66

a. Source: Ministry of Education.

b. Grade 1 is the highest. Students raw SAT scores are normalized. Each student's percentile is matched to one of nine preset grades in Table 1.1. When there are ties around the grade cutoff, the higher grade will be given to all.

Figure A.3: SAT Math B Grade Distribution



	1	2	3	4	5	6	7	8	9
09 English	4.27	7.12	14.51	15.18	19.65	17.13	11.73	6.89	3.53
08 English	4.32	7.67	11.48	17.83	19.12	17.21	12.02	6.77	3.58
07 English	4.74	6.38	12.67	18.95	17.44	17.8	11.41	7.01	3.6
06 English	4.66	7.03	11.41	16.91	21.18	15.96	12.6	6.32	3.93
05 English	4.18	8.47	11.65	16.22	20.40	16.89	12.00	6.93	3.26

CSAT English Grade (1: Highest - 9: Lowest)

09 English 08 English 07 English 06 English 05 English

a. Source: Ministry of Education.
b. Grade 1 is the highest. Students raw SAT scores are normalized. Each student's percentile is matched to one of nine preset grades in Table 1.1. When there are ties around the grade cutoff, the higher grade will be given to all.

Figure A.4: SAT English Grade Distribution

A. TOP GPA STUDENTS

	(1) Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study	(5) Leader Role	(6) Academic Club
Post	0.976**	0.252	0.153	-0.893***	0.048	0.035
	(0.438)	(0.369)	(0.716)	(0.117)	(0.103)	(0.061)
SAT Taker	-1.725***	-1.451***	7.324***	0.907***	0.104	0.104**
	(0.272)	(0.270)	(1.005)	(0.175)	(0.091)	(0.051)
Post × SAT	-0.847*	0.113	-1.601	1.163***	-0.045	-0.108
	(0.489)	(0.405)	(1.269)	(0.314)	(0.122)	(0.072)
N	332	332	332	332	332	332
Mean	3.145	6.059	5.669	1.971	0.370	0.111

B. BOTTOM GPA STUDENTS

	(1) Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study	(5) Leader Role	(6) Academic Club
Post	1.348***	0.332	-0.284	-1.160***	-0.000	-0.025
	(0.363)	(0.228)	(0.341)	(0.083)	(0.030)	(0.020)
SAT Taker	-1.289***	-0.957***	4.540***	0.239***	0.029	-0.001
	(0.176)	(0.139)	(0.615)	(0.085)	(0.032)	(0.022)
Post × SAT	-0.618	0.218	-0.185	0.738	0.103	0.070
	(0.586)	(0.396)	(1.433)	(0.454)	(0.086)	(0.065)
N	540	537	540	540	540	540
Mean	3.921	6.402	2.693	1.227	0.087	0.039

[a] Cluster standard errors by school in parentheses.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Panel A is for the top 11 percentile GPA students (Grade 1 and 2) and Panel B is for the bottom 23 percentile GPA students (Grade 7 – 9).
[d] Units are hours per day for weekday leisure and self study, hours per week for private lesson, and binary for leader role and academic club activities.
[e] Mean outcome is the average of total observations including both cohorts and groups.

Table A.3: Policy Effects on Time Allocation by High School GPA (No Covariates)

MID GPA STUDENTS (No Covariates)

