The Effects of Physical Education on Student Fitness, Achievement, and Behavior

Analisa Packham *   Brittany Street†

November 2018

Abstract

Despite the mounting evidence that physical education (PE) has health and education benefits for elementary-aged children, much less is known on the effectiveness of such programs for older children. To study the effects of PE on adolescents, we analyze the impact of Texas Fitness Now (TFN), a four-year $37 million grant program that mandated daily PE for middle-school students in low-income schools. Using a regression discontinuity approach to exploit the cutoff in school eligibility, we find that daily PE mandates do not lead to overall improvements in student fitness, including cardiovascular endurance, strength, and flexibility. Although we show that program was ineffective at changing average student body composition, estimates indicate a reduction in the proportion of obese students. Using individual-level school records data, we find that PE does not lead to positive spillover effects in the classroom, including improvements in standardized test scores, or increases in attendance for 6th, 7th and 8th graders. Instead, we provide some evidence to suggest that PE reduces attendance rates and increases disciplinary incidents for middle-school students.

JEL Classification: I12, I18, I21 H75

Keywords: physical education, obesity, fitness, discipline, attendance

*Department of Economics, 800 E. High St. Oxford, OH 45056. Email: apackham@miamioh.edu.

†Department of Economics, Texas A&M University, 4228 TAMU, College Station, TX 77843. Email: brittany210@tamu.edu. This research uses confidential data from the State of Texas supplied by the Texas Education Research Center (ERC) at the University of Texas-Austin. We gratefully acknowledge the use of these data and pledge that the views expressed in this manuscript are those of the authors and not the ERC, the State of Texas, or Texas A&M University. We would like to thank Carmen Alfonzo and the Texas Education Agency for providing fitness grant and fitness testing data. We also thank the Texas A&M Department of Economics for institutional support, as well as Mark Hoekstra, Jason Lindo, Miguel Sarzosa, Manu Raghav, Patrick Reilly, Lori Taylor, participants at the 2016 and 2017 Meetings of the Southern Economic Association and 2017 Western Economics Association International meeting and seminar attendees at Ohio State University and Miami University for useful feedback on work in progress. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

1 Introduction

In the United States, the rate of childhood obesity has more than quadrupled in the past thirty years (Centers for Disease Control and Prevention 2016). One in three children are at risk of becoming overweight or obese, and among children of lower socioeconomic status, the risk is even higher (Centers for Disease Control and Prevention 2016; Let's Move 2016).

From a public policy perspective, policies that seek to target the inputs to obesity, like food and exercise, can reduce negative externalities imposed by higher health care costs in the long run (Cutler, Glaser, and Shapiro 2003; Finkelstein, Fiebelkorn, and Wang 2003; Finkelstein, Trogdon, Cohen, and Dietz 2009). Given that children between the ages of 5-18 spend approximately 40 hours a week at school and may eat several meals there, a natural policy solution to address childhood obesity and increase total social welfare is to encourage children to form healthy habits at school. The purpose of this paper is to analyze the effects of one such initiative, Texas Fitness Now, on student health and academic performance.

Due to the concern of rising health risks and costs of obesity in recent years, federal and state agencies have created new guidelines and implemented numerous programs to encourage physical activity. Recently, medical authorities including the Institute of Medicine, American Heart Association, and the American Academy of Pediatrics, have endorsed curricula that consist of at least 30 minutes of daily physical activity a day as a way to reduce obesity and overweight (Institute of Medicine 2013; Pate and O'Neill 2008; American Academy of Pediatrics 2006). Despite these recommendations, schools may not provide enough opportunities for students to meet this standard during the school day, due to resource or time constraints. Only 3.8% of elementary schools, 7.9% of middle schools, and 2.1% of high schools provide daily physical education (Centers for Disease Control and Prevention 2007).

Although physical education (PE) interventions are continually recommended by medical professionals as a strategy to increase physical activity and reduce childhood obesity, the results of such policies have been mixed. A literature review by Guerra, Nobre, da Silverira, and de Aguiar Carrazedo Taddei (2013) reports that only 1 out of 11 published studies that use randomized control trials to evaluate PE programs estimate significant reductions in body mass index (BMI). None find effects on body weight. And while a handful of studies document that increasing PE time can reduce obesity for young, elementary-school children (Centers for Disease Control and Prevention 2016; Cawley, Frisvold, and Meyerhoefer 2013; Waters, de Silva-Sanigorski, Bedford, Brown, Campbell, Gao, Armstrong, Prosser, and Summerbell 2011; Datar and Sturm 2004), there is less evidence to suggest that such programs are effective at reducing BMI for middle-school or high-school students (Cawley, Meyerhoefer, and Newhouse 2007; Wang, Yang, Lowry, and Wechsler 2003; von Hippel and Bradbury 2015; Knaus, Lechner, and Reimers 2018).

