THREE ESSAYS ON EDUCATION POLICIES AND CHILD HEALTH

By

LU YIN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

I dedicate this dissertation to my wonderful family. Particularly to my loving and supportive fiancé, Burhan Ogut, who has never left my side.

ACKNOWLEDGMENTS

I want to thank my dissertation committee members, including David Figlio (chair), Larry Kenny, Mark Rush, and Paul Sindelar for their invaluable guidance, support and encouragement throughout the years. Without them, this dissertation will not be possible. I am especially grateful for their understanding and patience, while I was living away from campus. I feel deeply indebted to my advisor David Figlio, who taught me to put my heart into research and motivated me to strive for better. I cannot hope for a better advisor.

As I executed my dissertation research, I received the help of many people. Alan Guttmacher Institute provided data for the second chapter of this dissertation. Vivian Weis generously shared the report “Between the Lines: States’ Implementation of the Federal Government’s Section 510(b) Abstinence Education Program in Fiscal Year 1998” issued by Sexuality Information and Education Council of the United States. My dissertation has greatly benefited from the comments of those who have read various versions of it, including David Denslow, Damon Clark, Peter Steiner, Steven Slutsky, Jon Hamilton, David Sappington, Ken Mease, and seminar participants at Northwestern University, University of Missouri, Abt. Associates, Urban Institutes, and the American Institutes for Research. I alone am responsible for all errors and oversights.

Finally, the acknowledgements would not be complete without a heartfelt thank you to my family. My fiancé Burhan has always believed in me even when I could not quite believe in myself and has provided tremendous help throughout my research. His patience and encouragement helped me to make the transition from where I came from to where I am now. My parents gave me the courage to pursue my degree in the United States five years ago and without their generous help, I would not have completed my

study. Over the course of this five-year journey, my aunt, Lijun Yang, has been not only my closest family in Florida but also my mentor. She walked me through the highs and lows of my entire doctoral life. Beyond this, I thank her for sharing her academic experiences with me, for listening to my complaints and frustrations, and for believing in me.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BMI	Body Mass Index
ECLS-K	Early Childhood Longitudinal Study Kindergarten cohort
SIECUS	Sexuality Information and Education Council of the United States
ITS	Interrupted Time Series
YRBSS	Youth Risk Behavior Surveillance System
NCLB	No Child Left Behind Act
CDC	Centers for Disease Control and Prevention
DID	Difference-in-Differences
AYP	Average Yearly Progress
POLS	Pooled Ordinary Least Square
BRFSS	Behavioral Risk Factor Surveillance System
NCES	National Center for Education Statistics
IRT	Item Response Theory
ARS	Academic Rating Scale
SES	Socioeconomic Status
PSS	Private School Survey
CCD	Common Core of Data

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

THREE ESSAYS ON EDUCATION POLICIES AND CHILD HEALTH

By

Lu Yin

August 2010

Chair: David Figlio
Major: Economics

Studies in education literature have been focusing on the impact of policies on children's academic performances. However, under single incentive, educators and parents pursuing higher test score gains may undertake ways that unintentionally harm children's health. Understanding how and whether certain education policies post negative impacts on children's health is therefore very important, and has become the main research focus of this dissertation. In the first essay, we assess whether adolescent obesity has, in part, been driven by a factor previously overlooked: the school accountability movement. Exploiting the variation in the timing of the introduction of state accountability systems and in the grades to which those systems applied across states and over time, we find robust and consistent evidence across models that school accountability systems significantly increase students' Body Mass Index and explain a significant amount of the growth in adolescent obesity. These findings suggest that while accountability systems may promote students academic performance, this could occur at the expense of students' physical health. Policies that mitigate these effects could help to slow the rise in adolescent obesity. In the second essay, we undertake a comprehensive study of the effects of teachers' evaluation standards on children's

probability of being obese and BMI using a rich student-level data from the Early Childhood Longitudinal Study-Kindergarten cohort of 1998 (ECLS-K) including interviews with parents, data from principals and teachers as well as direct child assessments. In models in which we control for student-level fixed effects, we find strong evidence that with the increase of teacher evaluation standards students tend to have higher BMI and are more likely to be overweight. The third essay examines the impact of State-Sex-Education policies as well as the new Personal Responsibility and Work Opportunity Reconciliation Act of 1996, the Title V, Section 510 Abstinence Education Program on adolescent risk behaviors. To account for the potential differential impacts of 1996 Title V Section 510 had upon state sex education mandates which will subsequently bias our analysis, we thus employ an Interrupted Time-Series design that first exploits the impact of 1996 reform on state sex education legislatives and identify their effects on adolescent sexual behaviors subsequently. First, we find that neither abstinence-only nor comprehensive sex education decrease the probability of being sexually active or increase the likelihood of performing safe sex. Instead, we find that abstinence-only lower the probability of using condoms and birth control pills relative to not using any birth control method. Second, using the ITS model, we find that the trend in percentage of students who had sex (percentage of students who had sex before 13) decreases by 0.5% in high-implement states relative to low-implement states post 1996 reform.

CHAPTER 1
INTRODUCTION

Education policies are designed to promote student academic performance, narrow racial and ethnicity achievement gap and improve student readiness for college. Therefore, studies in education literature have been focusing on the impact of policies on children’s academic performances. However, under such single incentive educators and parents may undertake ways that unintentionally harm children’s health. Understanding how and whether certain education policies post negative impacts on children’s health is therefore very important, and has become the main research focus of this dissertation.

The adolescent obesity rate is tripled over the last three decades to 17%. In the second chapter, we assess whether this has, in part, been driven by a factor previously overlooked: the school accountability movement. Three potential reasons may underlie this phenomenon, increased stress, reduced physical activity or worsened diets. Exploiting the variation in the timing of the introduction of state accountability systems and in the grades to which those systems applied across states and over time, this paper estimates the effects of school accountability on students' weight status using a rich panel of student-level data. Three identification strategies are adopted in the empirical analysis, a Difference-in-Differences design to identify the short run treatment effect, a state fixed effects strategy and a state fixed-trend effects model. We find robust and consistent evidence across models that school accountability systems significantly increase students' Body Mass Index and explain a significant amount of the growth in adolescent obesity. We further estimate the causal effects on the treated of different lengths and frequencies of exposure to the school accountability systems and find that the effects are increasing in the length/frequency of exposure to the systems with a decreasing marginal effect. “Placebo tests” are performed to gauge the credibility of these conclusions. Consistent with these findings, we also find that at least among females, school accountability decreases the number of times that students participate in PE classes. These findings suggest that while accountability systems may promote students academic performance, this could occur at the expense of students' physical health. Policies that mitigate these effects could help to slow the rise in adolescent obesity.

High performance standards have been advocated by school administrators, teachers, parents, and even by students themselves in education for centuries, however, little is known about their effects on student health outcomes. In the third chapter, we undertake a comprehensive study of the effects of teachers' evaluation standards on children's probability of being obese and BMI using a rich student-level data from the Early Childhood Longitudinal Study-Kindergarten cohort of 1998 (ECLS-K) including interviews with parents, data from principals and teachers as well as direct child assessments. To account for the potential endogeneity problem due to simultaneity between BMI and teachers' grading standards, we first matched students' direct assessments scores horizontally and then divide students into high standard cohort and low standard cohort based on their teachers' ratings of English, mathematics, and science skills. We then estimate the effects of teachers' heterogeneous standards on students' BMI and the probability of being overweight given that students perform equally in direct assessment which includes reading, math, and science. Furthermore, we evaluate the impact of students' self-assessment of their academic and social skills on their own BMI and the probability of being overweight using the same method.

The fourth chapter examines the impact of State-Sex-Education policies as well as the new Personal Responsibility and Work Opportunity Reconciliation Act of 1996, the Title V, Section 510 Abstinence Education Program on adolescent risk behaviors. While intended to prevent student risky behaviors and hence promote adolescent health and quality of education, there is little empirical evidence about the effectiveness and efficacy of such state level policies, especially after the implementation of Title V Section 510. Using a rich panel of student-level data between year 1993 and 2005, available from Youth Risk Behavior Surveillance System (YRBSS), we first explore the effects of different state sex education mandates on adolescent sexual behaviors using various panel data techniques. To account for the potential differential impacts of 1996 Title V Section 510 had upon state sex education mandates which will subsequently bias our analysis, we thus employ an Interrupted Time Series design that first exploits the impact of 1996 reform on state sex education legislatives and identify their effects on adolescent sexual behaviors subsequently. First, we find that neither abstinence-only nor comprehensive sex education decrease the probability of being sexually active or increase the likelihood of performing safe sex. Instead, we find that abstinence-only lower the probability of using condoms and birth control pills relative to not using any birth control method. Second, using the ITS model, we find that the trend in percentage of students who had sex (percentage of students who had sex before 13) decreases by 0.5% in high-implement states relative to low-implement states post 1996 reform.

CHAPTER 2
ARE SCHOOL ACCOUNTABILITY SYSTEMS CONTRIBUTING TO ADOLESCENT OBESITY?

Introduction

The striking rise in childhood obesity has attracted national attention over the last several decades. Between the 1970s and 1990s, the fraction of overweight children and adolescents tripled from 5 percent to 17 percent. There are several public health and economic reasons to be concerned with this sharp increase. Being overweight in childhood increases the risk of being an overweight adult (Guo et al. 2002), coronary calcification in adulthood (Freedman et al. 2001), and cardiovascular disease mortality (Must et al.1999; Lakka et al. 2002; Mokdad et al. 2003). Additionally, greater health, social, and economic costs are known to be associated with obesity. Each year, obesity causes at least 400,000 deaths in the United States and the costs of providing health care to obese American adults and children have been estimated to be approximately $147 billion in 2008 (Finkelstein et al. 2009).

The existing literature in medicine and social sciences on the causes of obesity can be divided into studies directly examining genetic[1] (Perusse et al. 2001; Katzmarzyk et al.1999) or behavioral factors (Stunkard et al. 1990; Dietz 1994), and studies investigating environmental factors, for example, family structure and socioeconomic status (Locard et al. 1992; Woolston 1987; Bar-Or et al. 1998). Recently, researchers have applied the tools of economics to explore the causes and consequences of rising obesity in the United States. Among the most prominent causes

[1] Research on human genetic mapping has identified several specific genes and gene mutations that are believed to cause human obesity (Perusse et al. 2001). Research found that a child born to a parent in the upper five percent of the BMI distribution has a 60 to100 percent greater risk of being obese compared to a child born to parents with normal, healthy weights (Katzmarzyk et al.1999).

of obesity examined in the literature, the leading three are lack of physical activity[2] (Lakdawalla and Philipson 2002), increased television viewing[3] (Eisenmann, Bartee and Wang, 2002; Reilly et al. 2005) and fast food consumption[4] (Cutler et al. 2003; Chou et al. 2004). Only a few studies focus specifically on childhood and adolescent obesity. Cawley, Meyerhoefer, and Newhouse (2006) explore the impact of state Physical Education (PE) requirements on youth physical activity and obesity. Although the state requirements do increase student PE time, they find no evidence that PE lowers BMI or reduces the probability that a student is overweight. Schanzenbach (forthcoming) finds that children who consume school lunches are more likely to be obese than those who brown bag their lunches. Anderson, Butcher, and Levine (2003) examine the increase in maternal employment as a potential source of childhood overweight. As the increased value of time may lead more people to shift to fast-food alternatives that are often calorie and fat laden. They conclude that the intensity of a mother's employment over a child's lifetime has a positive effect on a child's likelihood of being overweight, given the child is in a high income family, with a well-educated or white mother.

While the previous research provides important insight into the actions and forces that induce individuals to gain excessive weight over time, they are unlikely to be the sole causes of this epidemic. This paper considers a potentially important additional

[2] Lakdawalla and Philipson (2002) estimate that one-third of the increase in BMI in recent decades can be attributed to falling food prices and two-thirds due to decreased exertion in normal work activity.

[3] Eisenmann, Bartee and Wang (2001) examined Physical Activity, TV Viewing, and Weight in U.S. Youth using 1999 Youth Risk Behavior Survey and found that physical inactivity and frequent television viewing may contribute to obsity. Reilly et al. 2005 showed that sedentary activities such as watching television and playing video games significantly increase rates of obesity in children as young as three years old.

[4] Cutler et al. (2003) and Chou et al. (2003) found that body weight and obesity prevalence increase significantly as the per capita number of restaurants and the real price of cigarettes go up, suggesting that more eating out and less smoking may have contributed to the rise in obesity.

explanation: the increased high-stakes testing systems for students and schools. Increased accountability for students began in earnest in the early 1970s, and since then many states have used “high-stakes” achievement or proficiency tests to hold students accountable for meeting state-mandated educational achievement standards. Beginning in the early 1990s, this accountability based on high-stakes testing was applied to schools in addition to students, and many states started their own test-based accountability systems. The No Child Left Behind Act of 2001 furthered this national trend toward increased school accountability, one in which dozens of states formally grade or evaluate their schools based on student test performance. Rather than evaluating the academic achievement intended to be accomplished through the systems[5], this paper contributes a new angle by focusing on an overlooked issue associated with school accountability: the possibility that increases in implementation of school accountability across states raise the prevalence of overweight among adolescents.

Three potential reasons may underlie this phenomenon. First, school accountability may increase weight gain by imposing additional stress on students, which has been shown to be associated with obesity in small scale studies. Over the last several decades, social scientists have devoted a considerable amount of effort to identifying the effects of various psychological states on obesity (Friedman and Brownell 1995). Their findings revealed that obesity is generally the result of a lifestyle that is plagued by stress, pressure, boredom, and poor self image. With the increased course work at school and the unavoidable pressures from various tests, young people facing

[5] See Koretz and Barron (1998), Clark (2003), Haney (2000, 2002), Figlio and Rouse (2006), West and Peterson (2006) and Chakrabarti (2006).

school accountability are now under higher levels of stress, which may directly relate to their mental and physical health. A recent study shows that children residing in states with more stringent accountability laws are more likely to be diagnosed with Attention Deficit/Hyperactivity Disorder (ADHD) and consequently are more likely to be prescribed psychostimulant drugs for controlling the symptoms[6] (Bokhari and Schnedier 2009). Other studies assessing the relation between homework and school exams and stress indicate that pressure from teachers, grades, and homework is one of the primary sources of stress and depression for teens (Lindblad 2006; Kang 1997; Kouzma 2002). Depressed feelings in adolescence are, in turn, associated with an increased risk for the development and persistence of obesity over one year (Goodman and Whitaker 2002).

Second, students, facing likely punishment for poor test outcomes, such as repeating the same grade or graduating without honors, may sacrifice much-needed physical activity time for more study time. In addition, facing a need to respond to improve academic test scores and fend off potential funding losses, school administrators may change policies in order to stimulate student performance. Rouse, Hannaway, Goldhaber, and Figlio (2007 NBER working paper) show that schools under higher pressure to perform well tend to change their policies in order to improve students' performances, such as requiring summer school, before/after school tutoring and Saturday classes. The direct effect of these policies will be decreased physical activity time in or after school and longer study/homework time, all of which may be correlated with young people being overweight.

[6] One should notice that another potential reason for the increase in ADHD diagnosis and psychostimulant consumption rates may due to schools labeling marginal students such that to reshape the testing pool or students are now receiving more appropriate diagnosis.

Third, to respond to the academic improvement focused testing system schools may reduce the nutritional content of students' food consumption in schools. Figlio and Winicki (2005) find evidence that schools respond to accountability pressures, by substantially increasing the calorific content of school menus on testing days. Anderson and Butcher (2005) find that schools under financial pressure due to the accountability system[7] are more likely to make junk food available to their students.

To investigate the relationship between school accountability and adolescent obesity I exploit unevenness in the timing of the introduction of school accountability systems and the grade levels to which these applied using a rich panel of student-level data, available from the Youth Risk Behavior Surveillance System (YRBSS) developed by Centers for Disease Control and Prevention (CDC). Three identification strategies are adopted in the empirical analysis, a Difference-in-Differences design to identify the short run treatment effect, a state fixed effects strategy and a state fixed-trend effects model where I construct a state BMI trend index for each state to control for potential heterogeneous BMI growth trends. Regression results suggest that school accountability systems have positive and statistically significant effects on student BMI and growth in adolescent obesity controlling for individual-level measures, state policy covariates and state BMI growth trend. This effect is bigger in the short run (within 4 years) than in the long run (more than 4 years), among female students and among Asian and Hispanic students compared to White and African-American students. Additionally, school accountability decreases the number of times female students

[7] Schools that got failing grades will suffer consequences; hence schools may try to raise money in order to strengthen core academics to meet the achievement goals without cutting elective courses. They could achieve it through soft drink and vending contracts, or through other snack food sales.

participate in physical education classes, but has no statistically significant effect on male students. These results are consistent with a simultaneous work by Anderson, Butcher and Schanzenbach (2009) where they use school level data from Arkansas and find that schools that were below the average yearly progress (AYP) threshold under NCLB in year t-1 have a small, but statistically significantly higher rate of overweight and obesity in year t.

Findings of this study suggest that a part of the recent rise in adolescent obesity can be attributed to the school accountability movement. Although this represents only a small portion of the overall growth, it is a portion that may be responsive to policies. Since many of the determinants of this rise may be invariant to education or health policies, this is nevertheless an important conclusion. The rest of the paper is organized as follows. The next section presents a brief background of the accountability systems and section three describes the data set followed by the discussion of empirical framework and methods used in the analysis. Section IV presents the results followed by falsification tests and Section VI concludes.

Accountability Systems

Increased accountability for students began in 1970s and these testing programs are generally designed to measure the degree of skills possessed by students. For example, high school exit examinations were first integrated into the student accountability reform movement in 1970s after unfavorable international comparisons of mathematics and science achievement among U.S. students with that of students in other industrialized nations. Since then educators and policy makers are working to bring value to the high school diploma by raising the rigor of high school standards, assessments and curriculum and aligning expectations with the demands of postsecondary education and work. This accountability system was applied to schools in the early 1990s.

In 1993, Wisconsin, North Carolina, Taxes and Connecticut were the first four states to implement accountability at different grades. 12 states had accountability systems at the school level by 1996, and 39 states did so by 2000. The No Child Left Behind Act of 2001 (NCLB) provided more extensive federal requirements for states, student assessment, and school accountability systems than had ever previously existed. Each state develops its own accountability system based upon the state's content and achievement standards, valid and reliable measures of academic achievement, and other key indicators of school and district performance such as attendance and graduation rates. By 2003, all the states were required to implement school accountability that formally grade or evaluate their schools based on student test performance and make report cards publicly available. To better illustrate the different timing of school accountability introduction dates, these data are presented in the form of maps that appears in Figure 2-6.

Additionally, accountability system was not applied to all grades when it was initially introduced. Specifically, around 14 states introduced the system into 1 to 3 different grades, 13 states applied it to 4 to 6 grades and other states selected more than 7 grades to apply the system. Under NCLB, however, all states were required to administer assessments in reading and mathematics every year to all students in grades 3 through 8 and once during high school by the end of school year 2005-06 as illustrated in Figure 2-2.

In the current accountability environment, policymakers want tests to have stakes for test-takers attached to them so that students will exert greater effort to pass them and use test results to reward schools with additional funding from the government or to punish the schools that failed to meet the academic yearly progress (AYP) goals. Different from previous education policy mainly focused on providing resources, school accountability systems provide additional incentive for schools to maintain high academic performances by requiring schools publish their annual test scores to the public. To respond to school accountability pressures, schools may reallocate resources, both time and financial assets, toward testing subjects and unintentionally harm students' physical and mental health.

Data

To test the proposed hypothesis, I make use of the Youth Risk Behavior Surveillance System (YRBSS) for the years 1999, 2001, 2003, and 2005 which has been administered biennially since 1991 by the Centers for Disease Control and Prevention (CDC). Starting in year 1999, CDC collected data on students' weight and height; therefore data before 1999 are not included in this analysis. The YRBSS system, established in 1991, monitors high school students' risky behaviors, including those relating to obesity, eating habits, physical activities, and other risky behaviors. It includes national, state, and local school-based surveys of nationally representative samples of 9th through 12th grade students. In order to implement this across-state analysis, the restricted version of the YRBSS data with state identifiers was acquired. In the analysis, observations were dropped if a student's weight or height was missing, reducing the pooled sample size from 57,826 to 54,065, a reduction of around 7%. Schools with relatively high numbers of African-American and Hispanic students were oversampled to ensure adequate representation of these groups. A weight is applied to each record to adjust for non-response and the distribution of students by grade, sex, and race/ethnicity in each state.

Self-reported data on height and weight allow me to construct three outcome variables: the body mass index (BMI) of each respondent and indicators of whether he or she is overweight or underweight. Brener et al. (2003) show that self reported height and weight by high school students are valid proxies for measured values, but these might underestimate the prevalence of overweight. Nevertheless, weight status variables will be the main dependent variables in all the regression analyses, and measurement error present at left hand side will not affect the accuracy of the estimates. According to the CDC growth charts in the United States, the 97th percentile of student BMI, ages between 15 and 18 years old, is below 34. The 5th percentile of student BMI, ages between 15 and 18 years old, is above 14 (Centers for Disease Control and Prevention, 2000) Therefore, observations with BMI>40 as well as BMI<10 are excluded from the data set which may be subject to measurement errors and this again reduced the sample size to 53,499, around a 1% fall.

The trends in student BMI and the percentages of the student obese and overweight[8] in the YRBSS are plotted in Figure 2-3[9]. Between 1999 and 2005, BMI increased by 0.4 kg/m^2 or by 2.8% for all students while the average BMI is higher for male students and the percentages of overweight and at the risk of becoming

[8] Children who are overweight or at risk of becoming overweight are defined as a BMI above the 85th percentile for children of the same age and gender in 2000 (defined by Centers for Disease Control and Prevention).

[9] The values of BMI and overweight are computed based on YRBSS sampling weights which produce nationally representative figures as of each survey year.

overweight increased over 3% for female students. One might notice that we observe a small dip in the BMI and obesity trends in year 2001 which might due to the fact that not every state participated in YRBSS in every round of the survey as showed in Figure 2-5[10]. Therefore, I plot the same graphs only using data from states that conducted all 4 waves of YRBSS. Reassuringly, the BMI and obesity trends increase consistently over the 4 survey years in this case. Relative to Asian students, White, Hispanic, and African-American students have higher BMI while White students have lower BMI compared to African-American and Hispanic students. On average, 7.8% of Asian students are overweight, yet this number is significantly higher, 20.5%, for Hispanic and African-American students. From the graph, we can see that BMI increased monotonically as the student progressed from grade 9 to grade 12, while the percentage of student overweight did not show any systematic pattern across grades. In 1999, 15.5% of ninth graders were overweight, which was higher than any other grades. In 2005, however, 18% of twelfth graders were overweight, highest among all grades. In addition to main individual control measures, such as gender, race, and grade, four variables related to food consumption[11] are also included in the later regressions under the hypothesis that unhealthy eating habits contribute to extra weight gain. I also examine three variables pertaining to students' physical activities: numbers of times a student participated in vigorous exercise, moderate exercise, and PE classes in the last seven days. These variables are used as outcome variables under the hypothesis that

[10] States apply for funding to conduct the YRBS every survey year. However, for various reasons not every state will apply for the money.

[11] Questions on the YRBSS survey are: during the past 7 days, how many times did you eat fruit/green salad/potatoes/carrots etc.? Regression results excluding these variables are not significantly different. (Results are available upon request.)

with the introduction of accountability, students may reduce the amount of physical activity. Following prior studies, I also added TV viewing, measured as number of hours of TV watched per day for each individual, into the accountability and BMI equation to control for its potential positive effects on adolescent obesity.

Definitions, means, and standard deviations of all variables employed in the regressions appear in Table 2-1. Except where noted, they are based on the sample of 53,499 that emerges when observations with missing values are deleted. The summary statistics in the table and all regressions in this paper are computed based on YRBSS sampling weights and are representative of the population at large. According to YRBSS, the target population consisted of all public, Catholic, and other private school students in grades 9 through 12 (Methodology of the Youth Risk Behavior Surveillance System 2004). The main concern with this is that only public schools are legally bound by accountability related policies; including both public and private schools in the dataset will potentially underestimate the impact of accountability on student BMI and obese prevalence. As CDC does not release school-level identifiers and the data does not contain any internal information that may help me to identify school types[12], to determine in which states' private schools were part of the sample becomes impossible. However, sampling schools with any of grades 9 to 12 were selected with probability proportional to school enrollment size. Thus most of the surveyed schools are public schools due to the bigger enrollment size compared with private schools. In addition, YRBSS is the only nationwide survey that includes questions on BMI, eating habits, and

[12] For instance, we may use school size or grades span as proxies for public schools. Unfortunately, the YRBSS does not include such information.

physical activities starting from year 1999, which makes YRBSS the most appropriate dataset for this study.

The key components of the analysis are the variables used to measure the accountability systems. I obtain the date of introduction of school accountability system from the Consortium for Policy Research in Education (CPRE) and then collect data on testing grades from the website of Department of Education for each state, shown in Table 2-2 column 2 and 3. As treatment may differ not only in the length of accountability treatment, but also vary by the intensity of treatment, exploiting the variations in the introduction dates of school accountability and in testing grades across states, I construct two micro level independent variables which equal to the number of years an individual has been exposed to the system (LENGTH_grade) and the number of times an individual has been tested (FREQUENCY_grade) at each survey year since the first time she/he has been tested for the accountability purpose, respectively. For example, in Texas where school accountability was implemented in 1993 for grades 3 to 8 and grade 10, if a student from 9th grade was surveyed in year 2001, the number of years she has been exposed to the system equals to 7 and the number of times she has been tested for the accountability purpose, in this case, equals to 6[13]. One potential problem with the LENGTH variable is that the official implementation date of school accountability policies for each state was used in calculating the exposure variable; however some states adopted an unofficial but similar system at an earlier time. For instance, the Florida school accountability system was announced to be implemented in

[13] A 9th grader in 2001 would be in 1st grade in 1993, so the first time she/he be treated under SA would be in 1995 when she/he was in 3rd grade. LENGTH_grade=2001-1995+1=7. FREQUENCY_grade=6 (grade 3 in 1995 to grade 8 in 2000. This student is not tested in grade 9 in 2001.)

1999, yet a similar system actually started running several years earlier, in 1994. If this is the case, the estimated effects of school accountability might be downward biased or I might not observe an adjustment period of student BMI for these states. One should also notice that to comply with the NCLB 2005-06 deadline, many states introduced new tests or tests in additional grades during the past few school years. Without taking into account of such changes, both LENGTH and FREQUENCY variables will be misspecified. I therefore collected data about major changes made in the states testing grades since 2002 from State Achievement Profiles conducted by Center on Education Policy and calculate the two main independent variables to account for these changes. Further, it is likely that not only students who are directly tested under school accountability systems experience higher weight gain, but also students who anticipate future tests. Therefore, using the variation across states and over time in introduction of accountability laws, I construct a separate variable which equals to the number of years a state had implemented the school accountability system to capture the overall effects on the entire student body (LENGTH_general).

My research design exploits the substantial variation in students' weight status across states in the timing of the enactment of pre and post school accountability laws. Controlling for state trend in the growth of BMI becomes crucial when trying to identify a cross state effect of school accountability on students BMI. On one hand, the fact that not every state participated in YRBSS in every round of the survey hinders the estimation of state trend in years 1999 to 2005 directly using YRBSS. On the other hand, BMI trend calculated using adolescent data will be plagued by the impact of school accountability and hence result in biased estimates. Instead, I employ data from the Behavioral Risk Factor Surveillance System (BRFSS), a large-scale state-based system of health surveys of approximately 350,000 adults annually that is conducted by the CDC and by 1994, all states were participating the BRFSS. The BRFSS is a consistent source of information on health risk behaviors at state level, including respondent's height and weight information, which allow the author to calculate her/his BMI and information on demographic factors, education and eating habits which are the same as the questions asked in YRBSS. I use BRFSS from 1999 through 2005 which spans the same period as YRBSS, to estimate a BMI growth rate for each state which is later used as a control variable in the accountability-weight status analysis.