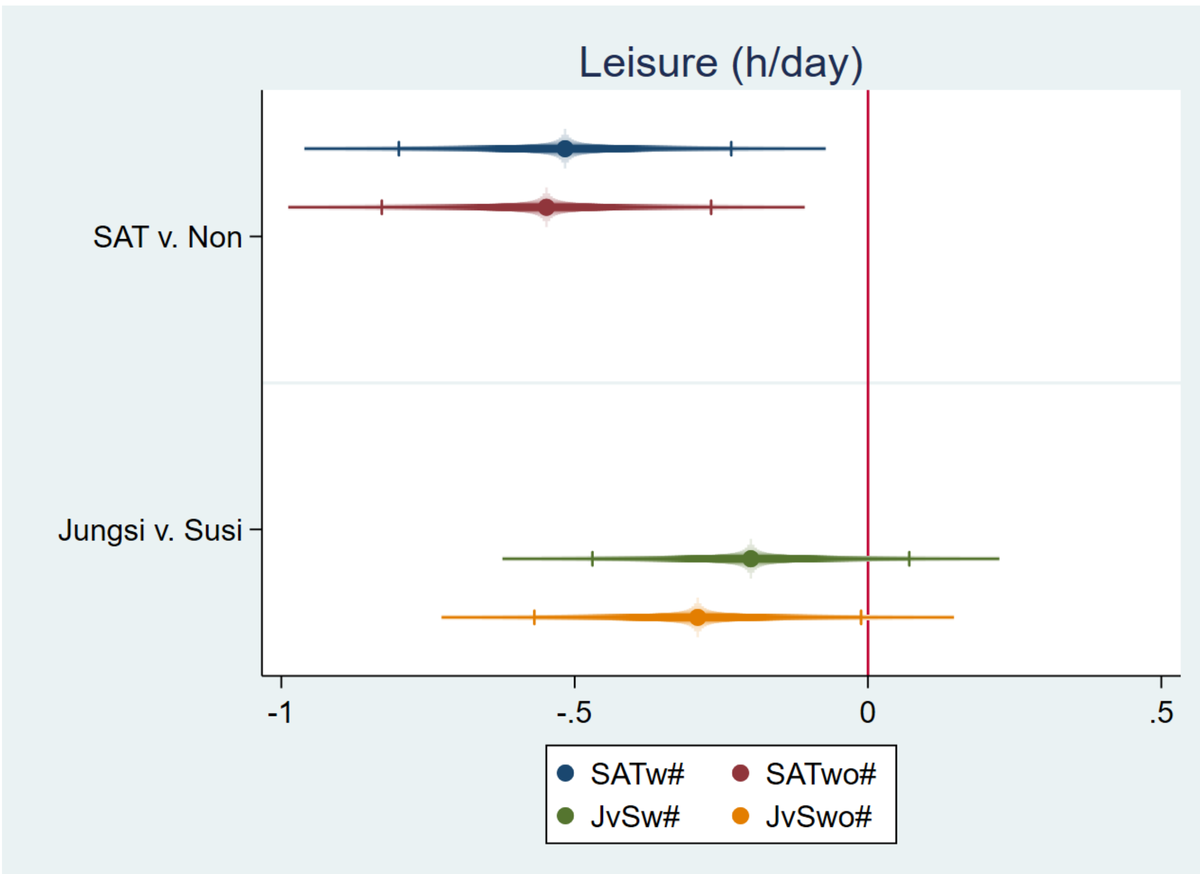
	(1) Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study	(5) Leader Role	(6) Academic Club
Post	0.266*	0.029	0.188	-0.911***	0.028	0.003
	(0.143)	(0.094)	(0.287)	(0.087)	(0.028)	(0.016)
SAT Taker	-1.834***	-1.275***	5.678***	0.592***	0.018	0.063***
	(0.084)	(0.071)	(0.363)	(0.054)	(0.019)	(0.013)
Post × SAT	-0.290	0.277**	1.057*	1.321***	0.037	-0.031
	(0.178)	(0.120)	(0.582)	(0.179)	(0.034)	(0.023)
N	3135	3129	3135	3135	3135	3135
Mean	3.184	6.136	4.800	1.777	0.188	0.090

[a] Cluster standard errors by school in parentheses.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Regression results with no covariates for the mid-class GPA students (Grade 3 and 6).
[d] Units are hours per day for weekday leisure and self study, hours per week for private lesson, and binary for leader role and academic club activities.
[e] Mean outcome is the average of total observations including both cohorts and groups.

Table A.4: Policy Effects on Time Allocation with no Covariates (Mid GPA Students)