Separate from the effects on health, PE proponents argue that increasing physical activity yields large academic benefits by improving cognition, focus, and memory. There is a growing body of research implying that this may indeed be the case.[1] In a recent report, the CDC describes analyses that link school-based physical activity, including physical education, to academic behaviors such as cognitive skills, academic attitudes, attendance, and achievement, and provides suggestive evidence of a positive relationship between physical activity and academic performance. (Centers for Disease Control and Prevention 2010).[2] Moreover, studies evaluating increases PE time in schools appear to offer some affirmation that such programs can improve student outcomes (Tremarche, Robinson, and Graham 2007; Carlson, Fulton, Lee, Maynard, Brown, III, and Dietz 2008).[3]

That being said, one concern is that increasing PE requirements takes away important instructional time, which could lead to less learning and poorer student outcomes. In a review of 7 quasi-experimental studies, which focus on academic outcomes for students up to grade 6, Trudeau and Shephard (2008) finds that physical activity can be

[1]For evidence on the relatinoship between physical activity and cognition, see, for example, Tomporowsk, Davis, Miller, and Naglieri (2008).

[2]Out of the 43 studies, nearly all estimates testing the relationship between academic performance and physical activity are positive (98.5%), and approximately half are statistically significant.

[3]Specifically, Tremarche, Robinson, and Graham (2007) estimates the effects of a randomized control trial, and concludes that students in an elementary school with more PE time had higher reading test scores. Carlson, Fulton, Lee, Maynard, Brown, III, and Dietz (2008) uses Early Childhood Longitudinal Study and finds that increasing PE time raises test scores for girls.

added to school curriculum without hindering student achievement. Dills, Morgan, and Rotthoff (2011) similarly explores this hypothesis and finds no statistically significant or economically significant impact of weekly PE on test scores for elementary-aged children, suggesting that PE at worst has no effect on academic achievement.

Based on the above research, we would expect that policies targeting physical activity have the ability to positively affect student behavior and performance, implying that there may be some scope for school-level services to play an even larger role. However, nearly all of the literature to date focuses on elementary-aged children, while little to no evidence exists on the effects of PE on middle-school students. Accordingly, a fundamental policy question remains unanswered: how much can PE affect adolescent fitness and health, and how much do these programs translate to changes in attendance, disciplinary action, and academic performance?

To answer this question, we present new evidence on the effects of physical education requirements and contribute to a growing literature on how policies can address childhood obesity and student achievement. In particular, we estimate a model that exploits a discontinuity in eligibility criteria for Texas Fitness Now (TFN), a four-year physical education grant program targeting low-income students with the aim of improving overall health and well-being. Program-eligible schools included campuses teaching grades 6, 7, and/or 8 with a large majority of economically disadvantaged students. Participating schools received funds contingent on the agreement that they: (i) spent funds on new athletic equipment or services related to PE and (ii) ensured that students attend PE classes for 30 minutes each school day.

The Texas Education Agency (TEA) has since pointed to the positive improvements in fitness and body composition as evidence of the program's success; however, the fact that fitness scores were increasing in each subsequent year of the program suggests that other factors probably contributed to the average increase observed across some Texas schools (Texas Education Agency 2011).[4]

[4]In particular, the TEA compared the year-to-year differences in test scores in grantee schools only. They report that TFN led to statistically significant increases of 3.6-6.2 percentage points in aerobic capacity, trunk lift, upper body strength and endurance, and body composition between 2007 and 2009 (Texas Education Agency 2011).

Similarly, von Hippel and Bradbury (2015), uses a fixed effects model instrumenting for program participation over time, and estimates that TFN improved some measures of fitness for some groups, although they find no effects on BMI.[5] However, the authors model estimates by gender as well as groups of years of the program separately, making both the overall average effects and local average effects of the program difficult to distinguish, and they do not provide any support for their identifying assumption, casting doubt on the validity of the research design.