Since the school accountability systems are ultimately measured at the state level, I incorporate additional state-specific measures in the analyses to capture time-varying trends within areas. Specifically, three state level control variables are augmented with YRBSS and state level measures pertaining to school accountability policy and testing grades in later analysis. They are percentage of individuals below poverty line used to control for states' socioeconomic status and the percentage of the state population with a bachelor's degree used to control for parents' education effects on youth BMI (both obtained from the US Census Bureau, Population Division), both of which in some studies show negative relation with adult BMI. The hypothesis is more schooling of a parent leads to lower levels of obesity for her children as she may influence kids' choice of food and lifestyle. Last, I obtain data on percentage of public schools providing AP courses for each state in each survey year (from College Board's Advanced Placement (AP) Program website)[14]. States established various high school honor diploma

[14] College Board's Advanced Placement (AP) Program was started nearly four decades ago to enable students to complete college-level studies while still in high school and to obtain college credit or

programs based on students' regular classes as well as AP course performance to encourage high school students to take challenging and rigorous courses. Therefore, the variable, which is coded as percentage of public schools providing AP courses, measures the effect of increased high school academic options on student BMI.

Empirical Strategy

The primary aim of this study is to evaluate the impact of school accountability on students' weight status. As the implementation map showed, twenty-seven states implemented school accountability after 1999 and among them ten states did not adopt this policy until 2003. This uneven timing allows me to employ a Difference-in-differences (DID) design to identify the short run effect that school accountability had upon the student BMI of those who are in the treated states. Specifically, using data from years 1999 and 2003, I classify states with school accountability implementation date between 1999 and 2002 as a treatment group. On average, students in the treatment group are exposed to the school accountability systems for 2.4 years. The choice of comparison group is crucial in DID analysis. I use students who have not been affected by the school accountability systems and reside in states which started the policy in 2003 as a control group based on the assumption that students' weight status will not change immediately after the implementation of school accountability systems. A DID estimator exploits the existence of a comparison group in an attempt to estimate the impact of the treatment (in this case, school accountability) on the eligible group (in this case, states which employed accountability systems before 2003). This relies on the assumption that there are no other contemporaneous shocks affecting the relative

placement. Today more than 500,000 students in about half of the nation's high schools take at least one AP course.

BMI of treatment and comparison groups. This assumption is fairly satisfied because during the time period there were no major health or education reforms that might affect student BMI to a great extent. Even if the assumption of common shocks is not satisfied fully, it does become more reasonable once I condition upon observable characteristics. The DID model estimates:

(1) $$BMI_{ist} = \alpha + \beta_1 Treat_{is} + \beta_2 Year_t + \beta_3 (Treat_{is} \times Year_t) + \delta' X_{ist} + \gamma' State_{st} + \varepsilon_{ist}$$

Equation 1 posits that the BMI, for the ith student residing in state s during year t, is a function of pre-post school accountability indicators, measured here by Treat a dummy for treatment group (1 if treated, 0 if control), Year a dummy for before or after policy (1 if after, 0 if before) and the interaction term capturing the variations to the treated students (relative to the untreated students) in the year after the policy (relative to the year before the policy). In addition, BMI depends on a vector of individual characteristics (X) and a vector of time-varying state characteristics (State). This DID estimator illustrates the unique identification strategy made possible by the YRBSS data. This approach differences out the time-invariant cross-sectional heterogeneity and relies instead on comparing the changes within the groups that did or did not face school accountability treatment over this period.

For this estimator to identify the impact of school accountability, it must be true that the average change in student BMI for the treatment group without the introduction of school accountability systems would have been on the same trajectory as the student BMI for the control group. In this case, both control and treatment groups must be randomly selected. Students should not be self selected to be in either group, yet the implementation of school accountability must be exogenous to their eligibility. However, the exogeneity assumption, $E(\varepsilon_{ist} | Treat_{ist}) = 0$, may not hold in this case. For instance, most states in the southeast regions adopted school accountability relatively earlier than other states, and this is the case partially because students in the South are traditionally have ranked at the bottom on National Assessment of Educational Progress (NAEP) scores, college entrance tests, and other indicators of educational achievement. If this is the case, the DID design will not be the most reliable strategy. To overcome the difficulties stated above to the most extent, the data obtained from Youth Risk Behavior Survey enables the use of panel data techniques.

Consider a simple model of adolescent BMI:

(2) $$BMI_{ist} = \beta_0 + \alpha\ SA_{ist} + \delta' X_{ist} + \gamma' State_{st} + \varepsilon_{ist}$$

(3) $$\varepsilon_{ist} = \mu_s + \varphi_{ist}$$

Where BMI is student i's body mass index in state s at time t, which is later replaced by the probability of an individual is overweight or underweight when using a Probit model, SA represents either the number of years or number of times student i has been exposed to school accountability or tested in state s at time t and ε_{ist} is the error term.

In this study, if the introduction of accountability is exogenous and uncorrelated with any local policies or socioeconomic status, then Pooled Ordinary Least Square (POLS) regression of student BMI on school accountability in each state would produce an unbiased estimate.

However, if under (3) and μ_s is correlated with any state policies other than accountability which may be correlated with student BMI, either POLS or GLS (RE) estimates will be biased, which to a great extent, would be the case. For example, various physical education programs implemented across states over the years which may coincide with school accountability implementation. Between years 1991 and 2003, the percentage of high school students enrolled in daily PE classes declined by 13% from 41.6% to 28.4% (Gunbaum et al. 2004). Because physical activity level is believed to be associated with body weight (Heitmann et al.1997; Stone et al. 1998), state PE policy may bias SA in a POLS model. Second, as mentioned earlier, states with relatively worse performance on national tests have a higher possibility of adopting a school accountability system before 1999 which results in the endogeneity problem in the variable of SA, $E(\varepsilon_{ist} | SA_{ist}) \neq 0$. Nevertheless, if the correlation is restricted to the state-specific component, i.e. $E(\varepsilon_{ist} | SA_{ist}, X_{ist}, State_{st}, \mu_s) = E(\varepsilon_{ist})$, fixed effects estimation is the preferred procedure to eliminate any other bias from the policies that lead to up or down trends in student BMI in each state.

Fixed effects model estimates:

(4) $$BMI_{ist} = \beta_0 + \alpha \; SA_{ist} + \delta' X_{ist} + \gamma' State_{st} + \phi_s + \varepsilon_{ist}$$

The various dates of introduction of the accountability systems and testing grades over the periods across states during which student BMI are observed enable the fixed effects analysis. In equation 4, ϕ_s refers to the state fixed effects.

One potential problem with the fixed effects model is that it fails to control for the potential state BMI trend over time which causes endogeneity problem due to omitted variable. For instance, Mississippi has been the ‘fattest’ state in years 2005 to 2007, with the average percentage of obese adult population 31.6%, while the adult obesity rate is about 13% lower in Colorado. Between year 2007 and 2008, the obesity rate increases 0.8% from 32% to 32.8% in Mississippi while the obesity rate decreases by 0.2% from 18.7% to 18.5% (Overweight and obesity Trends by State 1985-2008, CDC

2008). Figure 2-3 further validates the existence of differential state trends by plotting densities of BMI for states Tennessee and Missouri in year 1999 and 2005 using data from YRBSS. As one can see that the kernel density is more peaked than the normal for Missouri while the kernel density is very right-skewed for Tennessee. From 1999 to 2005, the density of BMI for Missouri noticeably shifted to the right where the shift is much milder for Tennessee. The fixed effects model would fail to address this state-specific growth pattern and leads to an overestimation of the causal effect of SA. I thus use a fixed-trend model as specified in equation (5) to solve this potential problem to the most extent.

(5) $$BMI_{ist} = \beta_0 + \alpha \; SA_{ist} + \delta' X_{ist} + \gamma' State_{st} + \phi_s + \tau_t + \phi_s \times \tau_t + \varepsilon_{ist}$$

where τ_t represents time dummies incremented in years and $\phi_s \times \tau_t$ captures different state trend over time effects. The fixed-trend model can resolve the endogeneity concern since the adolescent BMI depends only on historical level of state BMI rather than its growth.

One drawback of including the full set of year dummies, however, is that the SA measured by LENGTH will not be accurately estimated. Particularly, LENGTH variable is constructed as the number of years a student has been exposed to the system, in which case the incremental change will be exactly two years in the following survey period for students from a same state progressing from one grade to the next. The multicollinearity between LENGTH measures and the year dummies made it very difficult to disentangle their separate effects. Therefore the only measure left is FREQUENCY. Furthermore, time trend picks up the effect of permanent factors in national and state shocks and leaves only transitory variation from its trend to be explained by changes in SA, the coefficient of FREQUENCY may be underestimated with larger standard errors in comparison to a fixed effects model. Alternatively, to control for the unobserved nationwide events not measured by the other independent variables, I could firstly estimate the state by year trend for each state and then include them into the main regression. To estimate the unbiased state BMI trend, I use BRFSS data for adults, aged younger than 35 years-old, and calculate BMI trend for each state which is then incorporated in the main accountability-weight status equations. I choose adults age 18-35, because adults BMI in this age group are less likely to be affected by medical or other social status factors that could not be observed and hence controlled by the regression. Also, this age group is most likely to resemble the BMI trend of students' for each state. I further divide this sample into two subgroups aged younger or older than 25 years to test if the state trends differ across age groups. The original fixed effects model becomes:

(6) $$BMI_{ist} = \alpha\ SA_{ist} + \delta' X_{ist} + \gamma' State_{st} + \beta_1 Year_t * Trend_s(age < 25) + \phi_s + \varepsilon_{ist}$$

(7) $$BMI_{ist} = \alpha\ SA_{ist} + \delta' X_{ist} + \gamma' State_{st} + \beta_1 Year_t * Trend_s(25 \le age \le 30) + \phi_s + \varepsilon_{ist}$$

Where $Trend_s$ is obtained from:

(8) $$BMI^a{}_{ist} = \beta_0 + \delta' X^a{}_{ist} + \gamma' State_{st} + Trend_s \phi_s + \tau_t + \varepsilon_{ist}$$

where BMI^a is the outcome variable for adult *i* residing in state *s* during year *t*, X^a is a set of adult characteristics and *State* is a set of state characteristics which include the same controls as *X* and *State* in equations (1)-(7).

So far, I assume that school accountability has linear effects on student weight status once controlling for individual characteristics and various state policy characteristics, however, it might be the case that students' weights respond to school accountability systems in a non-linear fashion, i.e. $E(BMI_{ist} | SA_{ist}, \varepsilon_{ist}) \neq \alpha SA_{ist} + \varepsilon_{ist}$ (assuming $E(SA_{ist} | X_{ist})$ is linear in function of *X*). If this is the case, the linearity assumption is violated and hence jeopardizes the analysis. Additionally, students' weight status may not change dramatically in a relatively short period, for instance one or two semesters, rather an adjustment period for students to respond to the policy reform may exist. One might also be worried that state level programs implemented earlier or historical state socioeconomic status whose effects become apparent at the same time as the accountability policy was instituted. I thus employed an additional method by creating a set of dummy variables to signify separate length ranges for school accountability, namely: Exposed to School Accountability 1-2 years, Exposed to School Accountability 3-4 years, etc. Specifically, I then divided the school accountability LENGTH variable into six two-year intervals, which created five dummy variables (using students had not been or had been exposed to the system for less than one year as baseline), to test for nonlinearity in the effects.

Similarly, to test the differential impacts of the various intensity of SA on students' weight status, I use the same method separating FREQUENCY_grade into 5 mutually exclusive groups where the higher number of times a student had been tested represents a more stringent accountability system.

Last, I run separate regressions using the probability of being overweight and the probability of being underweight as dependent variables for each proposed equation, though I use BMI as the outcome variable in every equation showed above. Marginal effects are presented in the next section. Various probit regressions are estimated:

(9) $\Pr(Y_{ist} = 1) = \Phi[\alpha \; SA_{ist} + \delta' X_{ist} + \gamma' State_{st} + \beta_1 StateTrendMeasure + \phi_s + \varepsilon_{ist}]$

where the dependent variable is a dummy variable which equals one if student *i* is overweight or underweight in state *s* during year *t*. *StateTrendMeasure* is substituted with several measures described above in different models.

Results

Do school accountability systems increase adolescent BMI?

A Difference-in-Differences Design

In this section, I present the DID results based on the model discussed in Section III. The means of the data are presented in Table 2-3 for the accountability states and control states (both for the "before policy" year and "after policy" year). There are not many striking differences across groups except that non-accountability states have a lower percentage of African-American students and a higher percentage of White students than the accountability states. One reason is that African-American students usually have worse academic performances relative to other racial groups and school accountability systems, in turn, are more likely to be implemented in states where academic performance is poor. Hence, states with higher percentages of African-American students may adopt school accountability systems in an earlier year to promote students academic performances. Another difference between accountability and non-accountability states worth mentioning is that most states implemented school accountability system in grades 3 to 11, therefore the post policy percentages of eleventh graders who have not been exposed to the systems in the non-accountability states decreased significantly lower to around 7% compared to their accountability counterparts, around 25%.

Table 4 illustrates DID estimation of the short run effect of school accountability on student BMI. The panel compares the change in BMI for students in the states that implemented the policies to the change for students in the control states. Each cell contains the mean average student BMI, along with the standard error and the number of observations. There is a 0.17 kg/m^2 decrease in the BMI of students in the non-accountability states over this period, compared to a 0.2 kg/m^2 increase in the BMI of students in other states. Thus there was a total 0.37 kg/m^2 relative increase in the BMI of states implemented the policies. However, the difference is not statistically significant.

As the summary statistics revealed, there is an imbalanced distribution of racial groups across accountability and non-accountability states; hence a regression framework will allow me to control for these observables that affect the outcome variables of interest.

The third row of Table 2-5 shows the estimates of the interaction term from equation (1) and for brevity, some of the regression output has been omitted from the presentation. The coefficients indicate that student BMI increased significantly by 0.79 kg/m^2 for the treatment group with the average 2.4 years treatment. Female students' BMI increased by 1.17 kg/m^2 for the treatment group while there is no statistically significant increase of BMI for male students due to the treatment.

POLS and FE models

Starting with the specification using continuous variables, I tested the effects of school accountability on student BMI, along with variables that control for individual characteristics and state characteristics. The first column of Table 2-6 to Table 2-8 report the estimated effects of school accountability on student BMI and the probability of being overweight and underweight using the POLS of equation (2) where top panel used LENGTH measures and bottom panel used FREQUENCY measure. They are provided here primarily as a benchmark to gauge the possible bias in comparison to the more reliable state fixed effects models discussed below. First, note that both the LENGTH and FREQUENCY variables have positive effects on BMI and the probability of being overweight while the LENGTH_general which measures the overall school accountability effects presents slightly higher effects implying that not only students in the tested graded were affected by the school accountability systems. On the other hand, I do not observe any statistically significant effects of school accountability on the probability of being underweight. As one would suspect, POLS is not the ideal model under the circumstances where there may be omitted state characteristics that also affect BMI and the prevalence of obesity, or when the strict exogeneity assumption is violated. Therefore, columns (2) and (3) in Table 2-6 to 2-8 present the fixed effects model of the relationship between student weight status and school accountability from equation (4) without or with state covariates, respectively. All the coefficients of interest across model specification are positive and statistically significant at the 1% level except for the effects on the probability of being underweight. The sizes of the coefficients, though estimated very precisely, also increase compared to POLS. The inclusion of state covariates as showed in Model (3) address the potential problem that unobservable conditions of state economic status might give rise to the increase of BMI. The coefficients decrease slightly but are not statistically different from Model (2) which assures me that the effects of SA are not driven by unobserved state time-varying effects. Evaluated at sample means, an additional year of exposure to the school accountability system or an additional test is associated with a 0.06 kg/m^2 increase in the BMI from 22.93 to 22.99 kg/m^2 and a 0.5 percentage point increase in the percentage overweight from 14.2 to 14.7%. The impact of accountability on the probability of being underweight is only statistically significant when using LENGTH_general and the probability of a student is underweight decreases by 0.2% when she/he exposed to the system for an additional year implying that if SA is related with the current eating disorder phenomenon, I do not observe it here.

Although these coefficient estimates present considerable and statistically significant effects of SA on adolescent obesity, one might suspect they are still somewhat overestimated due to the potential endogeneity problem elaborated earlier. So in column (6) in Tables 2-6 to 2-8, I show results from the fixed-trend specification adopted here to deal with the endogeneity problem due to omitted state BMI growth. As expected, the standard errors in all regressions increase to a great extent while the magnitudes for FREQUENCY measure did not change significantly from model (3). To further test the possibility of endogeneity bias associated with LENGTH measure, I next present results estimated using equation (7)-(9) in column (4) and (5) in Tables 2-6 to 2-8. I first use data from BRFSS for adults aged between 18 and 24 to estimate BMI trend for each state which is then incorporated in equation (7). Similarly, I again estimate BMI trend using BRFSS for adults aged between 25 and 35 and then include them into equation (8). The magnitude for the estimated effects of LENGTH and FREQUENCY only drop slightly relative to the estimates in fixed effects model implying that the endogeneity bias due to omitted state BMI trend does not exert practically important impact on my main results. The standard errors change little, and so do the estimates for the LENGTH and FREQUENCY measures when estimating the probability of being overweight. It seems reasonable, at this point, to conclude that students under NCLB tend to have higher BMI level and are more likely to be overweight, increasing BMI by

0.6 kg/m^2 and overweight rate by 0.5%. However, I do not observe any statistically significant effects of school accountability on students' probability of being underweight.

As discussed in the previous sections, early implementers may experience bigger effects on BMI and probability of being overweight compared to late implementers over time due to the potential non-linear growing effects over time. Apart from that, school accountability implemented before 1999 is significantly different from those adopted after 1999 and 2001 (NCLB) in either measurement of school performances based on student test scores, or rewards and sanctions for high-performing and low-performing schools. Furthermore, there might be a lag between school accountability implementation and the full response on student BMI and the probability of becoming overweight due to physical adjustment time. Using a set of dummy variables, I test the hypothesis stated above. Logically, for the linearity assumption to hold a monotonic relationship with invariant growth rate rather than an increasing growth rate is expected, i.e. the incremental across LENGTH/FREQUENCY categories should be relatively stable. The first specification using fixed trend effects, which includes five levels of dummy variables representing the LENGTH of school accountability exposure, is reported in Table 2-9 columns (1) and (2). Relative to the omitted category, states without accountability laws, the impact of the system on student BMI is statistically significant and positive, and the longer the length, the larger the coefficients. On average, there is around a 0.1 kg/m^2 increase in BMI and a 1% increase in the probability of being overweight for every two-year increase in the length of school accountability, suggesting a linear growth rate of BMI and overweight rate under SA. I then test FREQUENCY measures using the same model as displayed in columns 3 and

4. Coefficients across categories increase slightly but the growth rate of BMI is smaller than measured by LENGTH and present higher effects when a student tested less than 4 times under accountability system. With the increase in the length/frequency of exposure to the systems, I observe a decreasing marginal effect. However, the difference among growth rates across categories is not statistically significant. Similarly as measured by LENGTH, there is a 1% increase in the probability of being overweight for every two times increase in the frequency of school accountability. Results presented in Table 2-9 indicate that SA has slightly stronger short run effects on adolescent obesity and the marginal effect decreases gradually. Overall, however, linearity assumption holds.

To further investigate the differential impact of accountability across gender, ethnicity, and grade, I ran state fixed effects, fixed-trend effects and state trends regressions using various subsamples. Results from different specifications do not vary significantly, therefore I only present estimates obtained from fixed effects models and fixed effects with state*year trends (calculated using adult aged younger than 24) models. Table 2-10 presents estimates using a subsample of female and male students respectively. Consistent with results using the whole sample, FREQUENCY measure has slightly larger effects on both BMI and the probability of being overweight for female. It is also important to point out that, the magnitude of the estimates on female BMI are slightly higher, while the probability of being overweight due to SA is higher for male students when measured by LENGTH. I also observe differential impacts of SA on student BMI across racial groups as shown in Table 2-11. Note that school accountability has the strongest effect on BMI for Asian students, followed by White students and African American students. Hispanic students' BMI do not respond to the SA policy, though the probability of being overweight is highest for these students compared to all other racial groups under accountability systems. Last, using a subsample of students from grades 9 to 12, I adopted the same regression analyses to test if accountability was only bound to the affected grade. Since school accountability is mostly implemented in grades 10 and 11 in public high schools[15] and many states integrate high school exit exam into their school accountability systems, therefore students in these grades are more likely to be affected. In addition, since SA presents substantial lag effects on student BMI, I expect to see higher effects on 11th and 12th graders rather on 9th and 10th graders. Consistent with my expectation, as Table 2-11 shows, BMI is significantly higher for students from grade 11 to 12, yet the probabilities of being overweight are only significantly higher for 12th graders.

Do school accountability systems reduce adolescent activity levels?

With the increased strength of accountability, it might be the case that students substitute their physical activity time with studying time to avoid poor test outcomes. Hence, I tested the effects of the SA[16] on the number of days students participated in physical activities using fixed effects models. In Table 2-12 each panel represents a different version of the dependent variable: the number of PE classes in the last seven days, or number of days participated in exercise that is longer than twenty minutes that made the respondent sweat in the last seven days (vigorous exercise), or number of days participated in exercise longer than thirty minutes that did not make the

[15] 43 states do not implement school accountability systems in 9th grade.

[16] Both LENGTH and FREQUENCY present similar effects on students' physical activities, we only show results using LENGTH measure here.

respondent sweat in the last seven days (moderate exercise). With fixed-trend effects in place, the variable of interest is only negative and statistically significant for the female students' participation in PE classes, which suggests that for each additional year of school accountability the number of days participated in PE classes decreased 0.032 day for female students. As can be seen in Table 2-12, students who are under the system one year longer are slightly less physically active in both vigorous and moderate exercise, though none of these differences are statistically significant. The findings are important for policy makers since school accountability policies are designed to improve academic performance without harming students' physical health. Nevertheless, the results presented above suggest that with the implementation of accountability across states, students spend less time on physical exercise. Especially, female students cut back on PE class time.

Falsification tests

Testing school accountability system effects using adult BMI (ages 18 to 30)

I have argued that accountability systems result in higher student BMI, to further support this hypothesis I use adult BMI as dependent variable in a similar DID regression framework. One important distinction between adults and students is that adults are not directly affected by the school systems. Thus, the specifications with adult BMI as dependent variable should not produce statistically significant coefficients on accountability variables. With these hypotheses in mind, I again use adult (ages 18-30) BMI data obtained from BRFSS to test the effects of SA on adults. Although more parameterized models are preferred in the SA and adolescent obesity estimation, using the same LENGTH measure on adult obesity seems unreasonable. Especially, I will not be able to calculate the exact number of years an adult has been treated and hence make the comparison inaccurate. Instead, employing the same DID strategy discussed above I use data in years 1999 and 2003 and classify states into treatment and control groups based on the implementation date of school accountability. For the control variables, I include same individual and state covariates as in the YRBSS regressions. Table 2-13 contains the results of the estimated effect of school accountability on adult BMI. Different columns represent different regressions restricted by age. In terms of BMI, school accountability systems fail to show statistically significant effect on adults' BMI across all age groups. In other words, adult BMI, ages between 18 and 30[17], are not significantly affected by the implementation of school accountability which further supports the original hypothesis.

The Effects of School Accountability System on Adolescent Height

I examine next whether increases in the implementation of school accountability systems have similar effects on adolescents' heights. To gauge the credibility of these conclusions, this placebo test uses an outcome that cannot possibly be affected by the accountability systems. Using the preferred fixed-trend model, I perform this falsification test with the whole sample and with a series of subsamples. As showed in Table 2-14, I do not observe any statistically significant effects of school accountability systems on adolescents' height across models, implying that my conclusion of the effects of school accountability on student weight status is legitimate.

Specification test: identifying the effects of student accountability

"High-stake" testing systems were firstly implemented at the student level after unfavorable international comparisons of mathematics and science achievement among

[17] We also run several regressions with various age cutoffs, for instance 18-24 and 25-34. Results are not sensitive to how we group adults by age (not shown here).

U.S. students with that of students in other industrialized nations. Since then, educators and policy makers have been working to bring value to the high school diploma by raising the rigor of high school standards, assessments, and curriculum. For example, if students cannot pass the test at a certain grade by meeting a pre-calculated national average, they run the risk of being held back a grade level or not being allowed to graduate. Not until the 1990s was this system applied to the school level. Therefore, one might suspect that student accountability systems may also contribute to the increase of student BMI, since ultimately students are directly accountable for their grades and academic performances, and hence school accountability will be upward biased. In this section, I test the effects of student accountability on student BMI. The student accountability reform movement began in 1970s and one of the main parts of the policy change in this movement is state requirement of a high school exit exam. As shown in figure 2-6, before 1990, 10 states required the HS exit exam, and the number increased to 24 by 2005 (National Center on Educational Outcomes surveys; Guy et al 2002). Now more states are planning to implement this policy in the near future to promote high school quality.

To identify the potential effect of student accountability, this paper uses the date of introduction of high school exit examinations as the main independent variable. As student accountability systems are bound to students in grade eleventh or twelfth, I might not observe a cumulative effect on student BMI. Therefore, rather than using the same measure as school accountability, I generate a dummy variable coded as one if the state implemented high school exit exam before or in the survey year and zero otherwise.

Using similar analysis strategies as in the previous sections, I start with the POLS and later focus on fixed effects and fixed-trend effects regressions. The results presented in Table 2-16 indicate that even after including student accountability into the models, school accountability still shows significant and positive effects on student BMI and the probability of being overweight consistently across specifications while the effects are slightly smaller once controlling for state fixed effects and state fixed-trend effects. The high school exit exam requirement also present statistically significant effect on both student BMI and the probability of being overweight indicating that both school and student accountability systems may unintentionally increase the prevalence of adolescent obesity. Results do not change significantly compared to the estimates obtained from models excluding student accountability variable which provides further support for my hypothesis.

Conclusion

Existing literature on adolescent obesity blames lack of physical activity and the rise of calorie intake, while the evidence presented in this paper highlights three conclusions that the previous literature has not explored. First, this paper contributes to the sparse literature on adolescent obesity by identifying school accountability system as a strong factor that may contribute to higher adolescent BMI and the probability of being overweight. I construct three different measures to evaluate state school accountability over the years between 1999 and 2005. According to my estimates, all measures present statistically significant and positive effects on student BMI and the probability of being overweight. Specifically, an additional year of exposure to school accountability systems will lead to an increase in BMI by around 0.06 kg/m^2 and this effect is bigger among females, and among Asian and White students, as compared to

Hispanic and African-American students. One more year exposure to school accountability also increases the probability of being overweight by around 0.5% and this effect is slightly bigger among males and Hispanic students. Second, my empirical results show that school accountability systems present significant lag effects on adolescent BMI and the probability of being overweight. It appears that students need an adjustment period before the full response to changes of school environmental factor. Last, school accountability significantly decreases the number of times female students participate in PE classes. Although school accountability fails to show any statistically significant negative effects on students' participation in vigorous and moderate exercises, it shows a negative sign on both vigorous and moderate exercises.

The reasons why young people in the nation experience a significant growth in obesity are complicated and studies in the area usually can only answer the question from one of the angles. This paper attempts to shed new light on the possible cause of the adolescent overweight problem from the perspective of the change in education accountability policies. One limitation of studies on causes of adolescent obesity is the low explanatory power. The main reason might be the undesirable dataset which fails to include parents' BMI and other environmental variables. Therefore, obtaining an updated and richer dataset will enable researchers to get a better look of this currently very urgent health problem in future studies.

Of course, administrators have the goals to improve math, reading, and science test scores, but in the meantime we should not ignore the possibility that high stake testing systems in turn affect the test takers' physical health. Findings of this study imply that while school and student accountability systems may promote students academic performance, this could occur at the expense of students’ physical health. It is time for accountability advocates to consider incorporating policies addressing solutions to combat adolescents’ obesity while improving academic performance contemporaneously.