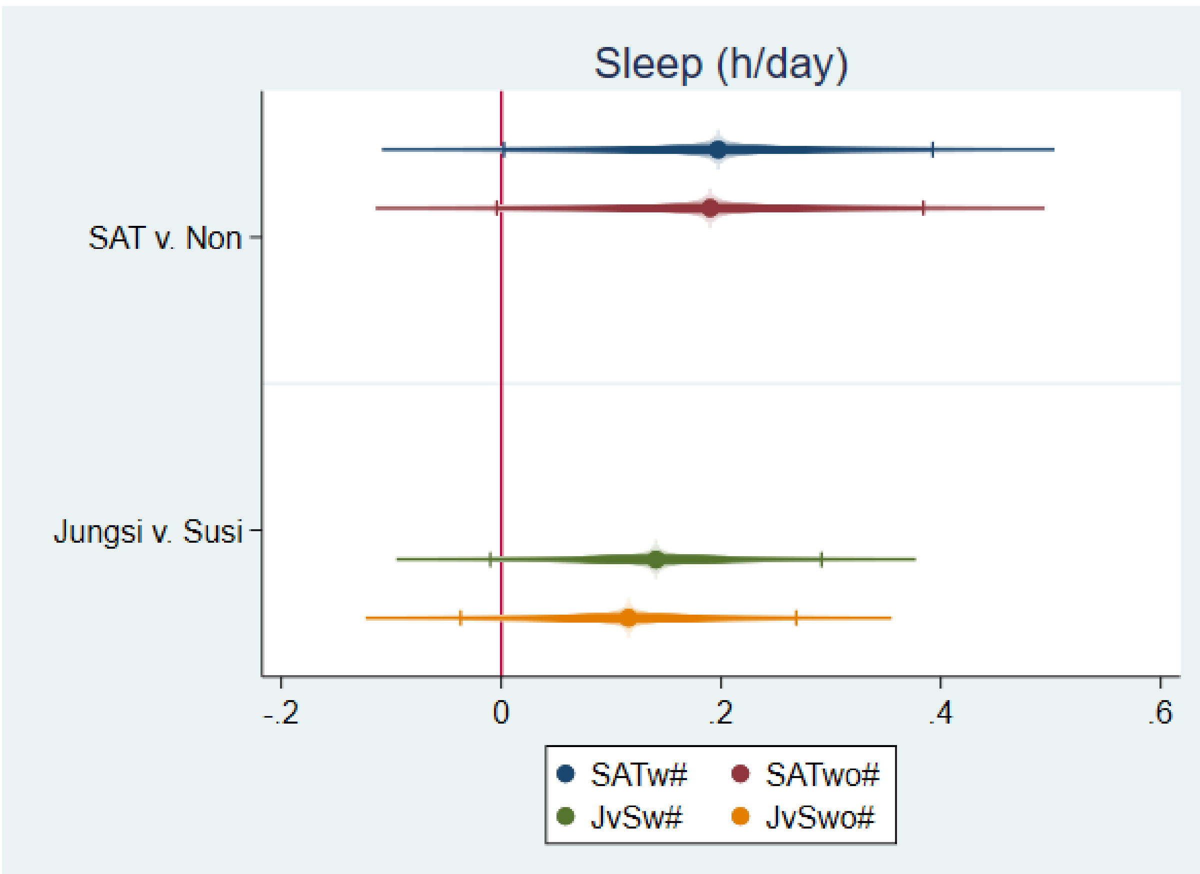
Sample Groups	**Pre**	**Post**	**Total**
Regular	1,631 (54%)	856 (49%)	2,487
Rolling	1,407 (46%)	902 (51%)	2,309
Total	3,038 (100%)	1,758 (100%)	4,796

* Rolling proportion in population: 44% (2005), 53% (2008).