This research addresses these shortcomings and builds on the existing literature in a number of ways. First, we employ a regression discontinuity design, using the eligibility criteria directly, to compare otherwise similar students across the TFN eligibility threshold. Under the plausible assumption that other determinants of student fitness and performance are smooth across the school grant-qualifying cutoff, this research design allows us to compare outcomes of students in schools just below the eligibility threshold to students just above the threshold. In doing so, we are able to provide evidence that any changes in student outcomes are a result of the program, and not an artifact of school selection or other unobservable characteristics. Below we provide evidence in support of this assumption showing schools eligible for physical education grant funding were similar to schools just below the cutoff in terms of size, other financial resources and student composition. Second, we use individual-level administrative data to study student outcomes, which constitutes an improvement on school-level data since it additionally contains information on student raw test scores, attendance and disciplinary behavior. Moreover, the granular nature of these data allows us to test for compositional changes and detect student attrition.

We find that Texas Fitness Now did not improve physical fitness, including overall body mass index (BMI), body fat, aerobic capacity, or strength and flexibility. However, we show that TFN was effective at reducing the number of obese students, implying that

[5] In particular, von Hippel and Bradbury (2015) finds that effects were greatest in measures of strength, and greater for girls than for boys, although they report no statistically significant effects on shoulder flexibility. They find that both boys and girls in high-poverty middle schools could complete more pushups and a faster shuttle run. Girls could also complete more curl-ups, a higher trunk lift, and had a better sit and reach.

such interventions may be most effective for high-risk students.

Using individual-level data on student academic outcomes for Texas middle schoolers, we estimate no effect of the program on student achievement. Conversely, we present suggestive evidence that compulsory PE classes reduce attendance rates and increase incidents of disruption and misbehavior. These findings imply that interventions encouraging daily physical activity have the potential to negatively impact students if adolescents have a strong aversion to physical education.

Given the existing literature documenting the beneficial effects of physical education on elementary-aged students, these findings may be somewhat surprising. However, there are several potential explanations why 12-14 year olds respond differently than young children to physical activity initiatives. For instance, middle-school students may have already formed lifetime exercise and eating habits, and therefore are more obstinate than elementary-aged children to abandon unhealthy habits. Another explanation is that for economically disadvantaged and/or overweight teens, PE class may serve as a class period where students that struggle with aerobic exercises experience bullying and teasing.[6] Finally, since middle-schoolers are less energetic, physical education could make teenagers more tired than younger students, which may contribute to more distractions or misbehavior in the classroom. Below, we explore these possibilities in an effort to shed light on the more comprehensive effects of PE requirements for middle-school students.

The rest of this paper is organized as follows. Section 2 describes the Texas Fitness Now program in more detail. Section 3 describes the data and empirical strategy. Section 4 presents the main results. Section 5 provides a discussion of the main results and potential mechanisms before concluding.

[6] In particular, students report being bullied more in middle school than at any other point during their academic career. Over 22 percent of middle schoolers experience bullying at least once per week, as compared to 11 percent of high schoolers, and these effects are largest in low-income schools (National Center for Education Statistics 2018).

2 Background on Texas Fitness Now

The Texas Fitness Now (TFN) program was, at the time of initiation, the second-largest physical activity grant program in the US.[7] From 2007-2011, with the goal of curtailing childhood obesity and Type II diabetes, the State of Texas allocated $37 million to the poorest Texas middle schools to be spent on physical education and activity. Although nutrition was included as a facet of the program, TFN primarily focused on increasing funds and requirements for physical education.[8] Schools that accepted the funding were required to have students participate in physical education classes for at least 30 minutes per day or 225 minutes every two weeks.[9] To ensure compliance, applicants detailed how they would feasibly incorporate more PE classes into their curriculum, and participants were required to conduct fitness assessments twice per school year for evaluation. While no data exists on how individual schools allot time for PE, during the course of the program, over 1/3 of participating campuses reported having difficulty finding time to fit more PE classes into the curriculum, indicating that the program's time constraints were binding for many schools (TEA 2011).[10]

Table 1 displays the total amount of grant funding distributed in each year of the program, as well as how eligibility changed over time. Schools serving 6th, 7th, and 8th grade students were eligible to participate in the grant program if in the previous school year they had reported having at least 75% economically disadvantaged students, although this cutoff was extended to include schools with 60% economically disadvantaged

[7]For reference, the largest grant program is the ongoing yearly Carol M. White Physical Education Program, which allocated 72.6 million in grants for physical education to 149 entities in 2007 (U.S. Department of Education 2013).