Table 2-1. Summary statistics of all variables

Variables	*Mean (Standard deviation)*	*Definition*
Body Mass Index	22.927 (4.216)	Weight in kilograms divided by height in meters squared
Overweight students	0.15 (0.357)	Dichotomous variable that equals 1 if body mass index ≥85th percentile for children of the same gender and age
Female	0.493 (0.5)	Dichotomous variable that equals 1 if respondent is Female
Asian	0.031 (0.175)	Dichotomous variable that equals 1 if respondent is Asian
African American	0.135 (0.342)	Dichotomous variable that equals 1 if respondent is African American
Hispanic	0.104 (0.305)	Dichotomous variable that equals 1 if respondent is Hispanic
White	0.636 (0.481)	Dichotomous variable that equals 1 if respondent is White
Grade 9	0.287 (0.453)	Dichotomous variable that equals 1 if respondent is in Grade 9
Grade 10	0.26 (0.438)	Dichotomous variable that equals 1 if respondent is in Grade 10
Grade 11	0.234 (0.424)	Dichotomous variable that equals 1 if respondent is in Grade 11
Grade 12	0.218 (0.413)	Dichotomous variable that equals 1 if respondent is in Grade 12
# Eat Fruit Last 7 days	2.856 (1.545)	Number of times a respondent eats fruit during the past 7 days
# Eat Green salad Last 7 days	2.109 (1.192)	Number of times a respondent eats green salad during the past 7 days
# Eat Potatoes Last 7 days	2.071 (1.076)	Number of times a respondent eats potatoes during the past 7 days
# Eat Carrots Last 7 days	1.709 (1.024)	Number of times a respondent eats carrots during the past 7 days
# Exercise-no Sweat Last 7 Days	4.719 (2.495)	Number of times a respondent exercises or participates in physical activity for at least 30 minutes that did not made her/him sweat and breathe hard
# Exercise Sweat Last 7 Days	3.653 (2.512)	Number of times a respondent exercises or participates in physical activity for at least 20 minutes that made her/him sweat and breathe hard
# Days have PE Class	3.21 (2.204)	Number of times a respondent exercises or participates in physical education class

Table 2-1. Continued

# Hours Watching TV per day	3.02 (1.74)	Number of hours a respondent watches TV on an average school day
State Level Variables		
LENGTH_grade	4.519 (3.064)	Number of years that the respondent has been exposed to school accountability system(based on testing grades)
LENGTH_general	5.285 (3.359)	Number of years that the respondent has been exposed to school accountability system
FREQUENCY_grade	2.984 (2.366)	Number of times that the respondent has been tested to school accountability system
Length of High Sch. Exit Exam	8.05 (9.404)	Number of years that High school exit exam has been implemented in respondent's state of residence
%Adult with Bachelor Degree	25.679 (4.185)	Percentage of adults with bachelor degree in respondent's state of residence
% Individual below poverty	10.909 (3.027)	Percentage of individuals below poverty line in respondent's state of residence
%Pub.Sch. with AP course	72.9 (0.188)	Percentage of public schools provide AP courses in respondent's state of residence
No. of Obs.	53,499	

Notes: *Summary statistics and standard deviation (in parentheses) from authors' calculations from the 1999,2001,2003,2005 Youth Risk Behavior Surveillance System.

* Africa American and Hispanics are oversampled in the YRBSS, results above are weighted means non-weighted means are similar for all variables except for the proportion of blacks and Hispanics

Table 2-2. Participation map and date of introduction of school/student accountability system

States	YRBSS participation map from year 1999 to 2005				Year of accountability system adoption	Testing Grades	Year of HighSchool Exit Exam adoption
	1999	*2001*	*2003*	*2005*			
Alabama	√	√	√		1997	3-8/3-8 & 11	1983
Arizona	√	√	√	√	2000	3-8 & 10	2001
Arkansas			√		1999	4,6,8/3-8	
California	√	√	√	√	1999	2-8/2-11	2005
Colorado		√			2002	3-10	
Connecticut	√			√	1993	4,6,8	
Delaware			√	√	1998	3-8, 10/3-11	
Florida	√	√	√	√	1999	4,5,8,10/3-10	1977
Georgia	√	√	√	√	2000	3-8, 11/1-8 & 11	1995
Hawaii	√				2001	3–8, 10/3-8 & 11	1983
Idaho		√		√	2003	3–8, 10	2000
Illinois	√	√	√	√	2003	4,8,10/3-8 & 11	
Indiana		√	√	√	1995	3,6,8/4,5,7,9/3-10	2000
Iowa				√	2003	3-8 & 11	
Kansas			√	√	1995	3-8	
Kentucky				√	1995	3-8,10-11/3-8	
Louisiana	√		√		1999	4,8,/4,8,10,11	1991
Maine	√	√	√		1999	4,8,11	
Maryland			√	√	1999	3-8	1979
Massachusetts		√	√	√	1998	4,8,10/3,4,7,6/3,5,7,8/3-8	2003
Michigan	√	√	√	√	1998	4,5,7,8,11/3-8 & 10-11	
Minnesota				√	1996	3-8 & 10-11	2000
Mississippi	√	√			1994	3-8/3-12	1989
Missouri	√	√	√	√	1997	3-8 & 10-11	
Montana		√			1998	4,8,10	
Nevada		√		√	1996	10-12/3,5,10	1980

Table 2-2. Continued

New Jersey	√	√	√		2003	3-8 & 11	1994
New Mexico		√	√		2003	3-9 & 11/3-8 & 11	1991
New York	√	√	√	√	1998	3-8 & 10/3-8 & 10-12	1979
North Carolina	√	√		√	1993	3-8 & 10	1981
Ohio	√	√	√	√	2002	3-8 & 10	1994
Oklahoma		√		√	1996	3–8, 10	
Oregon		√		√	2000	3-8 & 10	
Pennsylvania	√		√	√	2002	5,8,11/3-8 & 11	
Rhode Island	√				1997	4,8,10/3-8, 11	
South Carolina	√		√	√	1999	3-8/3-8,10	1990
South Dakota		√	√		2003	3-8 & 11	
Tennessee	√	√		√	1996	3-8 & 11	1983
Texas	√	√	√	√	1993	3-8 & 10	1987
Utah			√	√	2003	3-8 & 10/10-12	
Vermont			√	√	1999	3-8/3-8,10	
Virginia	√		√	√	1998	3,5,8/3-8	2004
Washington		√		√	1998	4,7,10	
West Virginia		√		√	1997	3-8 & 10	
Wisconsin	√	√	√	√	1993	4,8,10	

Notes:* The participation map is obtained from CDC YRBSS.

* The introduction dates of school accountability systems are obtained from the Consortium for Policy Research in Education (CPRE) and testing grades are collected from the state department of education website.

* To comply with the NCLB 2005-06 deadline, many states introduced new tests or tests in additional grades during the past few school years. I therefore collected data about major changes made in the states testing grades since 2002 from State Achievement Profiles conducted by Center on Education Policy.

* The introduction date of high school exit exam updated by author based on National Center On Educational Outcomes (NCEO) surveys.

Table 2-3. Summary statistics for difference-in-differences analysis

	State School Accountability Policy			
	Non-accountability states		Accountability states	
Variables	Before Policy	After Policy	Before Policy	After Policy
BMI	22.565	22.392	22.839	23.043
	(5.675)	(4.763)	(7.844)	(4.774)
Female	0.34	0.495	0.499	0.492
	(0.474)	(0.500)	(0.500)	(0.5)
Asian	0.054	0.015	0.046	0.027
	(0.226)	(0.12)	(0.210)	(0.162)
African American	0.035	0.086	0.366	0.278
	(0.184)	(0.280)	(0.482)	(0.448)
Hispanic	0.298	0.255	0.102	0.179
	(0.458)	(0.437)	(0.303)	(0.383)
White	0.576	0.591	0.405	0.440
	(0.495)	(0.492)	(0.491)	(0.496)
Grade 9	0.251	0.263	0.206	0.247
	(0.434)	(0.441)	(0.404)	(0.431)
Grade 10	0.221	0.334	0.189	0.250
	(0.415)	(0.472)	(0.392)	(0.433)
Grade 11	0.280	0.066	0.244	0.251
	(0.45)	(0.248)	(0.430)	(0.434)
Grade 12	0.248	0.335	0.360	0.253
	(0.433)	(0.472)	(0.480)	(0.435)
No. of Obs.	435	821	4,273	6,330

Notes: * Summary statistics and standard deviations (in parentheses) from authors' calculations from the 1999,2003 Youth Risk Behavior Surveillance System.
* See text for definitions of accountability and non-accountability states and before and after policy.

Table 2-4. Weighted average students' BMI in states with school accountability versus states without school accountability

Students' BMI	State School Accountability Policy		Time differences for locations
	Before Policy	After Policy	
Non-accountability states	22.565 (0.272) [435]	22.392 (0.166) [821]	-0.173 (0. 319)
Accountability states	22.839 (0.121) [4273]	23.043 (0.070) [6330]	0.204 (0.140)
Location difference at a point in time:	0.274 (0.298)	0.651 (0.180)	
Difference-in-Differences	0.377 (0.348)		

Notes: * Summary statistics, robust standard errors are clustered on state level (in parentheses) and number of observations [in squre brackets] from author's calculations from the 1999 and 2003 Youth Risk Behavior Surveillance System.
* See text for definitions of accountability and non-accountability states and before and after policy.

Table 2-5. Effects of school accountability on students' BMI: Results from DID controlling for individual characteristics

Variables	Whole sample	Female	Male
Year 2003	-0.690 (0.403)	-1.117 (0.554)*	-0.361 (0.429)
Treated States	0.50 (0.255)	-0.161 (0.272)	0.224 (0.376)
Accountability Treatment Effects	0.794^ (0.341)	1.171^ (0.553)	0.482 (0.396)
R^2	0.05	0.05	0.038
Sample size	11,381	5,587	5,794

Notes: * The table lists regression coefficients and standard errors in parentheses.
* Standard errors are cluster-corrected by state.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^: p<.1
* Additional controls include student's gender, grade, age square, race and variables related to students' eating habit: number of times eating fruit/ green salad/potatoes/carrots in the last 7days; the percentage of adults with bachelor degree in each state; percentage of individuals below poverty in each state and the percentage of public schools with AP course in each state.

Table 2-6. Estimated effects of school accountability on student BMI

Panel A	Number of years exposed to school accountability (LENGTH)					
Specification	(1)	(2)	(3)	(4)	(5)	(6)
LENGTH_grade	0.045* (0.020)	0.075*** (0.015)	0.062** (0.021)	0.058** (0.022)	0.056* (0.022)	-
LENGTH_general	0.042* (0.018)	0.085*** (0.014)	0.093*** (0.026)	0.094** (0.023)	0.094** (0.024)	-
State fixed effects	No	√	√	√	√	√
Year fixed effects	No	No	No	No	No	√
State-Specific Trends (State by year fixed effects)	No	No	No	No	No	√
State Covariates	No	No	√	√	√	√
State - Year Trends (using adults BMI age<=24)	No	No	No	√	No	No
State-Year Trends (using adults BMI 24<age<=34)	No	No	No	No	√	No

Panel B	Number of times exposed to school accountability (FREQUENCY)					
Specification	(1)	(2)	(3)	(5)	(6)	(4)
FREQUENCY_grade	0.057^ (0.030)	0.096*** (0.022)	0.069* (0.028)	0.063* (0.029)	0.060* (0.030)	0.080^ (0.045)
Adjusted R^2	0.044	0.05	0.05	0.05	0.05	0.051
Sample size	50,717	50,148	50,148	50,148	50,148	50,148

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression.

Standard errors are cluster-corrected by state.

* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^: p<.1

* Additional controls include student's gender, grade, age square, race and variables related to students' eating habit: number of times eating fruit/ green salad/potatoes/carrots in the last 7days; the percentage of adults with bachelor degree in each state; percentage of individuals below poverty in each state and the percentage of public schools with AP course in each state.

Table 2-7. Estimated effects of school accountability on the probability of being overweight

Panel A	Number of years exposed to school accountability (LENGTH)					
Specification	(1)	(2)	(3)	(4)	(5)	(6)
LENGTH_grade	0.003^	0.005***	0.005**	0.005**	0.005**	-
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	
LENGTH_general	0.003*	0.005***	0.004^	0.006*	0.006**	-
	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	
State fixed effects	No	√	√	√	√	√
Year fixed effects	No	No	No	No	No	√
State-Specific Trends (State by year fixed effects)	No	No	No	No	No	√
State Covariates	No	No	√	√	√	√
State - Year Trends (using adults BMI age<=24)	No	No	No	√	No	No
State-Year Trends (using adults BMI 24<age<=34)	No	No	No	No	√	No

Panel B	Number of times exposed to school accountability (FREQUENCY)					
Specification	(1)	(2)	(3)	(4)	(5)	(6)
FREQUENCY_grade	0.004^	0.006***	0.005^	0.005*	0.005*	0.004
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Adjusted R^2	0.019	0.025	0.026	0.026	0.026	0.028
Sample size	50,715	50,146	50,146	50,146	50,146	50,146

Notes: * Each column of the table presents coefficients from linear probability models and standard errors in parentheses from a different regression. Standard errors are cluster-corrected by state.
* All models include additional controls listed in the notes to Table 5.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^: p<.1

Table 2-8. Estimated effects of school accountability on the probability of being underweight

Panel A	Number of years exposed to school accountability (LENGTH)					
Specification	(1)	(2)	(3)	(4)	(5)	(6)
LENGTH_grade	-0.0002 (0.0005)	-0.0009 (0.0006)	-0.0002 (0.001)	0.00004 (0.0009)	0.00006 (0.0009)	-
LENGTH_general	-0.0004 (0.0004)	-0.002** (0.0006)	-0.002 (0.001)	-0.002 (0.001)	-0.002 (0.001)	-
State fixed effects	No	√	√	√	√	√
Year fixed effects	No	No	No	No	No	√
State-Specific Trends (State by year fixed effects)	No	No	No	No	No	√
State Covariates	No	No	√	√	√	√
State - Year Trends (using adults BMI age<=24)	No	No	No	√	No	No
State-Year Trends (using adults BMI 24<age<=34)	No	No	No	No	√	No
Adjusted R^2	0.028	0.032	0.032	0.037	0.037	0.035
Sample size	50,715	50,146	50,146	50,146	50,146	50,146

Panel B	Number of times exposed to school accountability (FREQUENCY)					
Specification	(1)	(2)	(3)	(4)	(5)	(6)
FREQUENCY_grade	-0.0002 (0.0008)	-0.001 (0.001)	-0.0003 (0.002)	0.0002 (0.001)	0.00006 (0.0009)	0.002 (0.003)
Adjusted R^2	0.028	0.032	0.032	0.032	0.032	0.035
Sample size	50,715	50,146	50,146	50,146	50,146	50,146

Notes: * Each column of the table presents coefficients from linear probability models and standard errors in parentheses from a different regression. Standard errors are cluster-corrected by state.
* All models include additional controls listed in the notes to Table 5.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^: p<.1

Table 2-9. Estimated effects of SA using fixed-trend model with dummy variables

Model	LENGTH		FREQUENCY	
Dependent Variables	BMI	Pr(Overweight)	BMI	Pr(Overweight)
1-2 years/times	0.238*	0.025**	0.296**	0.027**
	(0.115)	(0.0093)	(0.108)	(0.009)
3-4 years/times	0.459**	0.038**	0.436**	0.039**
	(0.136)	(0.014)	(0.132)	(0.012)
5-6 years/times	0.303^	0.025^	0.495**	0.044***
	(0.161)	(0.013)	(0.177)	(0.013)
7-8 years/times	0.541**	0.043**	0.477*	0.051**
	(0.170)	(0.015)	(0.213)	(0.016)
9-10 years/times	0.624**	0.065**	-	-
	(0.237)	(0.023)		
R^2	0.051	0.026	0.05	0.026
Sample size	50,148	50,146	50,176	50,146

Notes: * Each column of the table presents coefficients and standard errors (cluster-corrected by state) in parentheses from a different regression.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^: p<.1
* All models include individual characteristics, state covariates, state fixed effects and state by year trends (calculated using adults BMI data age<=24).

Table 2-10. Estimated effects of school accountability by gender

Main Indep. Var.	LENGTH				FREQUENCY			
	Female		Male		Female		Male	
Dep. Var.	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)
(1)	0.074*	0.004*	0.048*	0.005*	0.090*	0.004	0.040	0.005^
	(0.028)	(0.002)	(0.025)	(0.002)	(0.040)	(0.003)	(0.032)	(0.003)
(2)	0.073*	0.004*	0.042	0.005*	0.090*	0.005	0.027	0.005
	(0.028)	(0.002)	(0.025)	(0.002)	(0.039)	(0.003)	(0.033)	(0.004)
Sample size	25672	25672	24476	24476	25671	25671	24476	24475

Notes: * Each column of the table presents coefficients and standard errors (cluster-corrected by state) in parentheses from a different regression.

* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^: p<.1
* Model (1) includes individual characteristics, state covariates and state fixed effects.
* Model (2) includes individual characteristics, state covariates, state fixed effects and state by year trends (calculated using adults BMI data age<=24).
* Additional controls include student's gender, grade, age square, race and variables related to students' eating habit: number of times eating fruit/green salad/potatoes/carrots in the last 7days; the percentage of adults with bachelor degree in each state; percentage of individuals below poverty in each state and the percentage of public schools with AP course in each state.
* Separate regressions are run for each sex group to allow for heterogeneous treatment effects.

Table 2-11. Estimated effects of school accountability by racial group/grade

Panel A	LENGTH							
	African American		Hispanic		White		Asian	
Model	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)
(1)	0.058	0.001	0.039	0.007*	0.070*	0.005*	0.168***	0.008*
	(0.046)	(0.004)	(0.037)	(0.004)	(0.026)	(0.002)	(0.043)	(0.005)
(2)	0.049	0.003	0.039	0.007*	0.058*	0.005*	0.168**	0.009^
	(0.048)	(0.004)	(0.028)	(0.003)	(0.024)	(0.002)	(0.045)	(0.005)
Sample size	11803	11786	10812	10774	21629	21629	1497	1379
	9th Grade		10th Grade		11th Grade		12th Grade	
Dependent Variables	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)
(1)	0.045	0.003	0.055*	0.003	0.072*	0.003	0.105***	0.009***
	(0.049)	(0.005)	(0.027)	(0.002)	(0.034)	(0.003)	(0.027)	(0.002)
(2)	0.026	0.003	0.051^	0.002	0.067*	0.003	0.112***	0.011***
	(0.061)	(0.005)	(0.027)	(0.002)	(0.035)	(0.003)	(0.024)	(0.002)
Sample size	11,910	11,907	12,480	12,472	12,853	12,841	12,903	12,903
Panel B	FREQUENCY							
	African American		Hispanic		White		Asian	
Model	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)
(1)	0.119^	0.005	0.019	0.008**	0.066*	0.003	0.211	0.009
	(0.070)	(0.005)	(0.023)	(0.002)	(0.032)	(0.003)	(0.089)*	(0.007)
(2)	0.118^	0.009	0.005	0.007*	0.046	0.002	0.212*	0.010
	(0.065)	(0.006)	(0.032)	(0.003)	(0.033)	(0.003)	(0.086)	(0.007)
Sample size	11803	11786	10812	10774	21629	21629	1497	1379
	9th Grade		10th Grade		11th Grade		12th Grade	
Dependent Variables	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)
(1)	0.039	0.003	0.052^	0.001	0.101	0.006	0.212***	0.014**
	(0.053)	(0.005)	(0.029)	(0.003)	(0.067)	(0.006)	(0.047)	(0.004)

Table 2-11. Continued

(2)	0.017	0.002	0.045	0.001	0.095	0.007	0.261***	0.019***
	(0.069)	(0.005)	(0.028)	(0.003)	(0.074)	(0.006)	(0.051)	(0.004)
Sample size	11,910	11,907	12,480	12,472	12,853	12,841	12,903	12,903

Notes: * Each column of the table presents coefficients and standard errors (cluster-corrected by state) in parentheses from a different regression.

* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^:p<.1
* Model (1) includes individual characteristics, state covariates and state fixed effects.
* Model (2) includes individual characteristics, state covariates, state fixed effects and state by year trends (calculated using adults BMI data age<=24).
* Additional controls include student's gender, grade, age square, race and variables related to students' eating habit: number of times eating fruit/green salad/potatoes/carrots in the last 7days; the percentage of adults with bachelor degree in each state; percentage of individuals below poverty in each state and the percentage of public schools with AP course in each state.
* Separate regressions are run for each ethnicity group/grade to allow for heterogeneous treatment effects.

Table 2-12. Estimated effect of school accountability on students' physical activities

Variables	PE Classes	PE Classes (Female only)	PE Classes (Male only)
School Accountability	-0.012	-0.032*	0.008
	(0.01)	(0.013)	(0.015)
R^2	0.18	0.223	0.14
Sample size	46,409	23,972	22,437
Variables	Vigorous Exercise	Vigorous Exercise (Female Only)	Vigorous Exercise (Male Only)
School Accountability	-0.009	-0.013	-0.002
	(0.012)	(0.016)	(0.017)
R^2	0.134	0.125	0.08
Sample size	50,056	25,635	24,421
Variables	Moderate Exercise	Moderate Exe. (Female Only)	Moderate Exe. (Male Only)
School Accountability	-0.003	-0.014	0.010
	(0.013)	(0.017)	(0.019)
R^2	0.06	0.066	0.058
Sample size	50,054	25,633	24,745

Notes: * The table lists FE regression coefficients and robust standard errors in parentheses.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^: p<.1
* Vigorous Exercise: exercise longer than 20 minutes that make you sweat
* Moderate Exercise: exercise longer than 30 minutes that did not make you sweat
* PE class: school PE classes
* Additional controls include student's gender, grade, age square, race and variables related to students' eating habit: number of times eating fruit/green salad/potatoes/carrots in the last 7days; the percentage of adults with bachelor degree in each state; percentage of individuals below poverty in each state and the percentage of public schools with AP course in each state.

Table 2-13. Estimated effects of school accountability on adults' BMI: results from DID controlling for individual characteristics

Variables	Age 18-19	Age 20-24	Age 25-30
Year 2003	0.437	0.487	0.361
	(0.161)	(0.132)	(0.081)
Treated States	0.116	-0.176	-0.186
	(0.237)	(0.157)	(0.204)
Accountability Treatment	-0.119	0.17	0.159
Effects	(0.221)	(0.163)	(0.16)
R^2	0.046	0.054	0.053
Sample size	6,377	12,304	15,244

Notes: * The table lists regression coefficients and standard errors in parentheses.
* Standard errors are cluster-corrected by state.
* Statistical significance level: ***: p<.001; **:p<.01; p<.05

Table 2-14. Falsification test: If student accountability systems affect student height

Model	Whole sample	Female	Male
Dependent Variables	Height	Height	Height
LENGTH_grade	-0.00002	0.0001	-0.0003
	(0.0006)	(0.0008)	(0.0009)
FREQUENCY_grade	0.0002	0.0001	-0.0003
	(0.0007)	(0.0009)	(0.001)

Notes: * Each column of the table presents coefficients and standard errors (cluster-corrected by state) in parentheses from a different regression.
* Additional controls are listed in the notes to Table 5.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05
* Models includes individual characteristics, state covariates, state fixed effects and state by year trends

Table 2-15. Falsification test: If student accountability contributes to student overweight

Model	(1)		(2)		(3)	
Dependent Variables	BMI	Pr(Overweight)	BMI	Pr(Overweight)	BMI	Pr(Overweight)
School	**0.049*** **	**0.004***	**0.041** **	**0.003***	**0.041** **	**0.003***
Accountability	**(0.007)**	**(0.002)**	**(0.015)**	**(0.001)**	**(0.015)**	**(0.001)**
Student	0.019	0.009	**0.068*** **	**0.003***	**0.071***	**0.003***
Accountability	(0.025)	(0.011)	**(0.011)**	**(0.001)**	**(0.03)**	**(0.001)**
R^2	0.051	0.02	0.051	0.03	0.05	0.026
Sample size	50,717	50,715	50,148	50,146	50,148	50,146

Notes: * Each column of the table presents coefficients and standard errors (cluster-corrected by state) in parentheses from a different regression.

* Additional controls are listed in the notes to Table 5.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05
* Model (1) includes individual characteristics and state covariates.
* Model (2) includes individual characteristics, state covariates and state fixed effects.
* Model (3) includes individual characteristics, state covariates, state fixed effects and state by year trends (calculated using adults BMI data age<=24).