Table A.5: SAT Takers (Regular) v. Susi Students (Rolling) c 

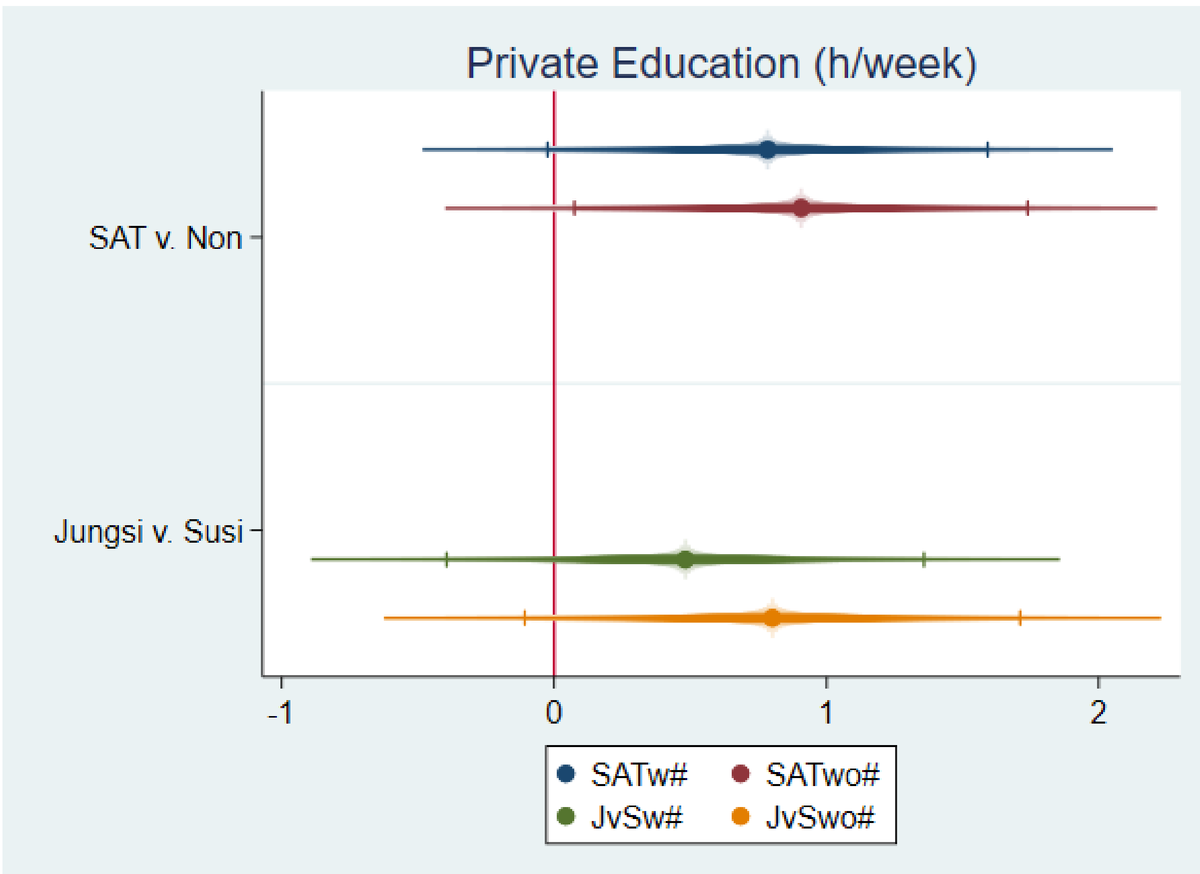
SAT v. Non: Main Specification
Jungsi v. Susi: Alternative Setting
$p < 0.1$ for outside of inner tick, $p < 0.01$ for outside of the end of the bar.

Figure A.5: Average Policy Effect: Daily Leisure c



SAT v. Non: Main Specification
Jungsi v. Susi: Alternative Setting
$p < 0.1$ for outside of inner tick, $p < 0.01$ for outside of the end of the bar.

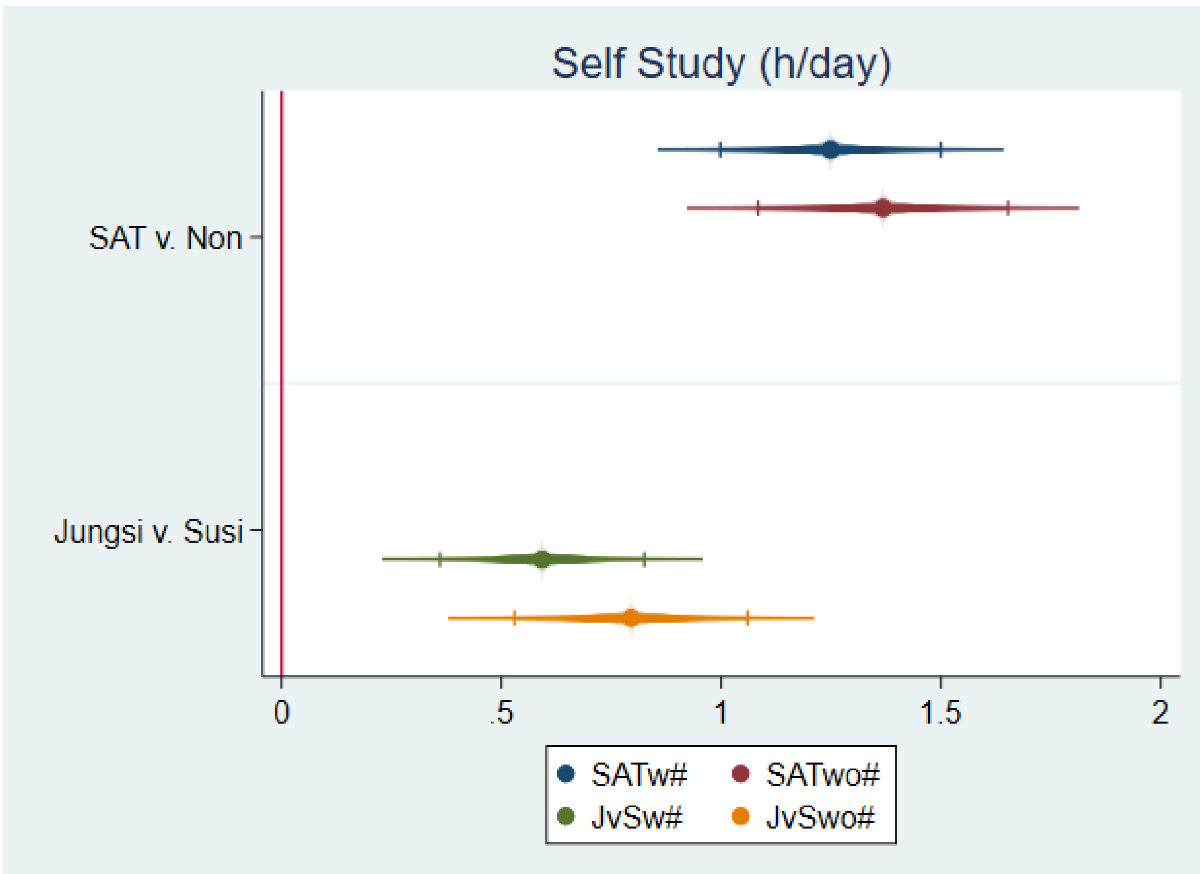
Figure A.6: Average Policy Effect: Daily Sleep