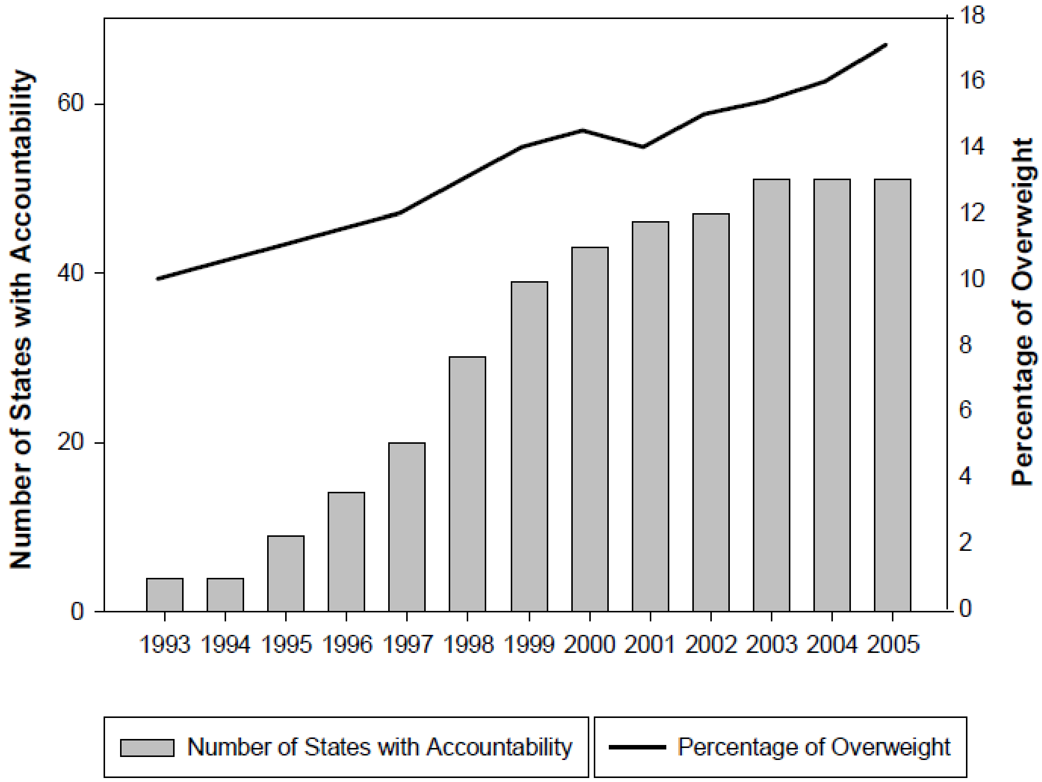


Figure 2-1 States with school accountability systems and rate of overweight, 1993-2005

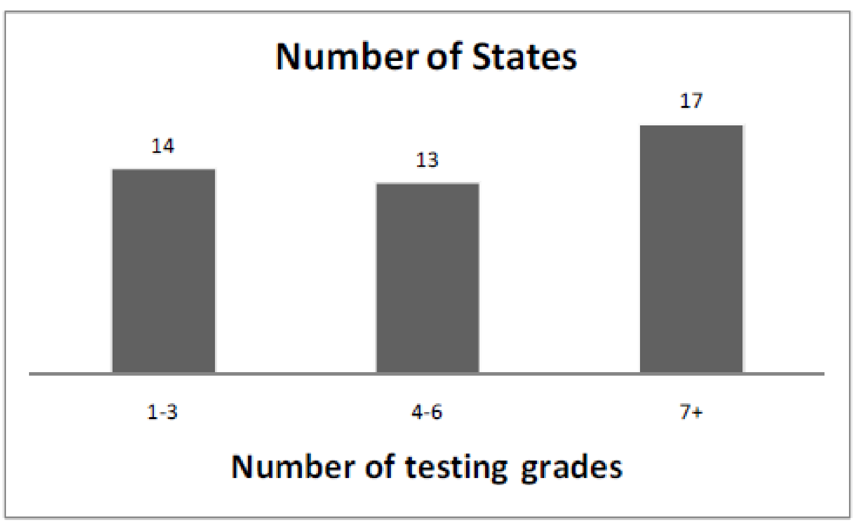


Figure 2-2 States with different testing grades when initiated the accountability systems

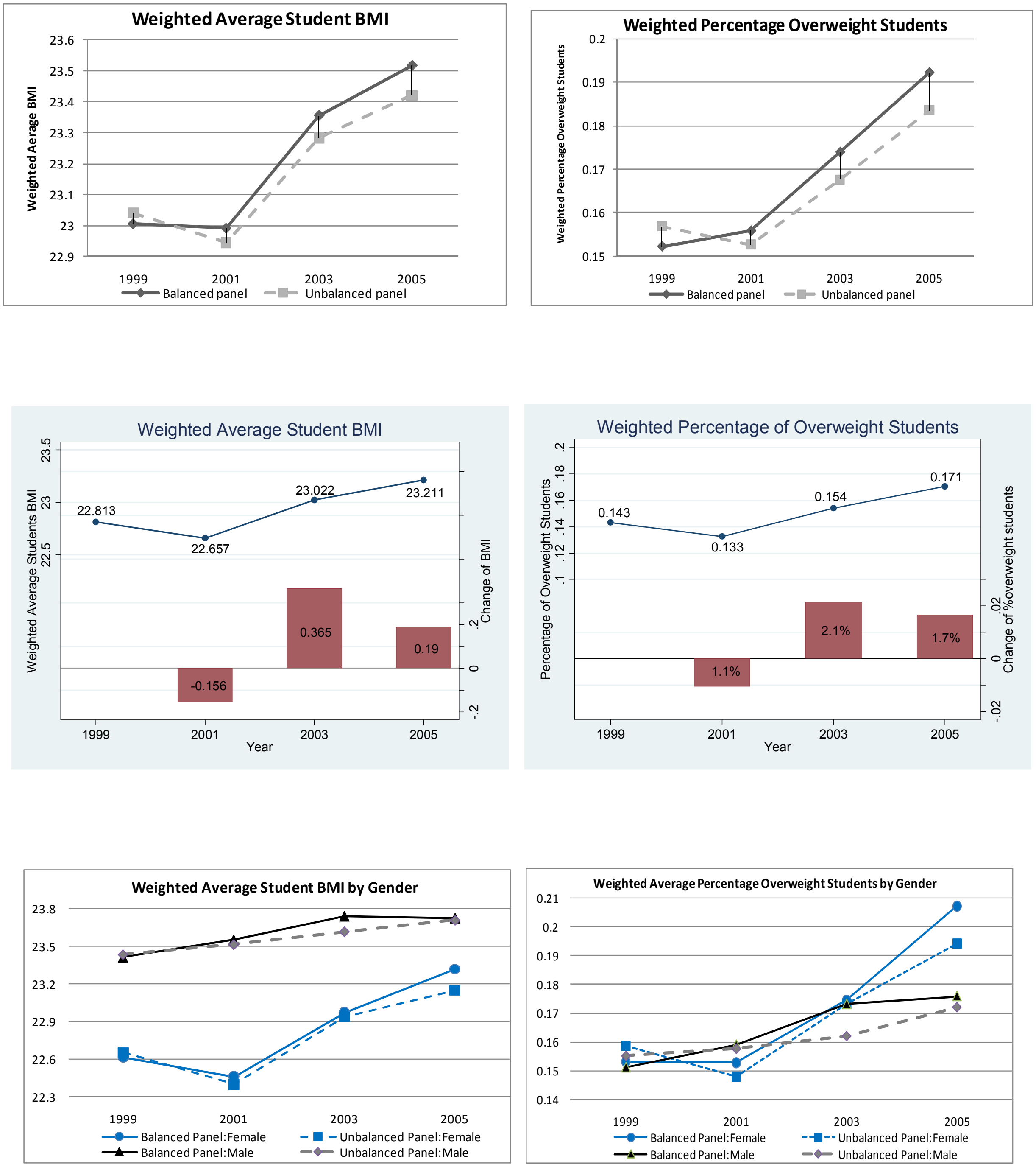


Figure 2-3. Trends in Body Mass Index and Percentage Overweight, 1999–2005

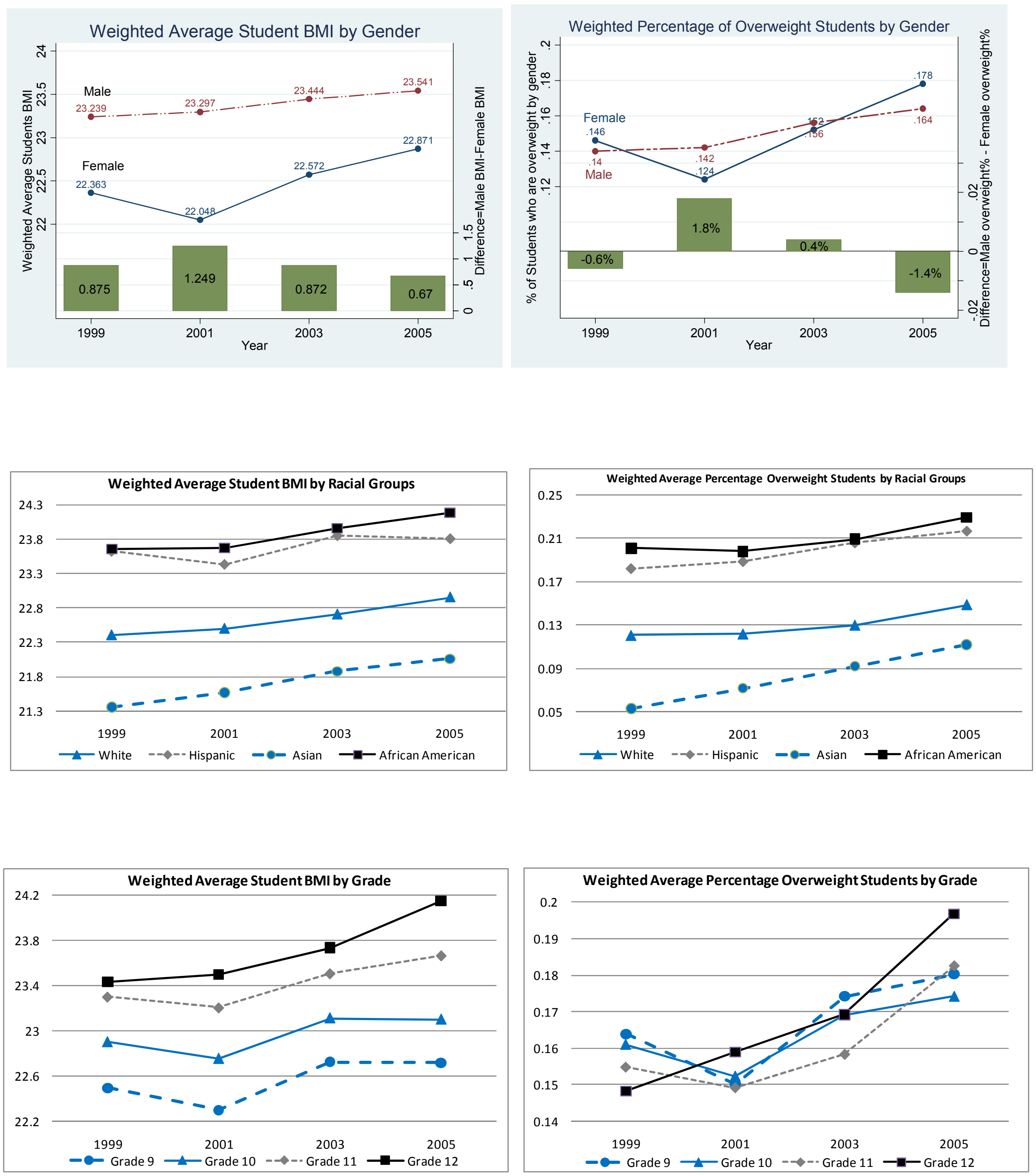
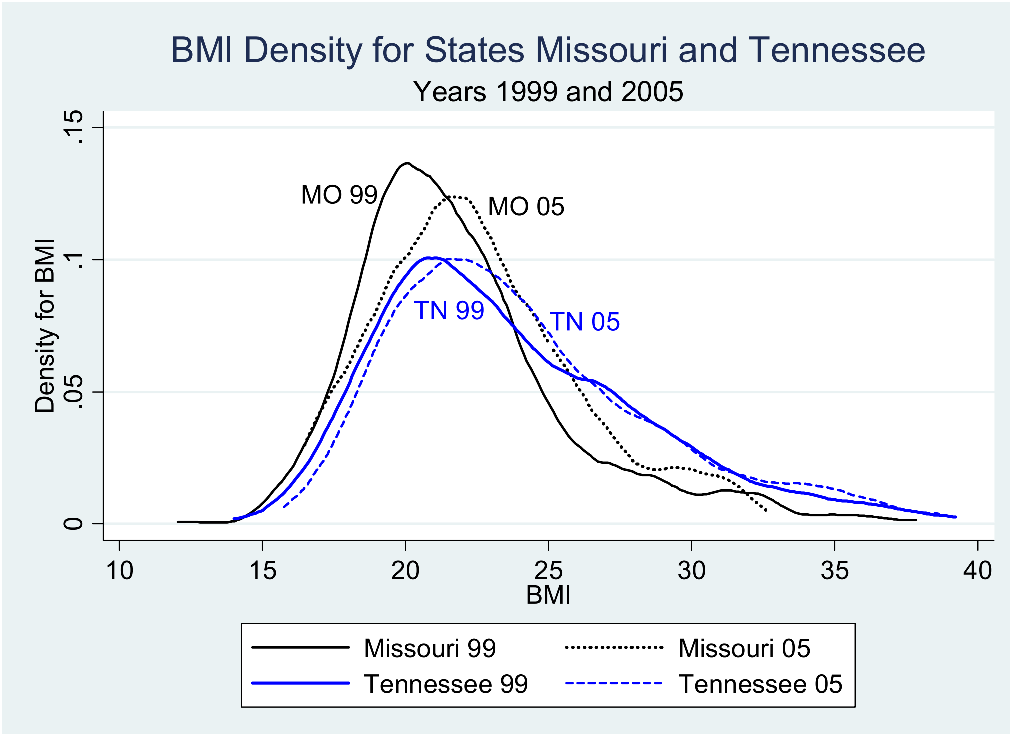


Figure 2-3. Continued



Note: Average BMI_{MO} 1999= 21.96 Average BMI_{TN} 1999= 23.74

Average BMI_{MO} 2005= 22.51 Average BMI_{TN} 2005= 24.15

Figure 2-4 BMI density for states Missouri and Tennessee, 1999 vs. 2005

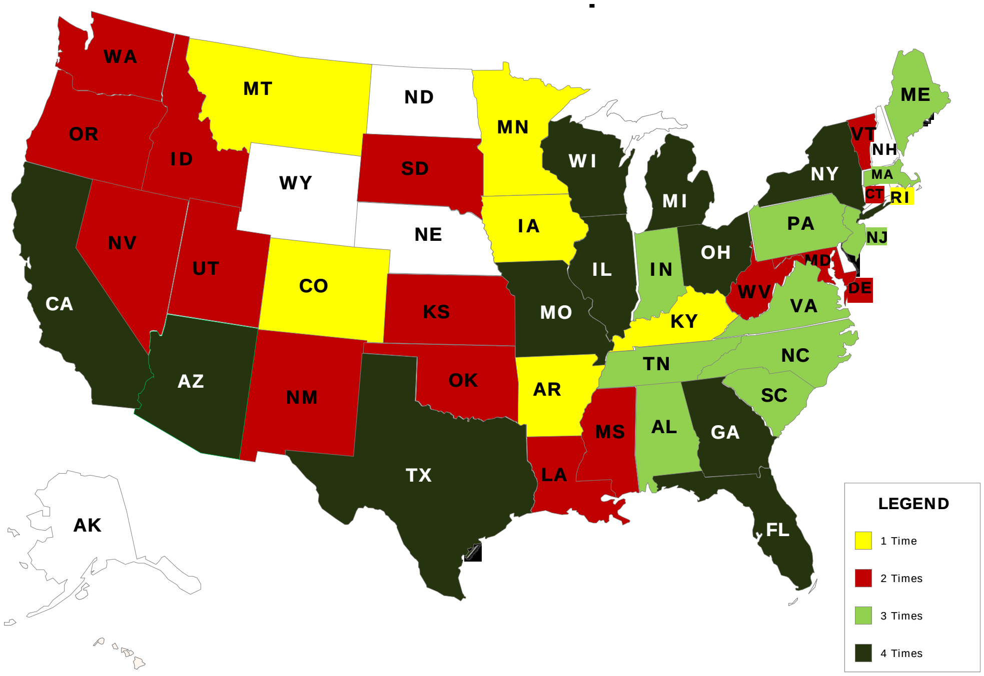


Figure 2-5 YRBSS Participation Map 1999-2005

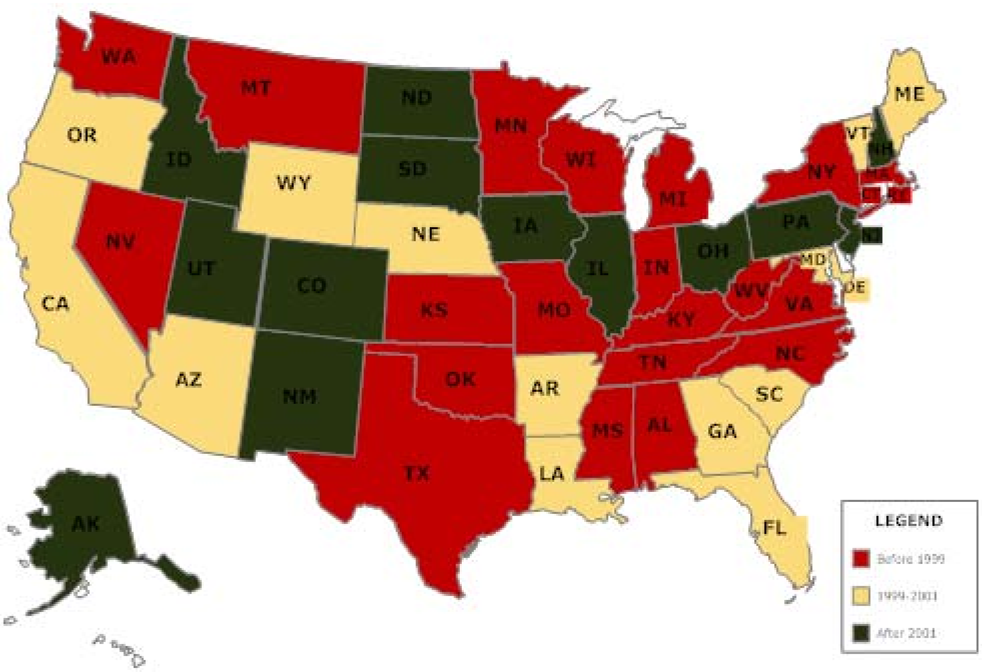


Figure 2-6. States with the Date of Introduction of School Accountability before 1999, Between 1999 and 2001 and After 2001 Shown in Different Legends

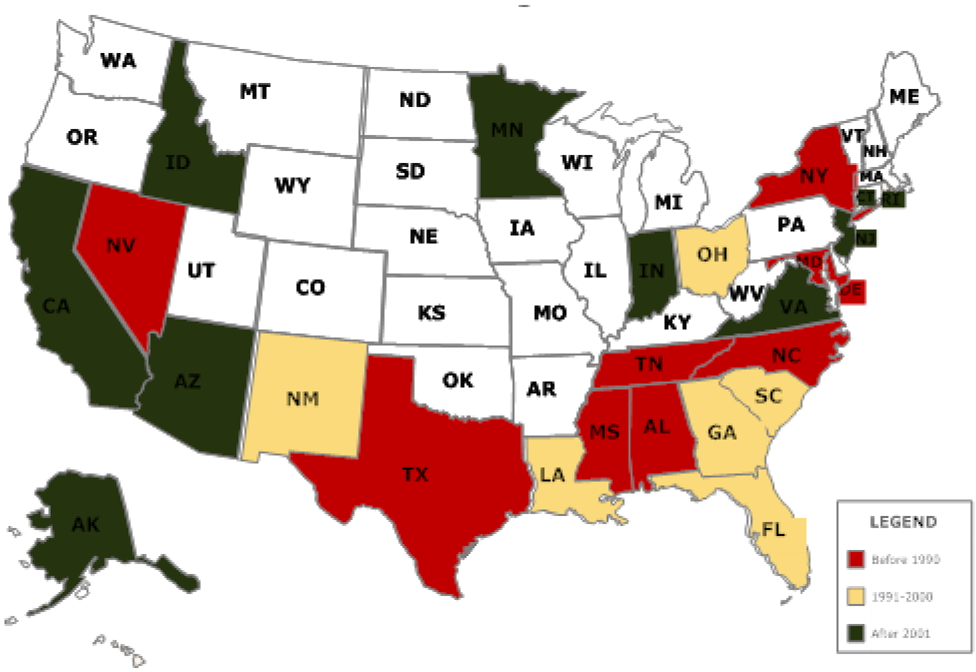


Figure 2-7 States with the Date of Introduction of High School Exit Exam Before 1999, Between 1999 and 2001 and After 2001 Shown in Different Legends.

CHAPTER 3
HIGHER STANDARDS MAY NOT ALWAYS BE BETTER: WILL HIGH EVALUATION STANDARDS CONTRIBUTE TO CHILDHOOD OBESITY

Introduction

High performance standards have been advocated by school administrators, teachers, parents, and even by students themselves in education for centuries. While economists have studied the theoretical (Becker and Rosen 1990; Betts 1998; Costrell 1994) and empirical (Figlio and Lucas 2004; Lillard and DeCicca 2001; Betts and Grogger 2000; Betts 1995) determination and impacts of educational standards on students' academic achievement, little is known about their unintended effects on student health outcomes. This lack of evidence is surprising considering the more than 20-year-long history of standard-based education reform, including the most recent reform plan of President Obama, uniformly raising academic standards in all states.

The existing literature on education standards presents mixed effects on student academic performances both theoretically and empirically. One stream of studies focuses on evaluating the impacts of differential institutional education standards on students' labor market outcomes. For instance, Bagues, Labini, and Zinovyeva (2008) find a significant negative correlation between departments' average grades and the labor market outcomes of their graduates using data from Italy, i.e. graduating from a high grading department leads to a higher unemployment probability and lower wages. Lillard and DeCicca examined the effects of graduation standards and find a positive impact on drop-out rates.

Another stream of studies instead is interested in grading standards at the level of schools or teachers. Figlio and Lucas evaluate the effects of teacher-level grading standards on student achievement and find that higher teacher grading standards tend to have large, positive impacts on student test score gains in mathematics and reading. Like Betts and Grogger (2000), they also find differential effects of grading standards, depending on student type. High-achieving students tend to benefit most from high standards in reading, while low-achieving students benefit from high standards only when their classmates are high-achievers.

Although the issue of educational standards has been widely discussed in the economics literature both from theoretical and empirical perspectives, there is so far no studies investigating the relationship between teachers' grading standards and students' health outcomes. Understanding and measuring the unintended effects of educational standards on children's health are important, because teachers, schools and states under standard based policies are only held accountable for test scores, but not for other student outcomes, such as children's health, schools facing pressure to improve students' academic performance may make decisions that may have unintended consequences. They are especially relevant in the current education reform debate because many of the policy decisions, for instance raising academic standards, adopting a national common standard and linking teacher pay to student achievement are all predicated upon the assumption of large positive effects of standard-based education policies. However, two recent studies by Anderson, Butcher and Schanzenbach (2009) and Yin (2009) find strong evidence that school accountability system, especially the No Child Left Behind Act, present significant negative effects on children's health, especially obesity. Anderson, Butcher and Schanzenbach (2009) use school level data from Arkansas and find that schools that were below the average yearly progress (AYP) threshold under NCLB in year t-1 have a small, but statistically significantly higher rate of overweight and obesity in year t. Yin (2009) examine the effects of school accountability movement on adolescent obesity, specifically age between 15 and 18, and find that school accountability policies lead to a higher obesity probability and higher BMI.

This paper, therefore, is the first to fill the gap in the existing literature by investigating how teacher-level grading standards may affect children's health, specifically childhood obesity. Between the 1970s and 2000s, the fraction of overweight children increased dramatically from 5 percent to 25 percent. Previous studies in childhood obesity literature focuses on genetic (Perusse et al. 2001; Katzmarzyk et al.1999) and behavioral factors (Stunkard et al. 1990, Dietz 1994) which did not provide much policy implications. Children spend on average 6 to 8 hours at school, it is critical to understand how the school environment may contribute to obesity. Cawley, Meyerhoefer, and Newhouse (2006) explore the impact of state Physical Education (PE) requirements on youth physical activity and obesity. Although the state requirements do increase students' PE time, they find no evidence that PE lowers BMI or reduces the probability that a student is overweight. Schanzenbach (forthcoming) finds that children who consume school lunches are more likely to be obese than those who brown bag their lunches.

While the aforementioned papers provide important evidence of how school environment may contribute to childhood obesity, there are remaining unanswered questions in the childhood obesity literature. First, the existing literature does not examine how teacher-level factors may affect children's health, considering students spend the majority of time with their teachers at schools. An extensive empirical literature has documented a positive link between teacher expectations and student achievement (Gauthier 1982; Proctor 1984). As grading standards reflect teachers' expectations to the students which in turn may change students' future performance as extensively described elsewhere (Jussim, 1989; 1991). Teachers under the impression that higher grading standards may improve students' academic performance may undertake more stringent standards that have unintended consequences on children's health. High grading standards may cause low self-esteem which has been shown in the obesity literature as one of the main causes of childhood obesity. Second, aforementioned studies investigating the impact of school accountability policies on childhood obesity either uses school level data or focuses on adolescent obesity, the current paper instead uses individual level data that tracks children from kindergarten to grade eight.

Specifically, we expand on the existing empirical work in educational standards and childhood obesity literatures in two important ways. First, we analyze the unintended effects of grading standards at the teacher level on childhood obesity. Second, using a large scale longitudinal dataset allows us to control for the potential endogeneity problem due to simultaneity by dividing direct students assessment test scores and teacher evaluation scores into ten deciles respectively and then categorizing teachers into high, median and low standards based on teachers' evaluation scores for students performed equally in direct student assessment tests. Conditional on this extensive set of controls, we find that relative to low grading standards, students with median or high grading standards teachers tend to have higher BMI by 0.07 kg/m^2 and are 12 percentage points more likely to become overweight. We then stratify the sample by gender, race/ethnicity and high/low achievers to analyze whether higher standards have adverse consequences for minorities. The results show that female, White and low achieving students present stronger response to teachers grading standards compared to their counterparts.

A straightforward policy implication follows from the above results. Teachers, schools and policy makers should be very cautious about raising evaluation standards. If, as shown in this paper, grading standards vary across teachers, rewarding teachers with high grading standards may lead to undesirable consequences.

The rest of the paper is organized as follows. Section 2 describes the data and the main variables use in the empirical part. Section 3 presents the empirical analysis method. Section 4 summarizes the main results and section 5 concludes and discusses policy implications.

Data

We investigate the potential casual relation between teacher grading standards and childhood obesity using Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) data, a very detailed nationally representative and individual-based study of children's early school experiences beginning with kindergarten through eighth grade, developed under the sponsorship of the U.S. Department of Education, National Center for Education Statistics (NCES). A total of 21,260 children throughout the country participated in the base year data collection in the fall of 1998 and spring of 1999. Five more waves of data were collected beyond kindergarten: fall and spring first grade, and spring third, fifth, and eighth grades. All data collection was completed in the spring of 2007 when most of the children were in eighth grade[1]. The current study used data from the weighted sample at baseline of spring 1999 and 4 other waves collected in springs of 2000, 2002, 2004, and 2007. Data were collected from parents, teachers, and schools to provide important contextual information about the environment for the sampled children which allowed the author to control for children, family and school characteristics that might contribute to childhood obesity. We exclude observations with missing teacher grading standard information for the purpose of the hypothesis testing. A weight is applied to each record to adjust for non-response and the distribution of students by grade, sex, and race/ethnicity in each state.

The design of the ECLS-K emphasizes the interrelationships between the child and family and the child and school and recognizes the importance of factors that represent the child's health status. Children sampled in the study are directly assessed in each round of data collection and also indirectly assessed by their teacher. The direct assessment that took approximately 50-70 minutes in the base year and 90 minutes in the following waves was designed to provide data on the developmental status of children in the United States at the start of their formal schooling through eighth grade. In all rounds, one-on-one direct child assessments were administered that included cognitive and physical components. The cognitive assessment scores include measures that can be compared across waves to study children's gains in reading and mathematics. We select Item Response Theory (IRT) Scale Scores as the indicator for children's academic achievements mainly because scores derived from the IRT model are based on all of the child's responses to a subject area assessment and can identify

[1] Sample also includes students who were either held back (e.g., seventh-graders) or promoted ahead an extra year or more (e.g., ninth graders) in 2007, and this sample frame applied to other survey years.

cross-sectional differences among subgroups[2]. Teachers were asked to complete a questionnaire for each of the sampled children in their classrooms. Specifically, teachers were asked to respond to 39 questions about the child's academic performance. Using this information, an academic rating scale (ARS) was developed to measure teachers' evaluations of students' academic achievement in reading and mathematics as the indirect cognitive assessment. The ARS Scores were rescaled to have a low of one and a high of five to correspond to the five-point rating scale that teachers used in rating children on these items.

Children's height and weight were recorded to measure their physical growth and development. A Shorr Board (for measuring height) and a digital bathroom scale were used to obtain the height and weight measurements, which were recorded on a height and weight recording form[3]. Children's BMI[4] was calculated based on height and weight. From these data, BMI percentiles were calculated using the 2000 Centers for Disease Control and Prevention growth charts. Obesity was defined as a BMI (kg/m^2) greater than or equal to the 95th percentile for age and gender, overweight as a BMI greater than or equal to the 85th percentile but less than the 95th percentile. According to the CDC growth charts in the United States, the 97th percentile of student BMI, ages between 4 and 16 years old, is below 28. The 5th percentile of student BMI, ages

[2] IRT uses the pattern of right, wrong, and omitted responses to the items actually administered in a test and the difficulty, discriminating ability, and "guess-ability" of each item to place each child on a continuous ability scale.

[3] Children's height and weight measurements were each taken twice to prevent error and provide an accurate reading.

[4] Composite Body Mass Index (BMI) was calculated by multiplying the composite weight in pounds by 703.0696261393 and dividing by the square of the child's composite height in inches.

between 4 and 16 years old, is above 17 (Centers for Disease Control and Prevention, 2000). Therefore, observations with BMI>30 as well as BMI<10 are excluded from the data set which may be subject to measurement errors. This reduced the sample size to 66,718. Between 1999 and 2007, BMI increased by 5.09 kg/m^2 or by 30% for all students while the average BMI is higher for male students and the percentages of overweight and increased over 10% for all students. According to the sample, approximately 25 percent children were overweight in 1999 and 35 percent in 2007 as shown in Table 3-1.

Additionally, to control for other individual characteristics that might have contributed to childhood obesity, we include age, age square, gender, and race into the analysis. In addition, we also include family, teacher and school characteristics to account for other observable demographics. For instance, ECLS-K calculated socioeconomic status (SES) at the household level using data for the set of parents who completed the parent interview in each survey year. The SES variable reflects the socioeconomic status of the household, which has shown to be linked with obesity, at the time of data collection and the components used to create the SES include father/male guardian's education, mother/female guardian's education, father/male guardian's occupation[5], mother/female guardian's occupation, and household income[6]. Thus, each component was converted to a z-score with mean of 0 and a standard

[5] Occupation was recoded to reflect the average of the 1989 General Social Survey (GSS) prestige score. This was computed as the average of the corresponding prestige scores for the 1980 Census occupational categories covered by the ECLS-K occupation.

[6] For income, the component i is the logarithm of the income for i-th household. The logarithm of income was used because the distribution of the logarithm of income is less skewed than the direct income values. Income was compared to Census poverty thresholds, which vary by household size. Households whose income fell below the appropriate threshold were classified as poor.

deviation of one. As described, the SES composite is the average of up to five measures, each of which was standardized to have a mean of 0 and a standard deviation of 1, hence the negative values. For analyses, we either include family characteristics or only include the continuous SES measure to avoid multicollinearity. To control for potential unobserved school factors that might affect children's weight status, we also include school type, percentage of students eligible for free/reduced lunch, and total school enrollment which was created using the school enrollment variable from the school administrator questionnaire. If this variable was missing, data for private schools were taken from the Private School Survey (PSS) and data for public schools were taken from the Common Core of Data (CCD) public school universe.

Table 3-1 depicts descriptive statistics for these key variables, including definitions, means, and standard errors. Except where noted, they are based on the sample of 66,718 that emerges when observations with missing values are deleted.

Empirical methods

An important concern in determining if teacher grading standards has an effect on childhood obesity is the fact that overweight children are more likely to be low performing students, and in turn more likely to receive lower ratings from their teachers (Datar and Sturm, 2006; Datar, Sturm, and Jennifer 2003). Therefore, if we construct grading standards directly based on teachers' ratings collected by ECLS-K, our estimates of grading standards might be biased. Instead, we first compare students' academic performance horizontally using direct student assessment scores and then define students with similar direct assessment scores but receiving a low rating from their teachers as high standard group (treatment group). This measure is different from the definitions used in previous literature (Figlio and Lucas 2004; Betts and Grogger

2000) where the standards are defined by aggregating teachers' grades across years to capture the time-invariant tendencies of high or low grading standards for the same teacher. The first reason why we define our standard differently is that the items and the metric for the eighth-grade teacher ratings are different from the Academic Rating Scale (ARS) ratings in earlier rounds of data collection, so the scores are not directly comparable to those for kindergarten, first, third, or fifth grades. Secondly, the average number of students rated by the same teacher across years is 2.5 and the total number of teachers sampled is around 18,176. Therefore, the average treatment effect of teacher grading standard is less likely to be biased by students' non-academic characteristics. Last, in the preliminary studies testing our standard measures, the standard deviation of common residuals is around 0.8, the standard deviation of individual teacher residual is around 0.5 and the fraction of variance due to common residual (ρ) is 0.683. Accordingly, we were able to reject the null hypothesis that all variance is due to within teacher variation.

Students in ECLS-K sample are given tests in reading and math in each survey year. Thus, we divide these direct assessment scores into deciles for each year and for reading and math, respectively. Teachers are asked to complete a survey which includes information about how they would rate the sampled students for both reading and math. We consider students fall into the same deciles perform equally in reading and math, and generate an indicator that equals to 2 if student received a lower teacher rating (high grading standard), 1 if student received a same teacher rating as their own score (median grading standard) and zero otherwise.

In Table 3-3, we present teachers characteristics by stratifying the sample into high, median, and low grading standards. One can see that teacher characteristics do not differ significantly except that teachers with regular certifications are more likely to perform high grading standards versus teachers with no teaching certifications. Student BMI is highest for high grading standard group and lowest for low standard group with approximately 4 kg/m^2 difference and overweight rate is 10% higher for high standard group relative to low standard group as shown in Table 3-4.

We begin our empirical work by estimating a simple pooled ordinary least squares (POLS) regression model, including an indicator for teacher grading standard and controls for other individual characteristics, family backgrounds and school characteristics. Grading standards (GS) are further measured by reading standards and math standards, respectively.

(1) $$BMI_{ist} = \alpha + \beta_1 GS_{ist} + \chi' X_{ist} + \delta' \, family_{it} + \phi' School_{st} + \varepsilon_{ist}$$

In this study, if teacher grading standard is exogeneous and uncorrelated with students' weight status, then POLS would produce an unbiased estimate. However, as mentioned earlier, simultaneity is the main concern for the current study. Although scholars in education have emphasized that teachers' contemporaneous perceptions of students' performance as well as their expectations for students' future performance are generally accurate (e.g., see Egan and Archer 1985; Good 1987; Hoge and Butcher 1984; Mitman 1985; Monk 1983; Pedulla, Airasian, and Madaus, 1980), one cannot guarantee that teachers grading standards will not be influenced by students' appearances or teachers' stereotypes will prevent them from forming inaccurate ratings for individual students.

To further circumvent this potential empirical problem, we utilize panel data from 1998 until 2007, five waves, so that we can control for unobserved, time-invariant individual fixed effects. Specifically, fixed effects model estimates:

(2) $$BMI_{ist} = \alpha + \beta_1 GS_{ist} + \chi' X_{ist} + \delta' \, family_{it} + \phi' School_{st} + \varphi_i + year_t + \varepsilon_{ist}$$

Equation 2 posits that the BMI, for the i^{th} student in school *s* during year *t*, is a function of teacher grading standard indicator, measured here by *GS* (2 if high standard, 1 if median standard, and 0 if low standard), and φ_i represents individual dummies and $year_t$ year dummies incremented in years.

As childhood obesity is defined as BMI greater or equal to 85th percentile for the same age and gender, we also test if teacher grading standard will affect children's probability of becoming overweight. As the probability of being overweight is a dichotomy, probit or logit model would be the appropriate. Therefore, we begin our empirical work by estimating a simple pooled logit model for the purpose of comparison to the more efficient estimators. Marginal effects are presented in the next section. Various logit regressions are estimated:

(3) $$\Pr(Y_{ist} = 1) = \Phi[\alpha + \beta_1 GS_{ist} + \chi' X_{ist} + \delta' \, family_{it} + \phi' School_{st} + \varepsilon_{ist}]$$

where the dependent variable is a dummy variable which equals to one if student i is overweight in school s during year t. GS is substituted with two measures described above in different models. If individual effect is invariant across time, pooled logit and probit will provide similar and consistent estimates. However, certain unobserved individual-specific characters may bias our estimates because teacher grading standard might not be exogenous. For instance, although not very likely teachers may give overweight students lower ratings due to certain stereotypes. If this is the case, a more sensible model is the fixed effects logit model. Chamberlain (1980) shows that a fixed effects logit model can be estimated by conditional maximum likelihood (ML) (conditioning on the fixed effects) consistently.

Results

Starting by examining some basic descriptive statistics on the final data set, one can see from Table 3-4 that students under high grading standards have significantly higher BMI compared to students from low grading standards and similarly with higher overweight rate. We also observe higher percentage African-American and Hispanic students in high standard group, implying that race of a particular student may cue the teacher to apply the generalized standards rather than developing specific standards tailored to individual students (Baron and Cooper 1985; Lightfoot, 1978). If this is the case, models without considering these factors will be biased as African American and Hispanic students tend to have higher BMI and are more likely to be overweight.

Moving across columns in Table 3-5, we present results showing the relationship between student BMI and teacher grading standards in the top panel and the probability of being overweight and grading standards in the bottom panel, both conditional on increasing number of controls. For the sake of brevity, the main tables in this paper present only the coefficients or marginal effects associated with the grading standards variables. Starting with simple POLS without any controls, we observe a statistically significant effect of high grading standards on student BMI while students under both higher math and reading standards show stronger effects in magnitude. Column 2 presents results including individual controls. We can see that all models remain statistically significant with slightly smaller magnitude regardless of which subject is used to evaluate standards. When adding family, school and teacher characteristics as well as school and year fixed effects, all models remain statistically significant. Comparing to students with low teacher grading standard, students' BMI are higher by 0.17 kg/m^2 and are 6 percentage points more likely to be overweight if having high math grading standards, 9 percentage points if having high reading grading standards and 3 percentage points if having both high math and reading grading standards.

There are all sorts of unobservable factors that predict the likelihood that a child is overweight. As a result, a naïve regression of grading standards on overweight may overstate the causal impact if all other related factors are not perfectly controlled. Although we have a relatively short panel, T=5, it is important to control individual unobservables. Column 6 in Table 3-5 shows estimates from individual and year fixed effects models. Again, we see a significant impact of grading standards on childhood obesity. With smaller magnitudes, high grading standards increase student BMI by 0.06 kg/m^2 and the probability of being overweight by 14 percentage points and 5 percentage points if both math and reading standards are high.

Given the childhood obesity rate differs across race, ethnicity and gender, the effects of grading standards will be evaluated separately for each of these groups to assess whether grading standards may have had heterogeneous treatment effects on the population using individual and year fixed effects model. We first stratify data by gender. In the case of female students, student BMI is significantly higher for high standard group versus low standard group for both math and reading, while high math grading standards do not increase the probability of being overweight for female students. However, high reading grading standards increase the probability of being overweight for female students by 13 percentage points and a combined high standards increases this probability by 5.5 percentage points.

Associations between teacher grading standards and children weight status by race and ethnicity are shown in Table 3-6 columns three to six. Interestingly, White students respond to the high grading standards greater than any other race/ethnicity groups. Though all estimates remain positive and are with comparable magnitudes, they are not statistically significant except that African-American and Asian students have higher BMI if they are under both high math and reading standards and for Hispanic students under high reading standards. None of African-American, Hispanic and Asian students present higher probability of being overweight when they are treated with higher grading standards.

Previous studies in this literature find evidence of differential effects of school and teacher grading standards, with initially high-performing students benefiting more from high grading standards (Betts and Grogger 2000; Figlio and Lucas 2004). Therefore, Table 3-6 column seven and eight shows results testing heterogeneous grading standards effects on children's health by dividing the sample into two subsamples based on children's direct assessment scores. Overall, we observe consistent results, with stronger effects on high-achieving students versus their low-achieving counterparts.

Studies in Childhood obesity literature have shown that obesity is a significant risk factor for adverse school outcomes (Datar and Sturm 2006; Cawley and Spiess 2008). If teacher rating standard is affected by students' non-academic characteristics, for instance students with higher BMI may receive a lower rating, the effects of grading standard will be overestimated. We use a unique method to define our standard measures in this study to avoid such endogeneity caused by simultaneity. Though we have a rather large sample size, with about 18,176 teachers sampled in the ECLS-K, we should be cautious when generalizing our results.

Conclusion

The existing literature has focused the majority of its attention on the treatment effects of grading standards or teachers' evaluation standards on student test scores. However, given that childhood obesity rate increased dramatically over the last decades and more researchers realized that education policies may post unintended effects on children's health, the purpose of this paper is to complement and extend the existing literature to determine if teacher level grading standards had any effects on child BMI and the probability of being overweight.

More specifically, we study teacher grading standards across two subjects and analyze how they relate to children's weight outcomes. We find that, conditional on a large set of individuals' observable characteristics that includes demographic information, family background, school characteristics, and teacher characteristics, students under higher teacher grading standards across subjects have higher BMI and are more likely to be overweight. Moreover, this effect is stronger for female and White students comparing to their counterparts. These results suggest that while high grading standard may help improve student academic performance, it may at the same time harm children's health, in this case children's weight.

Throughout the paper, the results consistently indicate that students' health respond uniquely to teacher grading standards. This suggests that school teachers should give careful thought to the grading standards when they are considering how these standards can be improved to promote academic performances. We certainly do not want to improve students' academic performances at the expense of their health. As childhood obesity has become one of the biggest health concerns in this nation, teachers, school and parents should realize that school environment is one of the main factors that may associates with this weight gain. More research is needed in this literature to examine the unintended impacts of other education policies on children's health outcomes. Policy makers should keep in mind that a single incentive policy that is designed to promote academic performance may unintentionally harm children's health as teachers and schools pursuing higher test score gains may undertake unwanted actions that cause irreversible impact on children.

Table 3-1. Summary Statistics of All Variables

Variables *ECLS-K Wave*	*Mean (Standard errors)* *1999 (W2)*	*Mean (Standard errors)* *2007 (W7)*	*Definition/Questions from ECLS-K*
Body Mass Index	16.358 (0.055)	22.843 (0.117)	Weight in kilograms divided by height in meters squared
Overweight students	0.262 (0.008)	0.374 (0.010)	Dichotomous variable that equals 1 if body mass index ≥85th percentile for children of the same gender and age
Male	0.477 (0.013)	0.498 (0.011)	Dichotomous variable that equals 1 if respondent is Male
Age	6.26 (0.01)	14.288 (0.008)	Child's age in years at the time the direct child assessment occurred.
White	0.662 (0.013)	0.627 (0.011)	Dichotomous variable that equals 1 if respondent is White
African American	0.147 (0.011)	0.141 (0.009)	Dichotomous variable that equals 1 if respondent is African American
Hispanic	0.133 (0.008)	0.164 (0.008)	Dichotomous variable that equals 1 if respondent is Hispanic
Asian	0.02 (0.002)	0.027 (0.002)	Dichotomous variable that equals 1 if respondent is Asian
Other race	0.038 (0.004)	0.04 (0.004)	Dichotomous variable that equals 1 if respondent is Other race
Mom's age	32.716 (0.218)	40.718 (0.209)	Age of resident mother, female guardian or mother figure
Dad's age	28.351 (0.464)	33.78 (0.483)	Age of resident father, male guardian or father figure
Mom_White	0.68 (0.013)	0.647 (0.011)	Dichotomous variable that equals 1 if mother is White
Mom_African American	0.138 (0.011)	0.132 (0.009)	Dichotomous variable that equals 1 if mother is African American
Mom_Hispanic	0.119 (0.008)	0.142 (0.007)	Dichotomous variable that equals 1 if mother is Hispanic
Mom_Asian	0.021 (0.003)	0.031 (0.003)	Dichotomous variable that equals 1 if mother is Asian
Mom_other race	0.023 (0.003)	0.024 (0.003)	Dichotomous variable that equals 1 if mother is other race
Dad_White	0.592 (0.013)	0.547 (0.011)	Dichotomous variable that equals 1 if father is White
Dad_Black	0.068 (0.008)	0.069 (0.006)	Dichotomous variable that equals 1 if father is African American
Dad_Hispanic	0.093 (0.007)	0.112 (0.006)	Dichotomous variable that equals 1 if father is Hispanic
Dad_Asian	0.018 (0.002)	0.024 (0.002)	Dichotomous variable that equals 1 if father is Asian

Table 3-1. Continued.

Dad_Other race	0.015 (0.002)	0.016 (0.003)	Dichotomous variable that equals 1 if father is other race
Mom's Education__highschool	0.306 (0.013)	0.211 (0.009)	Dichotomous variable that equals 1 if mother's education is high school
Mom's Education__some college	0.265 (0.012)	0.304 (0.011)	Dichotomous variable that equals 1 if mother's education is some college
Mom's Education__bachelor degree	0.174 (0.01)	0.19 (0.009)	Dichotomous variable that equals 1 if mother's education is bachelor degree
Mom's Education__graduate degree	0.017 (0.003)	0.025 (0.003)	Dichotomous variable that equals 1 if mother's education is some graduate degree
Mom's Education_master degree	0.051 (0.006)	0.072 (0.006)	Dichotomous variable that equals 1 if mother's education is master degree
Mom's Education_doctorate degree	0.018 (0.003)	0.022 (0.003)	Dichotomous variable that equals 1 if mother's education is doctorate degree
Dad's Education__highschool	0.257 (0.011)	0.19 (0.009)	Dichotomous variable that equals 1 if father's education is high school
Dad's Education__some college	0.181 (0.01)	0.173 (0.008)	Dichotomous variable that equals 1 if father's education is some college
Dad's Education__bachelor degree	0.139 (0.009)	0.15 (0.008)	Dichotomous variable that equals 1 if father's education is bachelor degree
Dad's Education__graduate degree	0.015 (0.004)	0.025 (0.004)	Dichotomous variable that equals 1 if father's education is some graduate degree
Dad's Education_master degree	0.046 (0.005)	0.062 (0.005)	Dichotomous variable that equals 1 if father's education is master degree
Dad's Education_doctorate degree	0.035 (0.005)	0.043 (0.004)	Dichotomous variable that equals 1 if father's education is doctorate degree
Family socioeconomic status (SES) status	0.074 (0.021)	-0.039 (0.018)	Family Socioeconomic scale
Familuy poverty level	1.832 (0.01)	1.829 (0.009)	Dichotomous variable that equals 1 if family below poverty threshold
Teacher_Male	0.019 (0.003)	0.157 (0.007)	Dichotomous variable that equals 1 if respondent is male
Teacher_Hispanic	0.033 (0.004)	0.043 (0.004)	Dichotomous variable that equals 1 if respondent is Hispanic
Teacher_Asian	0.013 (0.003)	0.015 (0.003)	Dichotomous variable that equals 1 if respondent is Asian
Teacher_African American	0.061 (0.007)	0.078 (0.006)	Dichotomous variable that equals 1 if respondent is African American

Table 3-1. Continued

Teacher_White	0.923 (0.007)	0.905 (0.007)	Dichotomous variable that equals 1 if respondent is White
Schooltype_Public	0.831 (0.01)	0.888 (0.007)	Dichotomous variable that equals 1 if respondent's school is a Public school
Schooltype_Private	0.031 (0.005)	0.019 (0.003)	Dichotomous variable that equals 1 if respondent's school is a Private school
Schooltype_Catholic	0.078 (0.005)	0.052 (0.004)	Dichotomous variable that equals 1 if respondent's school is a Catholic school
Schooltype_Others	0.061 (0.007)	0.04 (0.004)	Dichotomous variable that equals 1 if respondent's school is other types of schools
School_total enrollment	3.258 (0.03)	3.95 (0.024)	Total school enrollment
School_free lunch	29.061 (0.721)	29.199 (0.561)	Percentage of students eligible for free lunch in school
School_reduced lunch	8.382 (0.305)	8.446 (0.195)	Percent of students eligible for reduced price lunch in school
Region_Midwest	0.246 (0.011)	0.261 (0.01)	Dichotomous variable that equals 1 if respondent's is from Midwest
Region_South	0.415 (0.013)	0.386 (0.011)	Dichotomous variable that equals 1 if respondent's is from South
Region_West	0.176 (0.01)	0.176 (0.009)	Dichotomous variable that equals 1 if respondent's is from West

Notes: *Summary statistics and standard errors (in parentheses) from authors' calculations from the 1999,2000,2002,2004 and 2007 Early Childhood Longitudinal Study-Kindergarten Class of 1998-1999.

* Africa American and Hispanics are oversampled in the ECLS-K, results above are weighted means

Table 3-2. Comparison of BMI and Obesity Across Year and Gender

	1999		2000		2002		2004		2007	
Gender	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
BMI Mean	16.457 (2.268)	16.339 (2.352)	16.928 (2.829)	16.879 (2.912)	18.73 (3.89)	18.595 (3.88)	20.682 (4.782)	20.468 (4.73)	22.783 (5.898)	23.015 (6.653)
BMI Median	21.698	22.490	21.480	22.530	17.580	17.570	19.330	19.290	21.180	21.630
BMI 95th Percentile	20.790	20.903	22.580	22.710	26.770	26.230	30.100	29.650	33.530	33.630
Fraction Overweight	26.3%	24.6%	27.6%	25.6%	36.0%	33.3%	42.1%	36.7%	36.7%	34.9%
Number of Obs.	9,877	9,444	7,820	7,490	6,925	6,717	5,276	5,245	3,971	3,953

Notes: *Summary statistics and standard errors (in parentheses) from authors' calculations from the 1999, 2000, 2002, 2004, and 2007 Early Childhood Longitudinal Study-Kindergarten Class of 1998-1999.

* All the BMI related calculation excluded observations with 40<BMI<0

* Overweight is defined as a BMI above the 85th percentile for children of the same age and gender

* Total number of observations is 66,718

Table 3-3. Teacher grading standards and observed teacher characteristics

Teacher Characteristic	High Standards	Median Standards	Low Standards
Female	86.26	88.35	97.53
White	0.925	0.914	0.875
African American	0.044	0.043	0.079
Hispanic	0.040	0.044	0.055
Asian	0.019	0.021	0.022
Teacher has a Bachelors degree	23.59	27.16	20.22
Teacher has a Masters degree	37.81	30.38	40.17
Teacher has regular certification	69.58	48.23	3.39
Teacher has temporary certification	16.69	29.43	37.22

Notes: * Each column of the table presents weighted means.
* Statistical significance level: ***: p<.001; **:p<.01; :p<.05;
* Standard errors are cluster-corrected by state.
* All models include controls for individual race, gender and age.

Table 3-4. Means of dependent variables and selected student characteristics

Variable	High Standards	Median Standards	Low Standards
BMI	21.166	18.634	16.883
Overweight	0.381	0.304	0.261
Female	48.11	54.33	51.5
White	56.93	67.75	68.94
African American	12.71	6.83	7.91
Hispanic	19.68	12.84	12.47
Asian	4.84	7.52	5.85
Free lunch	32.445	25.556	22.36
Reduced lunch	8.067	7.593	7.535
Public school	82.72	82.16	75.77
Catholic school	11.38	10.48	14.67

Notes: * Each column of the table presents weighted means.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05;
* Standard errors are cluster-corrected by state.
* All models include controls for individual race, gender and age.

Table 3-5. Estimated effects of teacher grading standards on students' BMI and the probability of being overweight

Dependent Variable: BMI						
Model	(1)	(2)	(3)	(4)	(5)	(6)
High Math grading standards	0.219***	0.3756***	0.1946	0.1749	0.173***	0.0670^
	(0.040)	(0.0886)	(0.1063)	(0.1090)	(0.049)	(0.038)
High Reading grading standards	0.263***	0.2142***	0.1561*	0.1752*	0.179***	0.0694^
	(0.040)	(0.0606)	(0.0701)	(0.0764)	(0.038)	(0.0371)
High Math & Reading gs	0.301***	0.1322***	0.0889*	0.0802*	0.104***	0.0610**
	(0.025)	(0.0335)	(0.0386)	(0.0400)	(0.018)	(0.0210)
Dependent Variable: 1(Overweight)						
High Math grading standards	0.2928***	0.1362**	0.0444	0.0450	0.0568*	0.1012
	(0.0202)	(0.0460)	(0.0526)	(0.0551)	(0.0290)	(0.0608)
High Reading grading standards	0.1470***	0.0588	0.0469	0.0718	0.0947***	0.1405**
	(0.0203)	(0.0308)	(0.0355)	(0.0391)	(0.0222)	(0.0433)
High Math & Reading gs	0.0973***	0.0380*	0.0185	0.0212	0.0326**	0.0487*
	(0.0115)	(0.0160)	(0.0180)	(0.0195)	(0.0106)	(0.0226)
Student covariates	NO	YES	YES	YES	YES	YES
Family covariates	NO	NO	YES	YES	YES	YES
School covariates	NO	NO	NO	YES	YES	YES
Teacher covariates	NO	NO	NO	YES	YES	YES
School fixed effects	NO	NO	NO	NO	YES	NO
Student fixed effects	NO	NO	NO	NO	NO	YES
Year fixed effects	NO	NO	NO	NO	YES	YES
Sample size	40,663	32,689	22,164	19,385	30,493	7209

Table 3-5. Continued

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression.

* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^<0.1
* All regression analysis used ECLS-K longitudinal sample weight.
* Detailed covariates used in the models can be found in Table 1.

Table 3-6. Estimated differential effects of teacher grading standards on students' BMI and probability of being overweight

Model	Female	Male	White	African American	Hispanic	Asian	High Achievers	Low Achievers
Dependent Variable: BMI								
High Math grading standards	0.1112^	0.1150	0.0987^	0.1964	0.1452	0.4420	0.1800*	0.0225
	(0.0618)	(0.0665)	(0.0606)	(0.2631)	(0.1594)	(0.2319)	(0.0737)	(0.0631)
High Reading grading standards	0.2264***	0.1035*	0.1541**	0.1637	0.2975**	0.1923	0.1867***	0.1705***
	(0.0475)	(0.0490)	(0.0493)	(0.1608)	(0.1071)	(0.1740)	(0.0539)	(0.0431)
High Math & Reading gs	0.1049***	0.0480*	0.0806***	0.1375*	0.0716	0.2428**	0.0718**	0.1095***
	(0.0219)	(0.0225)	(0.0214)	(0.0608)	(0.0445)	(0.0804)	(0.0218)	(0.0275)
Dependent Variable: 1(Overweight)								
High Math grading standards	0.0560	0.0674	0.0264	0.1233	0.0952	0.2109	0.0611^	0.0455
	(0.0360)	(0.0401)	(0.0341)	(0.1262)	(0.0874)	(0.1721)	(0.0359)	(0.0490)
High Reading grading standards	0.1361***	0.0266	0.0926***	0.1137	0.1129	0.1453	0.0869***	0.0997**
	(0.0283)	(0.0293)	(0.0277)	(0.0773)	(0.0579)	(0.1299)	(0.0259)	(0.0329)
High Math & Reading gs	0.0551***	0.0090	0.0366**	0.0380	0.0067	0.0733	0.0211*	0.0546**
	(0.0114)	(0.0121)	(0.0113)	(0.0250)	(0.0212)	(0.0465)	(0.0107)	(0.0167)
Sample size	14,014	13,400	16,260	1,680	2,856	614	16,446	12,519

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05; ^<0.1
* All regression analysis used ECLS-K longitudinal sample weight and controlled for student and year fixed effects.
* Detailed covariates used in the models can be found in Table 1.