c

SAT v. Non: Main Specification
Jungsi v. Susi: Alternative Setting
$p < 0.1$ for outside of inner tick, $p < 0.01$ for outside of the end of the bar.

Figure A.7: Average Policy Effect: Weekly Private Education



c

SAT v. Non: Main Specification
Jungsi v. Susi: Alternative Setting
$p < 0.1$ for outside of inner tick, $p < 0.01$ for outside of the end of the bar.

Figure A.8: Average Policy Effect: Daily Self Study

Appendix B

CHAPTER 2 APPENDIX

	(1) Weekday Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study
Mean	-0.369**	0.304**	0.935	0.390
	(0.182)	(0.121)	(1.030)	(0.386)
10th	-0.068	0.945**	-0.271	0.307***
	(0.157)	(0.399)	(0.231)	(0.063)
20th	-0.137	0.047	-0.230*	0.368***
	(0.300)	(0.061)	(0.135)	(0.046)
30th	-0.757***	0.112	-0.046	0.452***
	(0.144)	(0.171)	(0.174)	(0.062)
40th	-0.141	0.509	-0.463	0.635***
	(0.277)	(0.397)	(1.594)	(0.058)
50th	-0.694**	0.121	-0.475	0.831***
	(0.335)	(0.303)	(1.692)	(0.160)
60th	-0.681**	0.066	0.667	1.426***
	(0.265)	(0.044)	(1.718)	(0.236)
70th	-0.103	0.129*	1.074	0.505
	(0.350)	(0.072)	(1.763)	(0.652)
80th	0.060	0.659*	0.478	-0.045
	(0.193)	(0.356)	(2.545)	(1.446)
90th	0.371	0.915***	1.445	-0.705
	(0.304)	(0.214)	(2.353)	(1.619)
N	3573	3564	3573	3573

[a] The variance is estimated by bootstraping the results 100 times.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Units are hours per day.

Table B.1: Changes-in-changes Estimates (Discrete)

	(1) Weekday Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study
Mean	-0.364**	0.309**	0.942	0.393
	(0.183)	(0.121)	(1.021)	(0.386)
10th	-0.068	0.945**	-0.263	0.307***
	(0.157)	(0.399)	(0.231)	(0.064)
20th	-0.137	0.047	-0.230*	0.368***
	(0.300)	(0.109)	(0.132)	(0.046)
30th	-0.757***	0.112	-0.007	0.452***
	(0.144)	(0.169)	(0.171)	(0.062)
40th	-0.141	0.509	-0.463	0.635***
	(0.278)	(0.396)	(1.600)	(0.057)
50th	-0.657*	0.132	-0.475	0.831***
	(0.338)	(0.312)	(1.696)	(0.160)
60th	-0.681**	0.066	0.667	1.443***
	(0.268)	(0.044)	(1.720)	(0.236)
70th	-0.103	0.129*	1.074	0.505
	(0.348)	(0.071)	(1.763)	(0.651)
80th	0.060	0.659*	0.478	-0.045
	(0.193)	(0.356)	(2.545)	(1.452)
90th	0.371	0.915***	1.445	-0.705
	(0.304)	(0.214)	(2.344)	(1.619)
N	3573	3564	3573	3573

[a] The variance is estimated by bootstraping the results 100 times.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Units are hours per day.