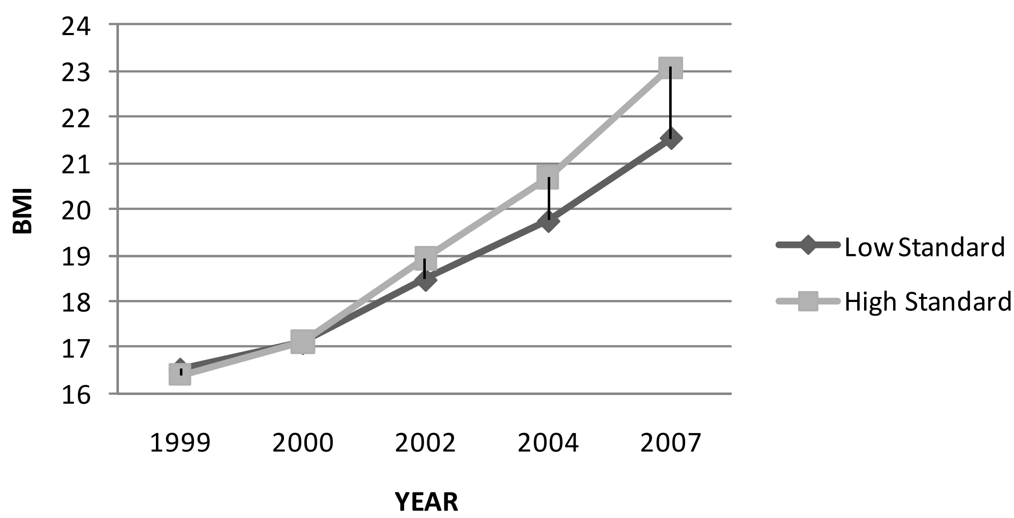


Figure 3-1. Student BMI by teacher grading standards

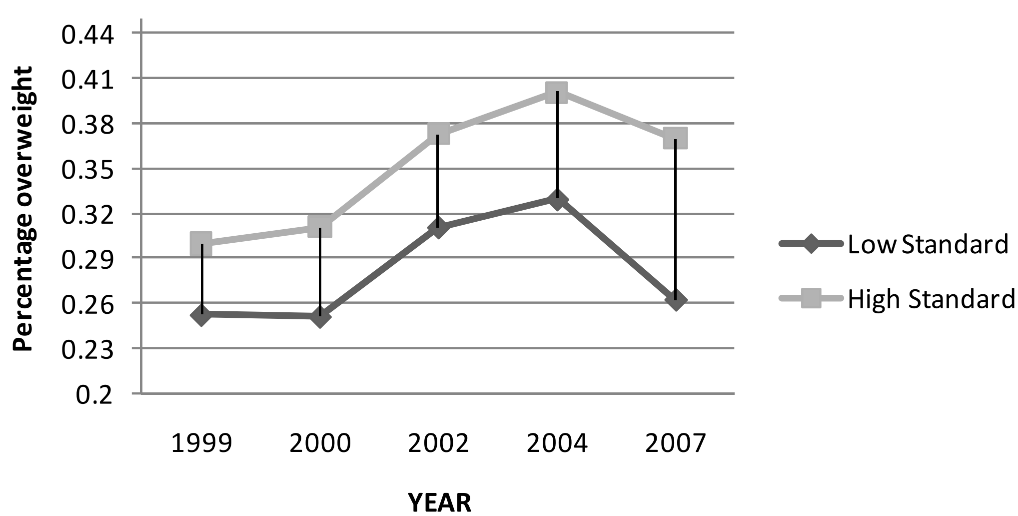


Figure 3-2. Student overweight rate by teacher grading standards

CHAPTER 4
IS IT WISE TO INVEST IN SEX EDUCATION? ABSTINENCE EDUCATION PROGRAM AND ADOLESCENT RISKY BEHAVIORS

Introduction

In 1996, the welfare reform law added Title V Section 510 (b) of the Social Security Act which established a new funding stream to provide grants to states for abstinence-only [1] programs. The new federal legislation requires states that choose to accept these funds to teach that abstinence is the only certain way to avoid pregnancy and sexually transmitted diseases (STDs), and may not in any way advocate contraceptive use or discuss contraceptive methods except to emphasize their failure rates. Beginning in fiscal year 1998, the Title V abstinence-only program has allocated $50 million annually to such programs that teach abstinence from sexual activity outside of marriage as the expected standard for school-age children. Under a matching block grant program, states must match this federal funding at 75 percent, resulting in a total of up to $87.5 million annually for Title V, Section 510 abstinence education programs (Trenholm et al. 2008).

With the official implementation of abstinence-only programs and the infusion of the federal funds across the nation, a new round of debate over the effectiveness and efficacy of abstinence-only sex education begins to heat up. A few decades ago, debate over sex education yet focused on whether schools should provide information on sex related matters. Supporters of no school sex education believe that providing such information to adolescents may promote their sexual behaviors, while advocates of sex

[1] Abstinence education focuses on teaching young people to abstain from sex until marriage. Comprehensive sex education does not focus either solely or so closely on teaching young people that they should abstain from sex until marriage. Although they do explain to young people the potential benefits of delaying having sex, they also make sure that they are taught about contraceptives.

education at the time argue that the probability of an adolescent becoming sexually active will not be affected solely due to the sexual information provided to her/him at school. Therefore, it is necessary for schools to provide accurate sex education which will help to prepare these will-be-sexually-active young adults to acquire the necessary knowledge about pregnancy and STD and HIV/AIDS prevention.

According to recent studies, in 2007 close to half of (47.8%) all teenagers 15 to 19 years old in the United States have had sexual intercourse at least once and 7.1% of them initially had sexual intercourse before age 13[2]. Coincidentally, the highest age-specific rates of reported Chlamydia and Gonorrhea, the most commonly-reported sexually transmitted disease (STD) in the United States, for women in 2007 were among those 15 to 19 years old of age (3004.7 and 647.9 cases per 100,000 females, respectively), which increased by 12.4% and 2.6% compared to 2003. While the number for adolescent males is substantially lower, both Chlamydia and Gonorrhea case rates among them have increased steadily since 2003[3]. Both of these diseases have also been associated with increased HIV transmission. Estimated numbers of cases of HIV/AIDS for the same age group increased 20 percent from 1010 cases per 100,000 in 2001 to 1213 in 2005[4]. In addition, the birth rate for U.S. teenagers 15 to 19 years rose 3% to 41.9 births per 1000 females in 2006, the first increase reported since 1991[5]. Given current patterns of adolescent sexual behaviors and the alarming

[2] Centers for Disease Control and Prevention, Youth risk Behavior Surveillance Summaries, June 2008. MMWR 2008; 57.

[3] Centers for Disease Control and Prevention, Sexually transmitted Diseases Surveillance, 2007

[4] Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2005. Vol.17.

[5] National Vital Statistics Report, 2007. Volume 56, Number 7

numbers of STDs and HIV incidences as well as teenage pregnancy rates, it is safe to believe that preventing teenagers from being sexually active is at best unrealistic and it is urgent to develop effective education programs to prevent the HIV/AIDS and other STDs epidemic.

The ongoing debate over sex education is now between advocates for abstinence-only-until-marriage programs, focusing on teaching young people to abstain from sex until marriage, and supporters of comprehensive sex education, which includes contraceptives information besides abstinence, in the nation's schools.

With the incessant increase of the number of STD/HIV infections among adolescents and millions of dollars flooding into abstinence-only programs, the unavoidable question is whether the money is well-spent.

Studies in this literature have provided fruitful evidence showing that sex education in general does not promote students' sexual behaviors in any way. Social scientists and lately economists have been longing for discovering further evidence of the potential differential effects, if there is any, of abstinence-only and comprehensive sex education have upon adolescent sexual behaviors. Credible estimates of such analysis, however, are limited. Specifically, most of the empirical work can be divided into two categories based on their research methods: quasi-experimental or pre-post approach and a randomized controlled evaluation technique. Much of the studies that adopted quasi-experimental approaches have used cross-sectional data (Lerner 2004; Doniger Riley, Utter, and Adams 2001; Ku, Sonenstein, and Pleck 1992, 1993). Though several studies found that exposure to abstinence-only sex education has positive effects on delay of first intercourse and increasing contraceptive use, failing to control for the unobservables between programs participants and comparison groups make these results susceptible (Ku, Sonenstein, and Pleck 1992, 1993). Another stream of studies, including the most recent study conducted by Trenholm et al. (2008), have used randomized experimental design to test the effects of a specific school program on students' sexual behaviors. They examined the impacts of four abstinence-only education programs on adolescent sexual activity and risks of pregnancy and STDs using survey data collected in 2005 and early 2006 from more than 2,000 teens who had been randomly assigned to either a program group that was eligible to participate in one of the four programs or a control group that was not. The findings show no significant impact on teen sexual activity, no differences in rates of unprotected sex, and some impacts on knowledge of STDs and perceived effectiveness of condoms and birth control pills.

So far, all the studies on sex education have been focusing on program-evaluation at school levels, although major sex education policies had been made at the state level. Therefore, our paper serves to fill this gap in the sex education literature and provides further insights on the current issue from at least three important angles. First, we collected detailed data on state sex education laws from various sources regarding when states have sex education mandates and if they do, whether they require comprehensive or abstinence-only sex education. Subsequently, employing a rich micro-level data set, we use various panel data techniques to identify the impact of state sex education programs has upon their students. We believe a study of outcomes at state level may paint a clearer picture of the benefits of comprehensive sexuality versus abstinence-only programs. However, critics of such method may argue that state sex education is endogenous due to the 1996 reform which may have altered some states' sex education focuses and hence bias our estimates in unknown directions. Therefore, we first estimate the heterogeneous impacts of the 1996 welfare reform on state sex education legislatives. Based on each state's policy change affected by the Title V 510, this is the first paper to employ the Interrupted Time-Series design to identify the causal effects of state sex education on students' sexual behaviors. Although all 50 states applied for the abstinence-only program funds in 1997, there are significant variability and creativity in the program planning and implementations. Without proper control for the observable and unobservable program variations across states and the potential pre-existing state trends, we will not be able to consistently estimate the impacts of 1996 reform through a simple pre-post mean differences analysis. In order to overcome the problem and to try to identify causal effects, we requested access to a detailed state-by-state survey conducted by Sexuality Information and Education Council of the United States (SIECUS) in winter 1997. SIECUS focus its research on the first-year implementation of the abstinence-only-until-marriage program by conducting a survey of state abstinence program administrators about a variety of issues related to the content and implementation of their Section 510(b) program. Especially, according to the report state plans vary in the age of target audience, media campaigns, in-school curricula, after-school curricula and etc. We then construct an index score for each state to evaluate the state-specific program implementation and then divide them into high implementation states and low implementations states based on their scores. This time series contrast with greater statistical power allow us to better identify the effects of Section 510(b).

Last, most previous cross sectional studies used relatively older data set or use only one or two-period data, while we employed Youth Risk Behavior Survey Surveillance System of years from 1993 to 2005, which not only allows us to control for state by year trend but also provides us information about the most current pattern of adolescent sexual behaviors. Additionally, this unique data set enables us to examine a variety of different outcomes with the most efficient estimators as well as to investigate the heterogeneity of effects across gender, racial groups and grades.

The remainder of this paper is organized as follows. Section 2 provides some background on the abstinence-only programs. Section 3 discusses the theoretical model followed by empirical strategy and Section 5 describes the data. Section 6 presents the main findings and Section 7 concludes.

A Brief Background on Abstinence-only programs

Federal funding for teen pregnancy prevention in the form of abstinence-only education began in 1981 with the passage of the Adolescent Family Life Act (AFLA). Ever since then, around 4 million dollars annually are invested in abstinence-only programs across the nation until 1996 when the 1996 welfare reform law added Title V, Section 510 (b) of the Social Security Act which established a new funding stream to provide grants to states for abstinence-only programs. Under the Title V abstinence-only program, states that choose to accept these funds may not in any way advocate contraceptive use or discuss contraceptive methods except to emphasize their failure rates.

Every state has at one time accepted Title V funds, though, in 2009, nearly half the states no longer participate in this program. Under Title V, states decisively coalesced around youth 10 to 14 years of age-in upper elementary and middle schools-as the intended audience for their efforts because students in this age group most likely have not begun to explore sexual relationships. If the intended or say the treated group are students aged between10 and 14, the first treated group observed from YRBSS are students in 9th grade in 1999, and the second treated group is students in 9th to 11th grades in 2001. Therefore, to include the entire data when analyzing the effects of 1996 reform will result in inconsistency of the estimates. We thus carefully define treated and control groups based on these information. Details are discussed in the method section. In October 2000, the federal government created another funding stream to support abstinence-only programs, Community-Based Abstinence Education (CBAE). Different from the 1996 reform, CBAE is managed directly by Family and Youth Services Bureau (FYSB), Administration for Children and Families, Department of Health and Human Services rather than providing funds to states directly to gain control over the programs and hence to adhere to the tenets of the welfare law more closely. In addition, majority of the funding is provided to public and private institutions for community-based abstinence education projects rather than school based programs. The objective of the grants is to reduce teen pregnancy rate and STD by teaching abstinence to adolescents ages 9 through 18 and by creating an environment within communities that support decisions to postpone sexual activity. Although a great amount of funds flooded into community abstinence programs, they did not alter the focus of state sex education or programs implemented at school level. We further control the implementation of CBAE, in case of any unobserved bias. Between 1996 and 2008, over $1.5 billion dollars were spent to promote abstinence-only programs through these three main tunnels and for fiscal year 2009 alone, the total amount allocated was just over $160 million.

Theoretical Framework

In this section, we provide the theoretical basis for linking state sex education mandate to sexual outcomes. The conceptual model of sexual decision making in this work is built on the model of sexual activity and pregnancy proposed by Oettinger (1999) and Tremblay and Ling (2005). We show that the relationship varies with individual preferences and with the type of sex education (risk-altering, risk-revealing, and utility-altering). Rational individuals become sexually active if the expected utility of intercourse exceeds the cost, for instance, becoming pregnant or contracting STD or HIV/AIDS. As we do not have data regarding STD or HIV/AIDS (only available in 2005 YRBSS) infection, we assume the only cost is the risk of becoming pregnant. Assume the perceived probability of becoming pregnant *P*, where $0 \le P \le 1$, depends on the optimal use of condom (k^*). The term U(y|k*) is defined as the net utility of sex in the present period plus the present value of expected future net utility. Net utility varies across pregnancy *y*, where *y=p* if the individual is pregnant from an incident of intercourse and *y=np* if the individual is not pregnant. An adolescent chooses to engage in sexual intercourse rather than abstinence (*V*) if and only if:

$$P(y|k^*)U(y|k^*) + [1 - P(y|k^*)][1 - U(y|k^*)] > V$$

Let $U(n|k^*) = 1 - U(y|k^*)$

we have

$$P(y|k^*)U(y|k^*) + [1 - P(y|k^*)]U(n|k^*) > V$$

or $P(y|k^*) < [U(n|k^*) - V]/[U(n|k^*) - U(y|k^*)]$

Equivalently, with $Z \equiv [U(n|k^*) - V]/[U(n|k^*) - U(y|k^*)]$, the individual chooses to be sexually active or not if and only if $P(y|k^*) < (>) Z$.

Similarly, a sexually active teen will choose to use a condom if and only if the expected benefit of using a condom exceeds the cost, i.e. becoming pregnant or causing someone pregnant:

$$P(y|c)U(y|c) + [1 - P(y|c)][1 - U(y|c)] > P(y|nc)U(y|nc) + [1 - P(y|nc)][1 - U(y|nc)] .$$

Rearranging them, we have

$$P(y|c) - P(y|nc) < [U(ny|c) - U(ny|nc)]/[U(y|c) - U(y|nc)] .$$

Sex education may influence an individual's ultimate decision regarding sexual activity by revealing or altering the probability of pregnancy and by affecting the relative utility of abstinence. ***Risk-altering sex education*** includes courses that provide information of alternative contraceptive methods which may allow sexually active teens to alter the risk of pregnancy (*P*). Comprehensive sex education including contraceptives would fall in this category. In our model, if comprehensive sex education provides information regarding condom use which in turn affects teens perception on the risk of pregnancy, the risk altering sex education will increase the probability of becoming sexually active conditional on the usage of condom, but has ambiguous effects on the probability becoming sexually active conditional on if condom is not used or the probability of abstinence conditional on the optimal condom usage. The probability of condom usage will increase if schools provide students information about safe sex and the benefits of condom usage. It is also likely that the probability will decrease if other forms of contraceptives are introduced to the students. Define *K=other* as if student chooses any other type of contraceptives besides condom, the probability of choosing other contraceptives conditional on sexually active and no condom usage increases $P(K = other | a = 1, nc)$ if other method introduced to the students. Next, if a student is already sexually active, we might be also interested in their other sexual behaviors, for instance, number of sexual partners. Under risk-altering sex education, if $P(p | k^*)$ decreases, we would observe an increase in the number of sexual partners.

Utility altering sex education changes the teens' perceived utilities (U or V) throughout their adolescence. Abstinence sex education that teach strategies for resisting sex (thereby increasing V), that highlight the costs of teen parenthood (presumably reducing U), or that provide information about abortion (likely increasing U) is more likely to be in this category. Therefore the expected effects depend on the relative change of utility of abstinence (V) and the utility of intercourse. Furthermore, if utility altering sex education affects the relative utility of having intercourse and abstinence but does not affect the probability of becoming pregnant with or without using condoms, the probability of using condom or any other contraceptives will not change. If the sex education increases the utility of having sex, we might also observe an increase in the number of sexual partners among sexually active young adults.

Finally, ***risk-revealing*** sex education provides accurate information to teens that initially may misjudge pregnancy risks. Instruction on contraceptive methods would be purely risk-revealing which again coincides with the focus of many comprehensive sex education programs. Having such sex education or not alters the probability of pregnancy conditional on the optimal condom usage. Additionally, if the probability of becoming pregnant increases due to the sex education, we would expect an increase of condom or other contraceptives use and decrease otherwise.

The expected signs of the impacts of these types of sex education on the probability of engaging in intercourse given the optimal condom choice are listed in Table 4-1. We test these hypotheses in the following sections.

Empirical Framework

The theoretical model portrays a simplified model without taking into account of the potential heterogeneity in individual preferences, the differential state policies and the frequency of sexual activity. In this section, we develop regression models incorporating these issues. To test for the existence of differential effects associated with state sex education programs, we estimate the marginal propensities of becoming sexually active and using condoms, respectively. As the probability of conducting sexual intercourse and the probability of using condoms given they are sexually active are dichotomies, probit or logit model would be the appropriate.

We start our analysis with a standard nonlinear panel model:

(1) $$\Pr(y_{ist} = 1 | x_{ist}, SexEd_{st}) = f(\alpha + \beta_1 x_{ist} + \beta_2 SexEd_{st}, \varepsilon)$$

where y_{ist}.represents a series of student sexual behaviors for student *i* residing in state *s* during year *t*, for instance if a student is sexually active or not, x_{ist} is a set of individual covariates, and $SexEd_{st}$ is our main independent variable of interest which measures different state level sex education policies. In addition, we assume the state effect is invariant across time and if this is the case, pooled logit and probit will provide similar and consistent estimates. However, certain unobserved state-specific characters may bias our estimates because x_{it} may contain lagged y_{it}. For instance, states with higher percentage of students who are sexually active in year 1999 might expect a similar outcome the following year. Similarly, states' sex education may have lagged effects on students' sexual behaviors or their effect might not become present until the next period. If this is the case, a more sensible model is the unobserved effects logit model. Model (1) thus becomes:

(2) $\Pr(y_{ist} = 1 | x_{ist}, \alpha_s) = f(\alpha_s + \beta_1 x_{ist} + \beta_2 SexEd_{st}, \varepsilon)$

where α_s is the unobserved state-specific effect. If we further assume that:

(3) $(y_{i1,...,} y_{it})$ is independent once conditional on $(x_{ist}, SexEd_{ist}, \alpha_s)$ and

(4) $\alpha_s | x_{ist}, SexEd_{st}$ ~ Normal $(0, \sigma_\alpha{}^2)$

The random effects model will produce consistent estimates. Nonetheless, the assumption that α_s and $x_{ist,} SexEd_{st}$ are independent may not hold. For example, states with higher teen pregnancy rates or higher STD infection cases are more likely to implement stricter sex education programs. If we relax this assumption by allowing α_s and $x_{ist,} SexEd_{st}$ to be correlated, we might consistently estimate βs using fixed effects logit. Before we characterize our model, one thing should be pointed out that we obtained information regarding state level sex education policies for years 1995, 1999, 2001, 2003, and 2005. Given this short panel, the only legitimate model that can consistently estimate the impact of sex education at state level is the fixed effects logit model (Chamberlain 1980). In particular, we consider the following underlying latent model:

(5) $y^*_{ist} = \alpha_s + \beta_1 x_{ist} + \beta_2 SexEd_{st} + \varepsilon_{ist}$

where y^*_{ist} is a continuous but unobserved index of utility of being sexually active of individual *i* residing in state *s* during year *t*, x_{ist} is a vector of individual characteristics, $SexEd_{st}$ represents the type of sex education implemented in state *s* in year *t* which equals to 1 if state requires sex education mandate with contraceptive, 2 if state requires abstinence sex education, with states without sex education mandates as the baseline, α_s is an idiosyncratic fixed effect which account for state-specific time invariant unobservables, and ε_{ist} is the stochastic error term. Rather than observing $y*_{ist}$, we observe:

(6) $$y_{ist} = \begin{cases} 1 & if\ y*_{ist} > 0 \\ 0 & otherwise \end{cases}$$

and hence the logit model is:

(7) $$\Pr(y_{ist} = 1 \mid x_{ist}, \alpha_i) = \frac{\exp(\beta_1 x_{ist} + \beta_2 SexEd_{st} + \alpha_i)}{1 + \exp(\beta_1 x_{ist} + \beta_2 SexEd_{st} + \alpha_i)}$$

Chamberlain (1980) shows that such a fixed effects logit model can be estimated by conditional maximum likelihood (ML) (conditioning on the fixed effects) consistently. For *T*=*5*, we consider the set s_i= $\sum y_{ist} = 1,2,...,4$, then

(8) $$\Pr(y_{i11},..., y_{i15} \mid x_{i11},...., x_{i15}, SexEd_{11},...., SexEd_{15}, \alpha_i, \sum y_{ist}) = \frac{\prod_{t=1}^{5} \exp[(\beta_1 x_{ist} + \beta_2 SexEd_{st}) y_{ist}]}{\sum_{d \in D_i} \prod_{t=1}^{5} \exp[(\beta_1 x_{ist} + \beta_2 SexEd_{st}) d_t]}$$

where d_i is the set of all possible combinations of s_i ones and 5-s_i zeros, is independent of state-specific effect.

The primary concern with identification of β_2 in fixed effects logit is that the state may not change their sex education frequently and hence will be cancelled out when differencing out the state fixed effects. If that is the case, SexEd will not be specified. On the other hand, if state does vary their sex educations from year to year, this could be endogenous due to the pre-existing state trend or due to certain state-specific characters, for instance high teen pregnancy rate/high percentage of teenage had sex. Last, the 1996 welfare reform may also bias the estimate because it might influence state sex education legislative differentially. In addition, one might be worried that other contemporaneous programs or certain state socioeconomic characters may also affect students' sexual behaviors which will again bias the estimate.

To overcome these concerns to the most extent, we use the SIECUS data, including information for 46 states that participated in the survey, to construct an index score for each state:

(9) $$Index_s = \alpha + \alpha_s Z_s$$

where Z_s is a set of variables representing state-specific policy differentials collected from the SIECUS. We categorize these variables into five subgroups. First, we evaluate if a state make-up an advisory panels[6] which may help plan and develop the state programs. Second, we document if states specify intended audience for their programs. We notice that 45 out of the 46 states have a specific age-intended audience. Third, we include data regarding states' media campaigns; especially we specify type of method in the campaign, the initiate year and etc. Fourth, majority of the states provide funding to community-based organizations and hence we further classify these programs based on their nature, for example if they provide education programs or recreational programs or other programs. Also, we include the absolute number of grants awarded for each state. Last, we collect data on if states provide grants to education agencies. According to SIECUS, 3 states provided grants to do abstinence programs at elementary school level, 7 states at junior high and 5 states at high school levels. A

[6] 26 states and the District of Columbia had advisory panels of some sort (SIECUS 1999)

detailed list of variables included in the model is shown in Appendix 1. In total, we include 48 variables to estimate an index score for each state. With the handful of data, we, however, do not know the pattern of how these numbers are related. To discover the unknown trends in the data, we apply the Principal Component Analysis (PCA) to identify the pattern of the data and reduce its dimensions. Specifically, we first standardize the data by subtracting the mean and dividing by the standard deviation and then perform the component analysis which basically tells us the most significant relationship between the data dimensions. We used the calculated "components" constructing two different scores for each state. As component 1, in our analysis, accounted for around 30% of the variance and 45 components were needed to explain all of the variance. Therefore, we use component 1 and the summation of all 45 components as our index score respectively. We then define states with index score higher than the median as high implementation states and low implementations states otherwise, as showed below.

$$(10) \quad group_s = \begin{cases} High\,Dose & if\ Index_s > Index1_{median} / IndexTot_{median} \\ Low\,Dose & otherwise \end{cases}$$

Using aggregated state level data, for instance percentage of students who had intercourse, we employ the interrupted time-series design to estimate the average treatment effects of 1996 welfare reform. For this model to consistently estimate the average treatment effects, we will need information regarding the pre-trend for each state. The data employed in this study allow us to go back to 1991 which leaves us 3-wave data prior to the introduction of 1996 reform. We use these data to calculate the potential pre-existing sexual behaviors trends and thus look for a sharp decrease/increase, if we assume Title V has negative effects on the probability of students engaging in sexual activities, in sexual behaviors/performing safer sex, increased condom usage for instance. Furthermore, to account for the potential bias from time-varying state-specific characteristics, we include a set of variables and state-fixed effects to capture such effects. Using state level data on students' risky behaviors, we estimate the following equation:

$$\begin{aligned} Y_{st} = \beta_0 + \beta_1 trend_{1t} + \beta_2 trend_{2t} + \beta_3 group_s + \beta_4 (trend_{1t} \times group_s) \\ + \beta_5 (trend_{2t} \times group_s) + \beta_6 State_s + \alpha_s + \varepsilon_{st} \end{aligned} \tag{11}$$

where $trend_{1t}$ takes on value of $(0,2,4,6,6,6,6,6,6)$ for year $(1991,...,2007)$ and $trend_{2t}$ equals to $(0,0,0,0,2,4,6,8,10)$ for year $(1991,...,2007)$ respectively[7], $group_{st}$ is a dummy variable which equals to one if the state is a high implementer and zero otherwise, β_4 captures the difference in average growth rate between groups prior to 1996 Reform and β_5 indicates how the comparison groups differ in the difference in growth rate from before to after the reform. This equation enables us to break up the potential nonlinear growth trajectories for high and low implementers into separate linear components. Specifically, we will observe the heterogeneous growth rates for high and low implementers during before and after 1996 reform respectively. Essentially, this is a difference-in-differences-in-differences (DDD) analysis where the first difference is a within high dose and low dose states change over time, the second difference is a within state pre-post reform change, and the last difference is between the high dose and low dose states pre and

---

[7] Another coding scheme is let $trend_{1t}$ equal to $(0,2,4,6,8,10,12,14,16)$ for year $(1991,...,2007)$ and $trend_{2t}$ equal to $(0,0,0,0,2,4,6,8,10)$. The only difference would be the interpretation. In this case, β_4 indicates the different growth rates between groups without the reform treatment and β_5 captures the heterogeneous increment (or decrement) to growth rates among comparison groups post reform treatment. Regression results, not shown here, present little difference and are available upon request from the author.

post reform. The advantage of this model over DDD is not only we observe the average treatment effects of 1996 on high implementers relative to low implementers; we will also observe the pre and post growth rates between these two types of states.

Data

Our dependent variables are drawn from the Youth Risk Behavior Surveillance System (YRBSS) for the years 1993 to 2005 which has been administered biennially since 1991 by the Centers for Disease Control and Prevention (CDC). The YRBSS system monitors six categories of priority health-risk behaviors among youth, including those relating to student sexual behaviors, information on HIV/AIDs education, and other risky behaviors. It includes national, state, and local school-based surveys of nationally representative samples of 9th through 12th grade students. Our main outcome variables include education programs delay the onset of sex, reduce the frequency of sex, reduce the number of sexual partners among teens, or increase the use of contraception. From Table 4-2 we see that, around 47% students in our data are sexually active and the average age of first time sex is 13.6. For students who had sexual intercourse, 59% of them chose to use a condom and on average they have 1.5 sexual partners. Other individual control variables are quite standard, for instance gender, race and grade. Results are shown in Table 4-2. Additionally, to control for school characteristics, we add two variables: if the respondent feels safe to be at school or if she/he carries weapons to school. We also include a few other individual control variables which may measure the respondent’s lifestyle, for example number of days drank alcohol/number of times used marijuana. Last, to control for time-varying state characteristics, we incorporate a set of state covariates to capture such differential effects, including state-level public school dropout data and student to teacher ratio in public elementary/secondary schools collected from Common Core of Data (CCD). These variables meant to control for the school quality which may in turn reflect students' tendency of behaving riskily. We also include data on percentage of people below poverty for each state (it is the below 100% poverty by weighted person count for all ages (in thousands)), total crime which equals to the sum of violent and property crimes both obtained from Census, and divorce rate obtained from CDC's National Vital Statistics based on provisional counts of divorces by state of occurrence (rates are per 1,000 total population residing in area) to control for state socioeconomic status which might influence teen behaviors.

Our data on state sex education is mainly collected from three sources: State Policies in Brief-Sex and STD/HIV Education for years 1995-2005 from Alan Guttmacher Institute (AGI), School Health Policies and Programs Study (SHPPS) conducted and maintained by CDC and Between the Lines: States' Implementation of the Federal Government's Section 510(b) Abstinence Education Program in Fiscal Year 1998 from collected by SIECUS.