Table B.2: Changes-in-Changes Estimates (Discrete Upper Bounds)

	(1) Weekday Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study
Mean	-0.374**	0.301**	0.927	0.383
	(0.182)	(0.121)	(1.031)	(0.387)
10th	-0.068	0.926**	-0.271	0.307***
	(0.158)	(0.401)	(0.229)	(0.064)
20th	-0.137	0.047	-0.237*	0.368***
	(0.298)	(0.062)	(0.138)	(0.046)
30th	-0.757***	0.112	-0.092	0.452***
	(0.143)	(0.171)	(0.178)	(0.062)
40th	-0.141	0.509	-0.463	0.635***
	(0.278)	(0.396)	(1.594)	(0.057)
50th	-0.694**	0.118	-0.650	0.831***
	(0.335)	(0.302)	(1.696)	(0.160)
60th	-0.681**	0.062	0.667	1.395***
	(0.265)	(0.043)	(1.723)	(0.234)
70th	-0.103	0.129*	1.074	0.505
	(0.352)	(0.073)	(1.755)	(0.651)
80th	0.060	0.659*	0.478	-0.045
	(0.193)	(0.356)	(2.545)	(1.439)
90th	0.371	0.915***	1.445	-0.705
	(0.304)	(0.213)	(2.353)	(1.619)
N	3573	3564	3573	3573

[a] The variance is estimated by bootstraping the results 100 times.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Units are hours per day.

Table B.3: Changes-in-Changes Estimates (Discrete Lower Bounds)

	(1) Weekday Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study
10th	-0.617***	0.036	-0.247	0.512***
	(0.151)	(0.280)	(0.234)	(0.069)
20th	-0.644***	-0.048	-0.088	0.509***
	(0.151)	(0.089)	(0.225)	(0.073)
30th	-1.270***	-0.026	0.050	0.330***
	(0.158)	(0.088)	(0.247)	(0.082)
40th	-0.658***	0.332	0.975*	0.368***
	(0.184)	(0.284)	(0.514)	(0.081)
50th	-0.618***	0.036	0.732	0.541***
	(0.151)	(0.172)	(0.547)	(0.145)
60th	-0.954***	-0.037	0.607	1.525***
	(0.246)	(0.092)	(0.403)	(0.172)
70th	-0.579***	0.016	0.668	1.989***
	(0.161)	(0.099)	(0.485)	(0.203)
80th	-0.401*	0.695***	0.187	2.293***
	(0.213)	(0.098)	(0.665)	(0.208)
90th	-0.144	0.824***	0.746	2.250***
	(0.323)	(0.224)	(0.877)	(0.148)
N	3573	3564	3573	3573

[a] The variance is estimated by bootstraping the results 100 times.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Units are hours per day.

Table B.4: Quantile DD Estimates

	(1) Weekday Leisure	(2) Sleep	(3) Private Lesson	(4) Self Study
Mean	0.031	0.163*	0.478	-0.006
	(0.127)	(0.093)	(0.643)	(0.154)
10th	0.102	0.306	-0.374	0.194***
	(0.127)	(0.274)	(0.313)	(0.047)
20th	0.058	0.098*	-0.242	0.231***
	(0.080)	(0.052)	(0.178)	(0.040)
30th	0.034	0.103	0.144	0.266***
	(0.165)	(0.100)	(0.196)	(0.043)
40th	-0.068	0.356	0.322	0.463***
	(0.115)	(0.325)	(0.757)	(0.056)
50th	-0.007	0.116	0.267	0.598***
	(0.076)	(0.128)	(0.946)	(0.179)
60th	0.047	0.053	0.429	0.933**
	(0.269)	(0.040)	(0.804)	(0.435)
70th	-0.017	0.098	0.535	-0.845***
	(0.137)	(0.136)	(0.861)	(0.326)
80th	0.194	0.266	1.216	-0.749
	(0.221)	(0.227)	(1.128)	(0.473)
90th	0.586*	0.806***	0.941	-0.622**
	(0.301)	(0.249)	(1.607)	(0.249)

[a] The variance is estimated by bootstraping the results 100 times.
[b] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[c] Units are hours per day.

Table B.5: Changes-in-changes Estimates (Continuous) (Jungsi v. Susi)

VITA

Taehoon Kwon is an applied microeconomist and a deputy director at the Ministry of Agriculture, Food, and Rural Affairs of Republic of Korea. He welcomes your comments to `taehoon.huskies@gmail.com`.