Firstly, we requested access to the surveys data collected by AGI and then hand-coded data which subsequently are augmented with YRBSS using Fips code. AGI collects data annually on state sex education legislative since 1988, however, data are not coded in a consistent way until 1995 and there is no survey conducted in 1997. Therefore, we obtained 5 years data from years 1995, 1999, 2001, 2003, and 2005. AGI report documented, in detail, the type of sex education mandate in each state including if the state has comprehensive sex education, abstinence sex education, or no sex education mandate, if state has sex education mandate, HIV mandate, both sex education and HIV mandate or only sex education or HIV mandate. As shown in Table 4-2, around 36% of students reside in states where comprehensive sex education is provided while 13% of students live in states where abstinence sex education is required. However, information might not be available for every state in each year. Based on the nature of the survey data, we construct two main independent variables: if state has comprehensive sex education or abstinence-only sex education, if state has sex education or HIV mandate, or both of them or one of them or none of them and then analyze each policy variable respectively.

As mentioned in the previous section, SIECUS conducted a thorough survey in 1998 concerning the impact of 1996 reform had upon state sex education programs. According to their report, California and New Hampshire did not conduct any abstinence programs and withdrew their application for the funding in fiscal year 1998, South Dakota and South Carolina chose not to participate in the survey and Louisiana was still in the development state in the survey year and hence chose not to participate which left 46 states in the survey.

Last, we make use of the School Health Policies and Programs Study (SHPPS), a national survey periodically conducted to assess school health policies and practices at the state, district, school, and classroom levels, to evaluate their effects on adolescent sexual behaviors conducted by CDC. Specifically, we employ state-level data which were collected by computer-assisted telephone interviews or self-administered mail questionnaires completed by designated respondents in state education agencies in all 50 states and the District of Columbia. Survey was conducted in 1994, 1996, and 2006. For the purposes of the current study, these data are merged with the database with a one year lag (i.e. 2006 is merged with year 2005) under the assumption that state will not change its health policy significantly within one year. Especially, we include information on if providing pregnancy prevention programs, STD prevention programs and sex education is a state mandate.

Results

Nonlinear Panel Data Analysis

We begin our empirical work by estimating a simple pooled Logit model for the purpose of comparison to the more efficient estimators. All models employed YRBSS survey weights and conditional on gender, race, and age. Result presented in Table 4-3 column 1 suggests that states with abstinence-only sex education is associated with a 12 percentage points decrease in the probability of becoming sexually active while states with comprehensive sex education programs increases the probability of becoming sexually active by around 11 percentage points. We further include a set of school covariates to control for differential impacts due to school environment on students' risky behaviors. As showed in Table 4-3 column 2, the negative effects of abstinence-only sex education disappeared while the effects of comprehensive sex education doubled. When further controlling for state-specific time-varying covariates which might affect respondents' behaviors in an unobserved fashion, we see that abstinence-only sex education again present negative sign on the probability of being sexually active, however it failed to present any statistical significance. On the other hand, comprehensive sex education presents consistent positive and statistically significant effects. However, without controlling for state fixed effects, the estimates might be overestimated due to the endogeneity of state sex education policy. It is likely that states with higher teen pregnancy rate or HIV/AIDS rates are more likely to implement sex education earlier or in a more aggressive way. We thus include year and state fixed effects in the rest of the models presented in Table 4-3. Abstinence-only sex education presents negative and significant effects on the probability of a student being sexually active, however the effect became insignificant when conditioning on a set of state covariates. Similar pattern exists for comprehensive sex education but with positive sign. Overall, abstinence-only (comprehensive) sex education suggests a negative (positive) impact on students' decision of being sexually active or not, however this effect might be due to unobserved state characteristics or other unknown state policies. Consistent with our theoretical model, comprehensive sex education serves as risk-altering and risk-revealing types of programs reduces the probability of becoming pregnant due to sexual intercourse which in turn increases the probability of adolescent being sexually active. On the other hand, if abstinence-only education promotes the utility of abstinence relative to the utility of having sexual intercourse, we will observe a negative effect on the probability of being sexually active. Although, the empirical model fails to present any statistically significant effect, it does show a negative sign.

We employed the same models substituting the abstinence/comprehensive sex education measures with if a state has sex education mandate or HIV mandate or have both. From Table 4-1, we see that states with both sex and HIV education mandates suggest a lower probability of being sexually active in fixed effects models, however this effect became insignificant once controlling for state time-varying characteristics.

We next test the hypothesis that if risk revealing and risk altering sex education, here comprehensive, reduces the probability of becoming pregnant by providing information on safer sex, we should see an increase in condom usage. From the results showed in Table 4-3, one notices that comprehensive sex education indicates positive effects in all three fixed effects models while abstinence-only education significantly reduces the probability of using condoms. When we use sex and HIV education mandates as independent variables, we do not observe any statistically significant causal effects of these programs.

As mentioned previously, we also obtained data from SHPPS. These policies variables are collected directly from school administrators, which might control for the variation of programs school levels. Top panel of Table 4-5 presents results for intercourse outcome across three models. Programs on pregnancy indicates a negative and statistically significant effect on the probability of having sex without controlling for state covariates while programs on STD significantly lower such probability. Because state policies on STD program do not vary across years we fail to estimate its effects in fixed effects models. Bottom panel of Table 4-5 shows results for condom usage. We find a significant positive effect of STD programs on the probability of using condoms, which is consistent with our theoretical model as it serves as a risk revealing program.

To measure the impact of sex education on the usage of other birth control methods we began by examining the differential effects of abstinence-only and comprehensive programs. Using no birth control method as the omitted group, we find that audiences of abstinence-only education are less likely to choose birth control pill and use condoms. Although comprehensive sex education fails to present any statistically significant effects on any type of birth control methods once conditioning on state time-varying variables and state fixed effects, we find positive signs on using condom and Depo-Provera[8]. Similarly, sex and HIV mandates fail to show any differential statistically significant effects on choosing different birth control method using fixed effects model with state and individual covariates.

Another indicator of safe sex is the number of partners a respondent had during her/his life time, therefore we further test the effects of sex education on the number of sexual partners a respondent had with results showed in Table 4-8. As can be seen in column 3, only abstinence-only sex education indicates negative and statistically significant effects on number of partners for male students. Specifically, male students residing in states with abstinence-only sex education mandate on average have 0.5 less sexual partners in their life time. Last, we present results testing the effects of sex education on the deferring of first time sex in Table 4-9. When both Sex and HIV education are required at state level, the policies significantly decrease the age of initial sex for female students by nearly 1.7 years and when only HIV is required the magnitude decrease slightly to 1.1 years. However, we do not observe any significant effects on male students.

However, because these analyses did not take into account of the 1996 reform as well as their potential differential impacts on states' policy choices, it is difficult to trust the credibility of the results and interpret these results as effects of state sex education policies.

Interrupted Time Series Model

As Congress allocated $50 million in federal funds for the abstinence-only program starting from federal fiscal year 1998 (October 1,1997 –September 30,1998), we

[8] Depo Provera (also known as DMPA or Depot Medroxyprogesterone Acetate) is a hormone injection that lasts for 3 months to prevent pregnancy.

therefore use 1998 as the policy intervention point to estimate the differential effects it has upon state policies and then the policy effects on their audiences. Before using a regression framework, it is worthwhile to take a look at the differential trends between high and low implementers in figures. Using year 1998 as the cutoff point, figure 4-1-5 present the state level percentage students who had sex, percentage students who used a condom last time had sex and percentage of students who had sex before 13 years of age before and after the cutoff point for high and low implementers which is decided based on the state index scores. Looking first at figure 4-1 where we plot state level percentage of students who had sex against year, we see some evidence of a negative effects of the treatment on both high and low implementers defined using component one as state index score. The slopes for both low and high implementers become more positive post 1998 implying the negative effects of 1996 reform on students' risky behaviors. In addition, the difference in trends between high and low implementers before 1998 is much bigger than post 1998 and converges toward 2007. We then aggregate state level data by years to get a clearer look at the changes in trends. Figure 4-2 shows same pattern as the pattern in figure 4-1, therefore in the rest of the figures we use aggregated level data. In figure 4-3 we use all components calculated from the PCA to identify high and low implementers, we observe slight changes in the shape of trends while the general pattern is the same. Figure 4-4 focuses on the differential trends in the percentage of condom usage between two groups. In the first of these two figures, we used component 1 only to identify two groups and one can see that the before trends are much wider relative to the post trends between two groups. In fact, the percentage of students who used condoms increases more for low implementers suggesting that abstinence-only education might reduce the probability of using condoms. The last two figures shows the differential trends in the percentage of students who had sex before 13 years of age. As was the case before, we see slopes become more positive for both low and high implementers. When using the total components to identify high and low implementers, the slope for high group in fact shifts up.

The advantage of using interrupted time series model is that we can break up the curvilinear growth trends into separate linear components and then estimate the trends difference. Table 4-10 provides results for the same model using different dependent variables. We see negative and significant effects on percentage of students who had sexual intercourse as well as the percentage of students who had sex before 13% for the high group post 1998, relative to the low group. Specifically, the percentage of students who had sex significantly decreased 0.5% and the percentage of students who had sex before 13 decreased around 0.5%, both at 5% significance level, for high-implement states after 1998, relative to low implementers.

One should also notice that ITS design controls for baseline level and trend when estimating expected changes due to the intervention. The regression method assumes linear trends over time, and the adolescent risky behavior data, in particular, had a poor fit, resulting in large standard errors in the post-intervention period. Therefore, the negative impacts of sex education policies post 1996 reform we observe could be because of the linearity and non-negativity assumptions. In addition, to conduct the ITS analysis, we aggregated data to state level with three data points before and after the intervention, respectively. The limited number of observations may significantly affect the power of ITS design which may be the reason why we could not detect any statistical significant effects for some of our analysis. Last, time series is strongest when the intervention produces an immediate effect. However, in our case, the sex education reform may present delayed effects and hence cannot be captured by our model which again may underestimate the treatment effects.

Conclusion

Our study provides evidence on the causal links between 1996 Title V 510 and students' sexual behaviors. However, we did not detect any statistically significant difference between state level comprehensive sex education and abstinence sex education mandates. One might be worried that school policies vary considerably and therefore we will not identify the effects of state sex education policies. However, we employ a rich dataset which allows us to control for individual characteristics, school characteristics and state covariates to the most extent and then identify the effects of sex mandate at state level. Overall, we do not find that abstinence-only and comprehensive sex education decrease the probability of being sexually active or increase the likelihood of performing safe sex. Instead, we find that abstinence-only lower the probability of using condoms and birth control pills relative to not using any birth control method.

We then use a unique identification strategy that allows us to create two groups based on how strict states implement their sex education policies and then test the effects of the differential trends in students' risky behaviors due to 1996 reform. We find that the trend in percentage of students who had sex (percentage of students who had sex before 13) decreases by 0.5% (0.5%) in high-implement states relative to low-implement states post 1996 reform.

Searching for an effective sex education policy which will reduce the incidence and prevalence of STD and HIV/AID and teen pregnancy rate has been the main goal for many policy makers, school administrators and parents for decades. However, credible studies on such issue are limited. As the first study evaluating the differential effects of 1996 reform at state level, our results indicate that a heterogeneous trend exists between high and low groups. With the implementation of 1996 reform, however, the decreasing/increasing trends in the likelihood of having sex (performing safer sex) slows down implying that such policies may result in negative effects on adolescent risky behaviors overall.

Table 4-1. Predicted Effects of Different types of Sex Education on the Probability of becoming sexually active and Condom Usage

Utility Ordering	Impact on	Risk-Altering* (Comprehensive)	Utility-Altering** (Abstinence-only)		Risk-Revealing*** (Comprehensive)	
			Case 1	Case 2	Case 1	Case 2
	(1)	(2)	(3)	(4)	(5)	(6)
V>U(np|k*)	P(a=1|k=c)	+	0 or +	0 or -	?	?
	P(a=1|k=nc)	-	0 or +	0 or -	-	+
	P(a=0|k*)	0	0 or -	0	?	?
V<U(np|k*)	P(a=1|k=c)	+	+	0	+	-
	P(a=1|k=nc)	?	+	0 or -	-	+
	P(a=0|k*)	?	-	0 or +	?	?
Overall	P(a=1|k=c)	+	+	0 or -	?	?
	P(a=1|k=nc)	?	+	0 or -	-	+
	P(a=0|k*)	?	-	0 or +	?	?
All	P(k=c)	?	0	0	+	-
	P(K=other|nc, a=1)	+	0	0	+	-
	P(n>1|k*,a=1)	-	+	-	-	+

* Risk-altering sex education reduces the P(p|k*)
**Case 1 utility-altering sex education increases U(np|k*) relative to V;
case 2 utility-altering sex education decreases U(np|k*) relative to V.
***Case 1 risk-revealing sex education increases P(p|k*)
case 2 risk-revealing sex education decreases P(p|k*)

Table 4-2. Summary Statistics of All Variables

Variables	***Mean (Standard errors)***	***Definition/Questions from YRBSS***
If had sex	0.466 (0.007)	Dichotomous variable that equals 1 if had sexual intercourse
If used condom	0.591 (0.004)	Dichotomous variable that equals 1 if used a condom the last time had sex.
Age of first time sex	14.187 (0.020)	How old were you when you had sexual intercourse for the first time?
Number of partners	1.43 (0.023)	During your life, with how many people have you had sexual intercourse?
Female	0.485 (0.004)	Dichotomous variable that equals 1 if respondent is Female
Asian	0.032 (0.003)	Dichotomous variable that equals 1 if respondent is Asian
African American	0.138 (0.006)	Dichotomous variable that equals 1 if respondent is African American
Hispanic	0.100 (0.005)	Dichotomous variable that equals 1 if respondent is Hispanic
White	0.643 (0.010)	Dichotomous variable that equals 1 if respondent is White
Grade 9	0.241 (0.428)	Dichotomous variable that equals 1 if respondent is in Grade 9
Grade 10	0.246 (0.431)	Dichotomous variable that equals 1 if respondent is in Grade 10
Grade 11	0.252 (0.434)	Dichotomous variable that equals 1 if respondent is in Grade 11
Grade 12	0.258 (0.437)	Dichotomous variable that equals 1 if respondent is in Grade 12
# Hours Watching TV per day	4.06 (0.027)	Number of hours a respondent watches TV on an average school day
Smoke regularly	0.214 (0.004)	Dichotomous variable that equals 1 if respondent smoked cigarettes regularly
# days drink of alcohol	1.74 (1.076)	Number of days a respondent had at least one drink of alcohol
# of times used marijuana	1.709 (0.022)	Number of times a respondent used marijuana
# Days have PE Class	3.170 (0.048)	Number of times a respondent exercises or participates in physical education class
# of times carry a weapon	0.559 (0.011)	Number of times a respondent carry a weapon such as a gun, knife.

Table 4-2. Continued

# Days unsafe at school	1.099 (0.004)	Number of days a respondent did not go to school because you felt you would be unsafe at school or on your way to or from school
State Level Variables		
% had comprehensive sex education	0.362 (0.036)	Percentage of students who had comprehensive sex education classes.
% had abstinence sex education	0.131 (0.025)	Percentage of students who had abstinence sex education classes.
% have no sex mandate	0.246 (0.033)	Percentage of students residing in states with no sex education mandate.
% had both HIV and SEX education	0.348 (0.038)	Percentage of students residing in states with both sex education and HIV education mandate.
% only had HIV education	0.398 (0.038)	Percentage of students residing in states with HIV education mandate.
% only had SEX education	0.007 (0.006)	Percentage of students residing in states with sex education mandate.
State high school Dropout rate	15.23 (7.07)	High school dropout rate in respondent's state of residence
Total Crimes	4.042 (1.74)	Sum of violent and property crimes per 1,000 total population residing in the area.
%Adult with Bachelor Degree	24.712 (0.272)	Percentage of adults with bachelor degree in respondent's state of residence
% Family below poverty	12.53 (0.011)	Percentage of family below poverty line in respondent's state of residence
State Divorce rate	72.9 (0.188)	State divorce rate in respondent's state of residence
No. of Obs.	105,724	

Notes: *Summary statistics and standard deviation (in parentheses) from authors' calculations from the 1991-2005 Youth Risk Behavior Surveillance System.
* Africa American and Hispanics are oversampled in the YRBSS, results above are weighted means non-weighted means are similar for all variables except for the proportion of blacks and Hispanics

Table 4-3. Estimated Effects of State Sex Education on adolescent Pr(sexual intercourse)

	Dependent Variable: 1(Sexual Intercourse)						
	Model	(1)	(2)	(3)	(4)	(5)	(6)
Panel A	Abstinence SexEd	-0.1229**	-0.0717	0.0821	-0.1751***	-0.1903***	-0.0380
		(0.0375)	(0.0716)	(0.0867)	(0.0382)	(0.0397)	(0.1822)
	Comprehensive SexEd	0.1162***	0.2267***	0.1451*	0.1699***	0.1551***	0.1158
		(0.0274)	(0.0371)	(0.0691)	(0.0292)	(0.0302)	(0.2224)
Panel B	Sex Education and HIVMandate	-0.0178	-0.0460	0.1709*	-0.1379**	-0.1486**	-0.6029
		(0.0448)	(0.0457)	(0.0756)	(0.0507)	(0.0530)	(0.4933)
	HIV mandate only	-0.0765	-0.1023	0.0296	-0.0535	-0.0305	0.4364
		(0.0564)	(0.0582)	(0.0844)	(0.0650)	(0.0705)	(0.3015)
	Sex Education Mandate only	-0.1157	-0.1439	0.0326	-0.1417	-0.1767	-0.7200
		(0.1707)	(0.1768)	(0.2026)	(0.1733)	(0.1801)	(0.5594)
	School covariates	NO	YES	YES	NO	YES	YES
	State covariates	NO	NO	YES	NO	NO	YES
	State fixed effects	NO	NO	NO	YES	YES	YES
	Year fixed effects	NO	NO	YES	YES	YES	YES
	Sample size	61354	40023	27184	61354	59523	27184

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05;
* Standard errors are cluster-corrected by state.
* All models include controls for individual race, gender and age.
* School covariates include number of times a respondent carry a weapon such as a gun, knife, or club and number of days a respondent did not go to school because you felt you would be unsafe at school or on your way to or from school
* State covariates include state dropout rate, percentage of adults with bachelor degree, Percentage of family below poverty line and state divorce rate

Table 4-4. Estimated Effects of State Sex Education on Performing Safe sex (using condoms)

Dependent Variable: 1(Used a Condom)						
Model	(1)	(2)	(3)	(4)	(5)	(6)
Abstinence SexEd	-0.0888* (0.0438)	-0.0686 (0.0448)	-0.0490 (0.0921)	0.0136 (0.0523)	0.0379 (0.0535)	-0.5133* (0.2470)
Comprehensive SexEd	0.0618 (0.0336)	0.0606 (0.0343)	-0.0744 (0.0701)	0.0135 (0.0393)	0.0131 (0.0401)	0.3096 (0.3142)
Sex Education and HIV Mandate	-0.0698 (0.0500)	-0.0433 (0.0496)	0.1389 (0.0783)	-0.0087 (0.0678)	0.0209 (0.0692)	0.3298 (0.3140)
HIV mandate only	-0.0938 (0.0587)	-0.0640 (0.0585)	0.0608 (0.0882)	-0.1308 (0.0922)	-0.1197 (0.0939)	0.3288 (0.5233)
Sex Education Mandate only	-0.2124 (0.2329)	-0.1784 (0.2355)	-0.1829 (0.2544)	-0.1031 (0.2480)	-0.0552 (0.2510)	-
School covariates	NO	YES	YES	NO	YES	YES
State covariates	NO	YES	YES	NO	YES	YES
State fixed effects	NO	NO	No	NO	NO	YES
Year fixed effects	NO	NO	YES	YES	YES	YES
Sample size	32341	31207	14371	32341	31207	14371

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05;
* Standard errors are cluster-corrected by state.
* All models include controls for individual race, gender and age.
* Additional controls are listed in the notes to Table3

Table 4-5. Estimated Effects of State Policies on Pr(Sexual intercourse) and Pr(Condom Usage)

Dependent Variable: 1(Sexual intercourse)			
Model	(4)	(5)	(6)
Program on Pregnancy	-0.4849***	-0.4849**	-0.1809
	(0.0875)	(0.1498)	(0.1815)
Program on STD	0.7263***	0.7264**	-
	(0.0790)	(0.2474)	
Program on Sex education	-0.1220	-0.1221	0.2939
	(0.1049)	(0.1529)	(0.2530)
Dependent Variable: 1(Used a Condom)			
Program on Pregnancy	-0.2223	-0.2223	-0.1688
	(0.1236)	(0.1928)	(0.2484)
Program on STD	0.5362**	0.5362	-
	(0.1955)	(0.3208)	
Program on Sex education	-0.2675	-0.2675	-0.2428
	(0.1902)	(0.2192)	(0.3709)
School covariates	NO	YES	YES
State covariates	NO	YES	YES
State fixed effects	NO	NO	YES
Year fixed effects	YES	YES	YES

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05;
* Standard errors are cluster-corrected by state.
* All models include controls for individual race, gender and age.
* Additional controls are listed in the notes to Table 3

Table 4-6. Estimated Effects of State Sex Education on Birth Control Method

Dependent Variable: Categorical variable-Birth control method				
Model		(4)	(5)	(6)
Birth Control Pill	Abstinence	0.1198	0.1526	-1.2025*
		(0.1002)	(0.1020)	(0.5947)
	Comprehensive	0.1375	0.1345	-0.5001
		(0.0740)	(0.0752)	(0.5403)
Condom	Abstinence	0.0479	0.0925	-0.8606*
		(0.0696)	(0.0714)	(0.3851)
	Comprehensive	0.0152	0.0213	0.0683
		(0.0531)	(0.0543)	(0.4981)
Depo-Provera	Abstinence	0.6503***	0.6827***	-0.1732
		(0.1219)	(0.1240)	(0.5651)
	Comprehensive	0.4872***	0.4804***	0.6120
		(0.1052)	(0.1066)	(0.8600)
Withdrawal	Abstinence	-0.2909*	-0.2490*	-0.3874
		(0.1145)	(0.1158)	(0.5393)
	Comprehensive	-0.1211	-0.1271	-0.4412
		(0.0791)	(0.0803)	(0.8900)
	School covariates	NO	YES	YES
	State covariates	NO	YES	YES
	State fixed effects	NO	NO	YES
	Year fixed effects	YES	YES	YES

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05;
* Standard errors are cluster-corrected by state.
* All models include controls for individual race, gender and age.
* Additional controls are listed in the notes to Table 3

Table 4-7. Estimated Effects of State Sex Education on Birth Control Method

Dependent Variable: Categorical variable-Birth control method

Model		(4)	(5)	(6)
Birth Control Pill	Sex Education and HIV Mandate	-0.0973	-0.0584	-0.4646
		(0.1287)	(0.1310)	(0.5401)
	HIV mandate only	-0.3824*	-0.3927*	-0.0965
		(0.1714)	(0.1741)	(0.7140)
	Sex Education Mandate only	0.3823	0.3937	-
		(0.4287)	(0.4296)	
Condom	Sex Education and HIV Mandate	-0.0621	-0.0222	0.1037
		(0.0887)	(0.0906)	(0.4979)
	HIV mandate only	-0.2960*	-0.2992*	0.1310
		(0.1301)	(0.1325)	(0.5835)
	Sex Education Mandate only	0.0941	0.1226	-
		(0.4049)	(0.4073)	
Depo-Provera	Sex Education and HIV Mandate	0.5869***	0.6125***	0.5962
		(0.1493)	(0.1520)	(0.8589)
	HIV mandate only	0.8257***	0.8015***	18.4641
		(0.2236)	(0.2270)	(298.5175)
	Sex Education Mandate only	0.1130	0.1040	-
		(0.7237)	(0.7239)	
Withdrawal	Sex Education and HIV Mandate	- 0.7041***	- 0.6880***	-0.4240
		(0.1550)	(0.1571)	(0.8897)
	HIV mandate only	-0.4810*	-0.5181*	-0.2508
		(0.2006)	(0.2032)	(0.7699)
	Sex Education Mandate only	-0.0840	-0.3033	-
		(0.6298)	(0.6686)	
	School covariates	NO	YES	YES
	State covariates	NO	YES	YES
	State fixed effects	NO	NO	YES
	Year fixed effects	YES	YES	YES

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression

* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05;

* Standard errors are cluster-corrected by state.

* All models include controls for individual race, gender and age.

* Additional controls are listed in the notes to Table 3

Table 4-8. Estimated Effects of State Sex Education on number of sex partners

Dependent Variable: number of sex partners in respondent's life time

Model	Full Sample	Female	Male
Abstinence	-0.1803	0.0334	-0.4957*
	(0.1560)	(0.2136)	(0.2311)
Comprehensive	0.0168	-0.2276	0.2639
	(0.2013)	(0.2897)	(0.2811)
Sex Education and	-0.6395	-0.8171	0.2827
HIVMandate	(0.4318)	(0.6255)	(0.2810)
HIV mandate only	0.3005	0.2323	0.0812
	(0.2642)	(0.3947)	(0.6448)
School covariates	YES	YES	YES
State covariates	YES	YES	YES
State fixed effects	YES	YES	YES
Year fixed effects	YES	YES	YES

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05;
* Standard errors are cluster-corrected by state.
* All models include controls for individual race, gender and age.
* Additional controls are listed in the notes to Table 3.

Table 4-9. Estimated Effects of State Sex Education on the age of first time had sex

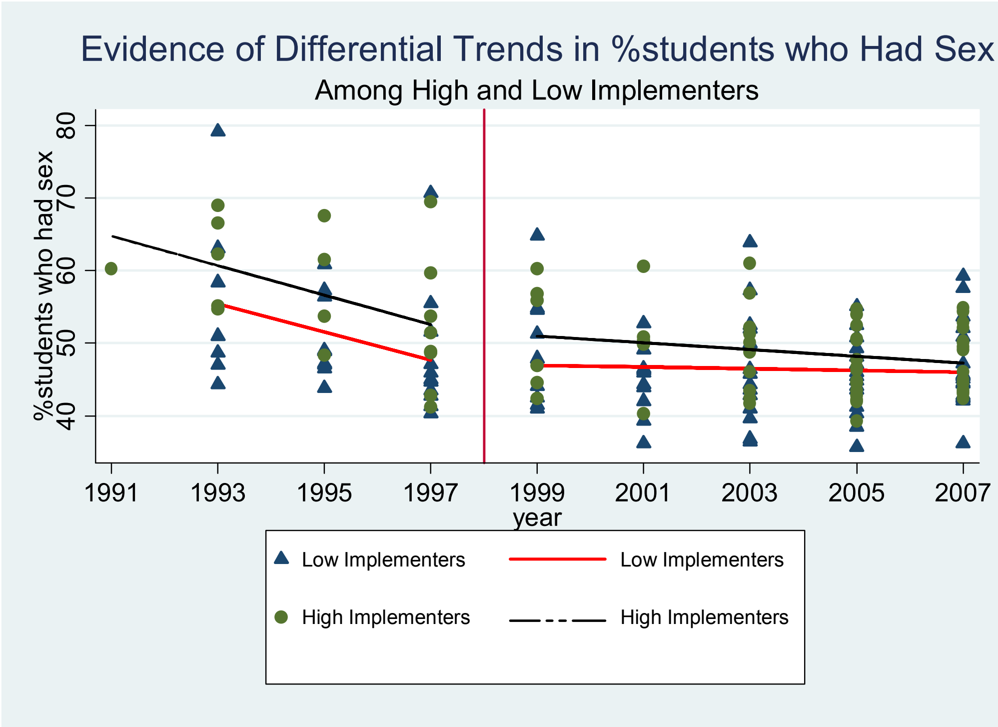
Dependent Variable: age of first time sex		-	-
Model	Full Sample	Female	Male
Abstinence	0.0057	-0.0422	0.0892
	(0.1936)	(0.2736)	(0.2771)
Comprehensive	0.0263	-0.3267	0.0673
	(0.2664)	(0.3811)	(0.3697)
Sex Education and	0.0260	-1.7415*	0.0632
HIVMandate	(0.2663)	(0.8469)	(0.3695)
HIV mandate only	0.7592	-1.1090*	0.5134
	(0.4386)	(0.5258)	(0.8204)
School covariates	YES	YES	YES
State covariates	YES	YES	YES
State fixed effects	YES	YES	YES
Year fixed effects	YES	YES	YES

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression.
* Statistical significance level: ***: p<.001; **:p<.01; *:p<.05;
* Standard errors are cluster-corrected by state.
* All models include controls for individual race, gender and age.
* Additional controls are listed in the notes to Table 3

Table 4-10. Estimated Effects of State Sex Education using ITS design

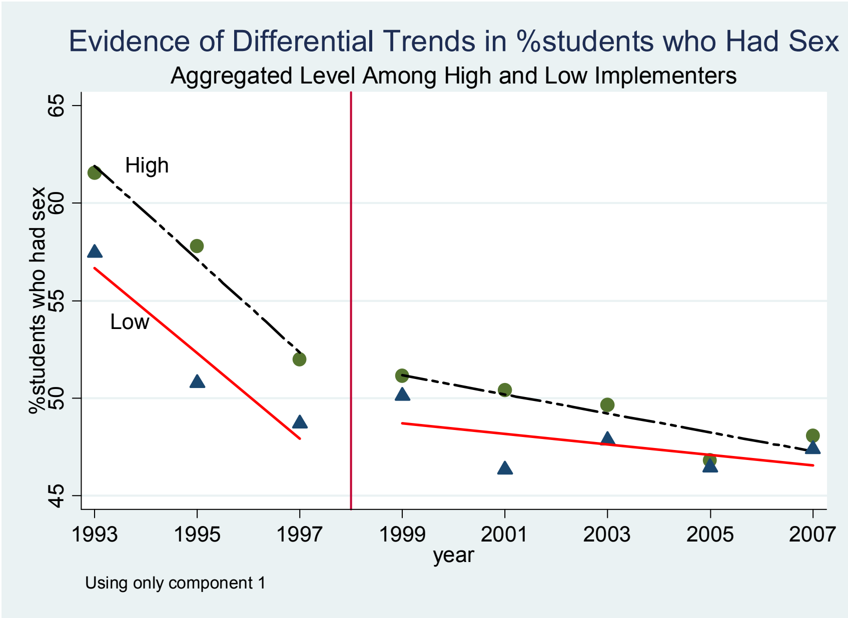
Dependent Variable:	%had sex		%had sex before 13	
Model		%used a condom		#of sex partners
Pre-Trend (Before Reform)	-1.2982** (-0.4672)	0.541 (-0.5573)	-0.5023 (-0.4137)	-0.9128*** (-0.2321)
Post-Trend (After Reform)	0.1509 (-0.4863)	1.2014*** (-0.2666)	0.2265 (-0.3955)	-0.2128 (-0.1744)
High Implementers	-10.2151 (-7.1281)	-16.0338 (-8.261)	-1.5822 (-7.114)	3.7087 (-2.6642)
Pre-Trend*High Implementers	0.1292 (-0.5326)	0.4061 (-0.6917)	-0.2596 (-0.5485)	-0.0992 (-0.4196)
Post-Trend*High Implementers	-0.5054* (-0.2678)	-0.2574 (-0.3441)	-0.4693* (-0.2532)	-0.2817 (-0.2341)
R^2	0.929	0.87	0.911	0.89
Sample size	131	137	137	138

Notes: * Each column of the table presents coefficients and standard errors in parentheses from a different regression.
* Statistical significance level: ****: p<.001; ***:p<.01; **:p<.05; *:P<0.1
* Standard errors are cluster-corrected by state.
* All models include controls for individual race, gender and age.
* Additional controls are listed in the notes to Table 3



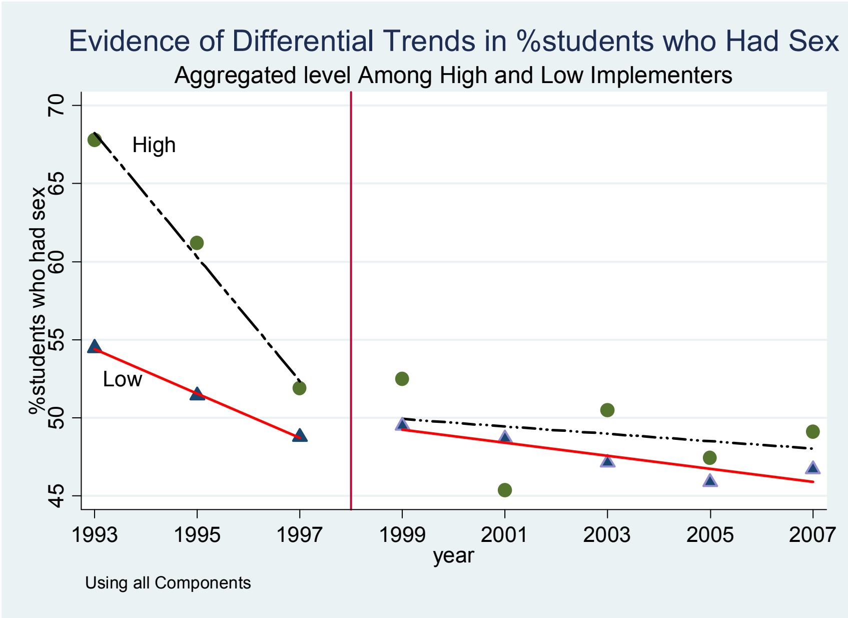
Define High Implementation> $Index_{comp1}$

Figure 4-1A. Percentage of students who had sexual intercourse Pre-Post 1996 Reform (State level)



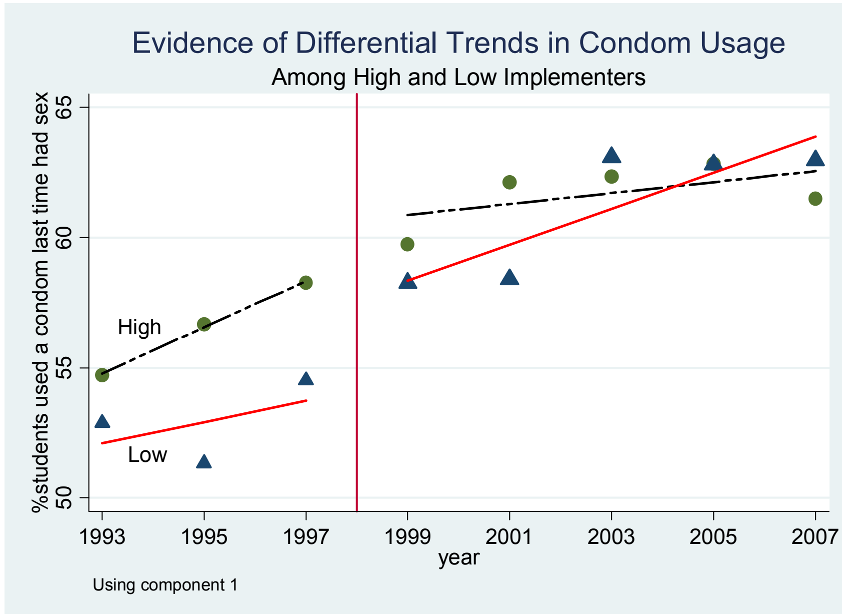
Define High Implementation> $Index_{comp1}$

Figure 4-2 Percentage of students who had sexual intercourse Pre-Post 1996 Reform (Aggregate level)



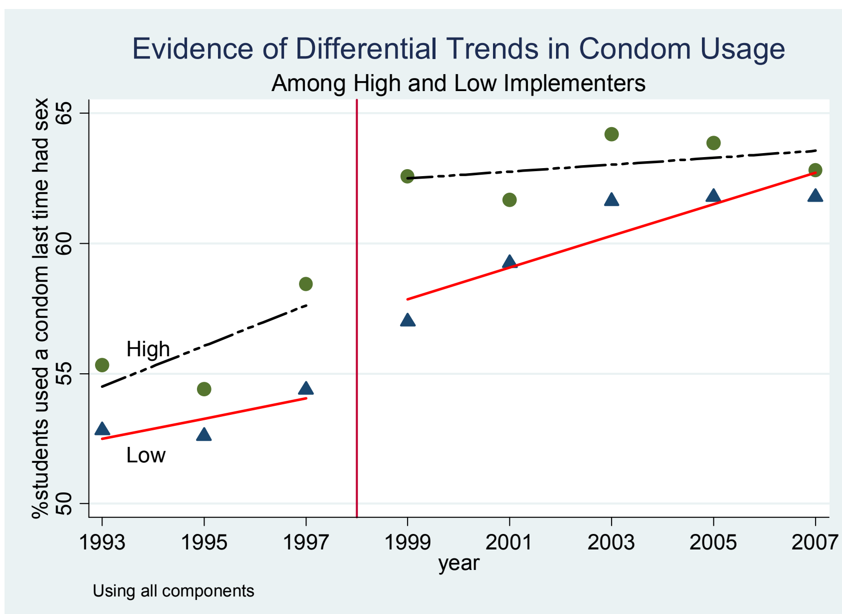
Define High Implementation> $Index_{compTotal}$

Figure 4-3 Percentage of students who had sexual intercourse Pre-Post 1996 Reform (Aggregate level)



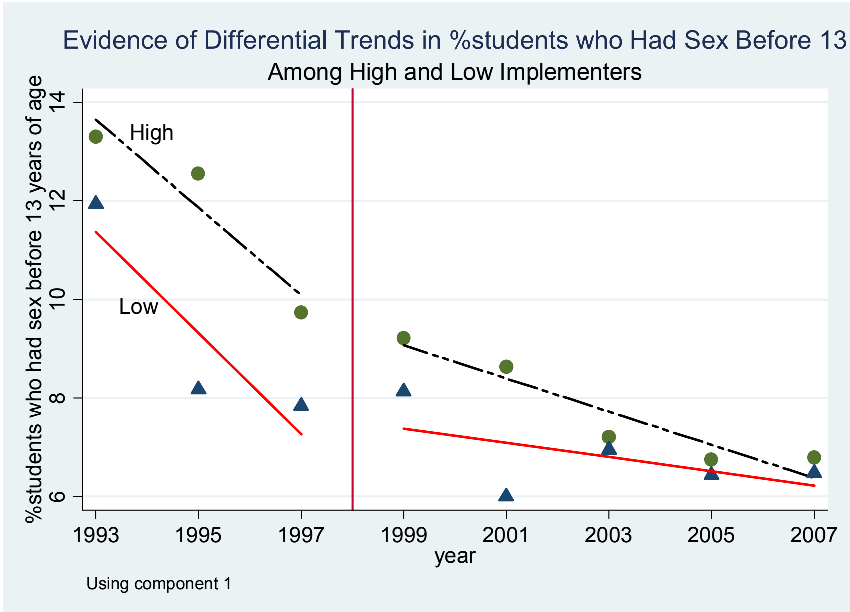
Define High Implementation> $Index_{comp1}$

Figure 4-4 Percentage of students used condoms last time had sex Pre-Post 1996 Reform (Aggregate level).



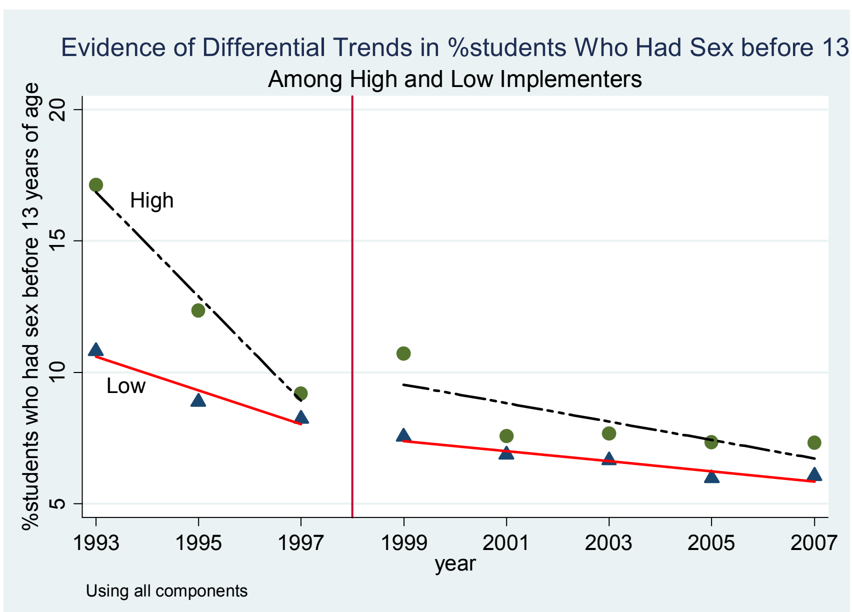
Define High Implementation> $Index_{compTotal}$

Figure 4-5 Percentage of students used condoms last time had sex Pre-Post 1996 Reform (Aggregate level).



Define High Implementation> Index_{comp1}

Figure 4-6 Percentage of students who had sex before 13 years of age Pre-Post 1996 Reform (Aggregate level)



Define High Implementation> $\text{Index}_{compTotal}$

Figure 4-7 Percentage of students who had sex before 13 years of age Pre-Post 1996 Reform (Aggregate level).

CHAPTER 5
CONCLUSION

The first two essays in this dissertation examine the causal relation between education policies and childhood obesity. The first study contributes to the sparse literature on adolescent obesity by identifying school accountability system as a strong factor that may contribute to higher adolescent BMI and the probability of being overweight. We construct three different measures to evaluate state school accountability over the years between 1999 and 2005. According to our estimates, all measures present statistically significant and positive effects on student BMI and the probability of being overweight. Specifically, an additional year of exposure to school accountability systems will lead to an increase in BMI by around 0.06 kg/m^2 and this effect is bigger among females, and among Asian and White students, as compared to their counterparts. One more year exposure to school accountability also increases the probability of being overweight by around 0.5 percentage points and this effect is slightly bigger among males and Hispanic students. Second, our empirical results show that school accountability systems present significant lag effects on adolescent BMI and the probability of being overweight. It appears that students need an adjustment period before the full response to changes of school environmental factor. Last, school accountability significantly decreases the number of times female students participate in PE classes.

The purpose of the second study is to complement and extend the childhood obesity literature to determine if teacher level grading standards had any effects on child BMI and the probability of being overweight. More specifically, we study teacher grading standards across two subjects, math and reading, and analyze how these relate to children’s weight outcomes. We find that, conditional on a large set of individuals’ observable characteristics including demographic information, family background, school characteristics, and teacher characteristics, students under higher teacher grading standards across subjects have higher BMI and are more likely to be overweight. Moreover, this effect is stronger for female and White students. These results suggest that while high grading standard may help improve student academic performance, it may at the same time harm children’s health, in this case children’s weight status. Throughout the paper, results consistently indicate that students’ health respond uniquely to teacher grading standards. This suggests that school teachers should carefully give thought to the grading standards when they are considering how these standards can be improved to promote academic performances. We certainly do not want to improve students’ academic performances at the expense of their health. As childhood obesity has become one of the biggest health concerns in this nation, teachers, school and parents should realize that school environment is one of the main factors that may associate with this weight gain. More research is needed in this literature to examine the unintended impacts of other education policies on children’s health outcomes. Policy makers should keep in mind that a single incentive policy that is designed to promote academic performance may unintentionally harm children’s health as teachers and schools pursuing higher test score gains may undertake unwanted actions that post irreversible impact on children.

The last study provides evidence on the causal links between 1996 Title V 510 and students’ sexual behaviors. However, we did not detect any statistically significant difference between state level comprehensive sex education and abstinence sex education mandates. One might be worried that school policies vary considerably and therefore we will not identify the effects of state sex education policies. However, we employ a rich dataset which allows us to control for individual characteristics, school characteristics and state covariates to the most extent and then identify the effects of sex mandate at state level. Overall, we do not find that abstinence-only and comprehensive sex education decrease the probability of being sexually active or increase the likelihood of performing safe sex. Instead, we find that abstinence-only lower the probability of using condoms and birth control pills relative to not using any birth control method.

We then use a unique identification strategy that allows us to create two groups based on how strict states implement their sex education policies and then test the effects of the differential trends in students' risky behaviors due to 1996 reform. We find that percentage of students who had sex (percentage of students who had sex before 13) decreases by 0.5% (0.5%) in high-implement states relative to low-implement states post 1996 reform.

Searching for an effective sex education policy which will reduce the incidence and prevalence of STD and HIV/AID and teen pregnancy rate has been the main goal for many policy makers, school administrators and parents for decades. However, credible studies on such issue are limited. As the first study evaluating the differential effects of 1996 reform at state level, our results indicate that heterogeneous effects exist between high and low groups. With the implementation of 1996 reform, however, the decreasing/increasing trends in the likelihood of having sex (performing safer sex) slows down implying that such policies may result in negative effects on adolescent risky behaviors overall.

APPENDIX
VARIABLES USED IN THE PRINCIPAL COMPONENT ANALYSIS

Variable	Variable
Who is the Primary authority that control over the program	Provided Grants to community-based organizations
If the state has a Advisory panel	Grants_ education programs
Number of grants granted to the community-based programs	Grants_recreational programs
Intended audience	Grants_mentoring programs
Less than 10 years old	Grants_life option or career planning programs
10 to 14 years old	Grants_motivational speaker
15 to 17 years old	Grants_less than 14
18 to 19 years old	If provide grants to education agencies
If state conduct media campaign	Education_state funding classroom
Status of campaign: continue an old program/developed new program	Education_Continued existing abstinence program
Media campaign initiated year	Education_introduced new abstinence program
Media_using radio programs	Education_parents optout choice
Media_using billboard	statefunds_provide afterschool programs
Media_using newspaperads	Afterschool_education programs
Media_using TV PSAs	Afterschool_recreational programs
Media_using posters	Afterschool_tutoring remedial edu
Media_using pamphlets	Afterschool_community service
Media_using Paid TV Radio	Afterschool_mentoring programs
Organization that responsible for media campaign:	Programs in Elementary schools
Health organization	Programs in Junior High
Education organization	Programs in High school
Social service	abstinence coordinated with state govt. initiatives
Faith-based	with private efforts

LIST OF REFERENCES

Anderson, P.M, Butcher, K.F., Levine, P.B. 2003. Maternal employment and overweight children. Journal of Health Economics 22(3), 477-504

Anderson, P.M., Butcher, K.F. 2005. Reading, Writing, and Raisinettes: Are school finances contributing to children's obesity? NBER Working Paper No. 11177

Anderson, P.M., Butcher, K.F., Schanzenbach, D.W. 2009. The effect of school accountability on children's health. Working Paper.

Bagues, M.,Labini, M.S., Zinovyeva, N. 2008. Differential Grading Standards and University Funding: Evidence from Italy, CESifo Economic Studies 54(2), 149-176.

Bar-Or, O., Foreyt, J., Bouchard, C., Brownell, K. D., Dietz, W. H., Ravussin, E., Salbe, A. D., Schwenger, S., St. Jeor, S., Torun, B. 1998. Physical activity, genetic, and nutritional considerations in childhood weight management. Medicine & Science in Sports & Exercise 30(1), 2-10.

Becker, W., Rosen, S. 1990. The Learning Effect of Assessment of Evaluation in High School. Discussion paper 90-7, Economics Research Center, NORC.

Betts, J. 1995. Do Grading Standards Affect the Incentive to Learn? Working paper, University of California-San Diego.

Betts, J. 1998. The Impact of Educational Standards on the Level and Distribution of Earnings. American Economic Review 266-275.

Betts, J., Grogger, J. 2000. The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings. NBER working paper 7875, September.

Bokhari, F.A.S. Scheneider, H. 2009. School accountability laws and the consumption of psychostimulants. Working Paper, wp2009_03_02, Repec.

Brener N.D., McManus T., Galuska D.A., Lowry, R., Wechsler, H. 2003. Reliability and validity of self-reported height and weight among high school students. Journal of Adolescent Health 32(4), 281–287.

Cawley, J., Meyerhoefer, C., Newhouse, D. 2006. The impact of state physical education requirements on youth physical activity and overweight. Health Economics 16(12), 1287-1301.

Cawley, J., Meyerhoefer, C.D., Newhouse, D.L. 2007. The correlation of youth physical activity with state policies. Contemporary Economic Policy 25(4), 506-517.

Centers for Disease Control and Prevention, National Center for Health Statistics. CDC growth charts: United States. http://www.cdc.gov/growthcharts/ May 30, 2000

Chakrabarti, R. 2006. Do public schools facing vouchers behave strategically? Evidence from Florida. Nashville, TN: National Research and Development Center on School Choice.

Chou, S.Y., Grossman, M., Saffer, H. 2004. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. Journal of Health Economics 23(3), 565-587.

Costrell, R. 1994. A Simple Model of Educational Standards. American Economic Review 956-971.

Cutler, E.L., Glaeser Shapiro, J. M. 2003. Why have Americans become more obese? NBER Working Paper No. 9446.

Dietz ,W.H. 1994. Critical periods in childhood for the development of obesity. Am J Clin Nutr. 59(5), 955-959.

Doniger, A. S., Riley, J. S., Utter, C. A., Adams, E. 2001. Impact evaluation of the "Not Me, Not Now" abstinence-oriented, adolescent pregnancy Prevention Communications Program, Monroe County, NY. Journal of Health Communication 6(1), 45-60.

Eisenmann, J.C., Bartee, R.T., Wang, M.Q. 2002. Physical Activity, TV Viewing, and Weight in U.S. Youth 1999 Youth Risk Behavior Survey. Obesity Research 10(5), 379-385.

Figlio, D. N., Rouse, C.E. 2006. Do accountability and voucher threats improve low-performing schools? Journal of Public Economics 90(1-2), 239-255.

Figlio, D., Lucas, M. 2004. Do High Grading Standards Affect Student Performance? Journal of Public Economics 88(9-10), 1815-1834.

Figlio, D.N., Winicki, J. 2005. Food for thought: the effects of school accountability plans on school nutrition. Journal of Public Economics 89(2-3), 381-394.

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., Dietz, W. 2009. Annual medical spending attributable to obesity: Payer-and service-specific estimates. Health Affairs 28(5), w822-w831.

Freedman, D.S., Khan, L.K., Dietz, W.H., Srinivasan, S.R., Berenson, G.S. 2001. Relationship of Childhood Obesity to Coronary Heart Disease Risk Factors in Adulthood: The Bogalusa Heart Study. Pediatrics 108(3), 712-718.

Friedman, M.A., Brownell, K.D. 1995. Psychological correlates of obesity: moving to the next research generation. Psychological Bulletin 117, 3–20.

Goodman, E., Whitaker, R.C. 2002. Prospective study of the role of depression in the development and persistence of adolescent obesity. Pediatrics 109, 497-504.

Guo, S.S., Wu, W., Chumlea, W.C., Roche, A.F. 2002. Predicting overweight and obesity in adulthood from body mass index values in childhood and adolescence. American Journal of Clinical Nutrition 76(3), 653-658.

Haney, W. 2000. The myth of the Texas miracle in education. Education Policy Analysis Archives 8(41). Retrieved December 20, 2007, from http://epaa.asu.edu/epaa/v8n41

Haney, W. 2002. Ensuring failure: How a state's achievement test may be designed to do just that. Education Week 21(42), 546-58.

Jussim, L. 1989. Teacher Expectations: Self-fulfilling Prophecies, perceptual biases, and accuracy. Journal of Personality and Social Psychology 57(3), 469-480.

Jussim, L. 1991. Social perception and social reality: A reflection-construction model. Psychological Review 98, 54–73.

Katzmarzyk, P.T., Malina, R.M., Bouchard, C. 1999. Physical activity, physical fitness, and coronary heart disease risk factors in youth: The Québec Family Study. Preventive Medicine 29, 555-562.

Koretz, D., Barron, S. I. 1998. The Validity of Gains on the Kentucky Instructional Results Information System (KIRIS). MR-1014-EDU, Santa Monica: RAND.

Ku, L., Sonenstein F.L., Pleck, J.H. 1993. Young men's risk behaviors for HIV infection and sexually transmitted diseases, 1988 through 1991. American Journal of Public Health 83(11), 1609-1615.

Ku, L., Sonenstein F.L., Pleck, J.H. 1992. Factors influencing first intercourse for teenage men. Public Health Report, 108, 680-694

Lakdawalla, D., Philipson, T. 2002. The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination. NBER Working Papers No. 8946.

Lakka, H.M., Laaksonen, D.E., Lakka, T.A., Niskanen, L.K., Kumpusalo, E., Tuomilehto, J., Salonen, J.T. 2002. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. The Journal of the American Medicine Association 288(21), 2709-2716.

Lerner, R. 2004. Can Abstinence Work? An Analysis of the Best Friends Program. Adolescent and Family Health 3(4), 185-192.

Lillard, D.R., DeCicca, P. 2001. Higher Standards, More Dropouts? Evidence Within and Across Time, Economics of Education Review 20(5), 459-473.

Locard, E., Mamelle, N,, Billette, A., Miginiac, M., Munoz, F., Rey, S, 1992. Risk factors of obesity in a five year old population. Parental versus environmental factors. Int J Obes Relat Metab Disord 16(10), 721-729.

Mokdad, A.H., Ford, E.S., Bowman, B.A., Dietz, W.H., Vinicor, F., Bales, V.S., Marks, J.S. 2003. Prevalence of obesity, diabetes, and obesity-related health risk factors." Journal of the American Medical Association 289, 76-79.

Must, A., Spadano, J., Coakley, E.H., Field, A.E., Colditz, G., Dietz, W.H. 1999. The disease burden associated with overweight and obesity. The Journal of the American Medical Association 282(16), 1523-1529.

Oettinger, G.S. 1999. The effects of sex education on teen sexual activity and teen pregnancy. Journal of Political Economy 107(3), 606-644.

Rouse, C.E., Hannaway, J., Goldhaber, D., Figlio, D.N. 2007. Feeling the Heat: How Low Performing Schools Respond to Voucher and Accountability Pressure. NBER Working Paper No. 13681.

Schanzenbach, D.W. Forthcoming. Does the Federal School Lunch Program Contribute to Childhood Obesity? Journal of Human Resources.

Stunkard, A.J., Harris, J.R., Pedersen, N.L., McClearn, G.E. 1990. The bodymass index of twins who have been reared apart. The New England Journal of Medicine 322(21), 1483-1487.

Tremblay, C.H., Ling, D.C. 2005. AIDS education, condom demand, and the sexual activity of American youth. Health Economics 14(8), 851-867

Trenholm C, Devaney B, Fortson K, Clark, M., Quay, L., Wheeler, J.. 2008. Impacts of abstinence education on teen sexual activity, risk of pregnancy, and risk of sexually transmitted diseases. Journal of Policy Analysis and Management 27 (2), 255-276.

West, M. R., Peterson, P. E. 2006. The efficacy of choice threats within accountability systems: Results from legislatively induced experiments. The Economic Journal 116(510), c46–c62.

Woolston, J.L. 1987. Obesity in infancy and early childhood. Journal of American Academy of Child & Adolescence Psychiatry 26, 123-126.

Yin, Lu. 2009. Are School Accountability Systems Contributing to Adolescent Obesity? Working paper, November.

BIOGRAPHICAL SKETCH

After earning her bachelor's degree in finance from Beijing Capital University of Economics and Business in China, Lu worked as an associate auditor in Pricewaterhouse Coopers for one and half years. In 2005, Lu started graduate school in economics at University of Florida. Her fields of specialization are economics of education, health economics and applied econometrics. She received her Ph.D. from the University of Florida in the summer of 2010.