[8]Only 7% of schools reported spending some money on nutrition initiatives (Texas Education Agency 2011).

[9]While Texas maintains baseline PE requirements for middle schools, students in grades 6-8 only need to participate in daily physical activity for 4 out of 6 semesters. "Physical activity" is defined at the district-level, but in many cases may include extra-curricular activities, such as marching band or cheerleading, although these activities would not meet PE requirements under TFN guidelines.

[10]One of the main limitations of our data is that we are unable to speak directly to how schools chose to reallocate timing for PE courses. After speaking to a few administrators, we note that the most popular route that schools took to implement the program was to reallocate elective course blocks to physical education. Therefore, it does not appear that many schools reduced time for math and reading as a result of the program. Instead, students would spend a semester in PE instead of courses such as art, theater, computer programming, or choir. Unfortunately, since our data do not contain any information on student courseload, we are unable to observe changes in student schedules.

students in 2009 and 2010.[11] Participating schools received an average of $10,000 to improve their physical education programs by purchasing equipment such as stop watches, pedometers, jump ropes, and free weights, as well as by adding more PE classes and hiring coaches and fitness instructors.[12]

The State of Texas required that the schools use the grant money as a supplement and not as a replacement for other academic programs. For example, TFN funds could not be spent on athletics or construction projects. During the first three years of the program, over 60% of schools spent money on traditional equipment, while, on average, 15% and 24% of schools added staff and classes, respectively. Nearly all of the participating schools reported that after receiving grant money they were able to provide opportunities for students to participate in physical activity at least 30 minutes a day or 225 minutes per two weeks (Texas Education Agency 2011).

3 Empirical Approach

This section describes the data and approach we use to estimate the causal effects of the Texas Fitness Now program on student health, fitness, test scores, discipline, and attendance.

3.1 Data

Data on fitness outcomes is from a statewide testing assessment for physical fitness, known as the FITNESSGRAM© test. These data are collected by health educators in the spring of each school year and are available from school years spanning 2007-2013. Given confidentiality concerns, FITNESSGRAM© data is available only by school, gender, and grade in the spring of each academic year. Notably, this limitation of the data means that we are unable to examine differences in physical ability across race, ethnicity, or

[11]Economically disadvantaged students are indicated as students that: (i) are eligible for free or reduced-price meals, (ii) are from a family with an annual income at or below the poverty line, (iii) are eligible for public assistance, and/or (iv) received any need-based financial assistance.

[12]Funding for eligible schools was determined by a fixed amount ($1,500) plus a proportional amount ranging from $11-$32 per 6th-8th grade student, depending on the school year.

fitness level. Students are tested in six main areas: body composition, aerobic capacity, upper body strength and endurance, abdominal strength and endurance, flexibility, and trunk extensor strength and flexibility.

Given their age and gender, results are measured relative to a range of acceptable scores for each test, known as the "healthy fitness zone" (HFZ). The HFZ is intended to reflect the level of fitness needed for good health. Students are not informed of the HFZ cutoff intentionally as a way to motivate them to perform their best. Since a majority of students are able to achieve their HFZ for all tests, any failures signal a need for more frequent exercise.

See Figure A1 for an official chart of healthy fitness zones for each fitness test, by age and gender. For body composition, HFZ levels represent a healthy weight. However, FITNESSGRAM© additionally indicates where students "need improvement"; these upper ranges correspond to overweight or obesity. Otherwise, HFZ ranges are intended to represent a typical level of fitness by age and gender. For example, according to Figure A1, a 13-year-old girl would need to complete 18 curl-ups and 7 push-ups to pass the corresponding fitness tests.

We additionally analyze effects of physical education on student performance and classroom behavior at the student level using data from the Education Research Center at UT-Austin. These data include student demographics, including economically disadvantaged status, as well as raw test scores, attendance, and disciplinary behavior for the full population of students enrolled in a Texas public school from school years spanning 2006-2011. One of the main advantages of these data is that we are able to examine effects on individual test scores and attendance as well as student discipline and suspensions, which are unavailable in the publicly available, school-level data. Additionally, these data allow us to analyze heterogeneous effects by subgroups, such as grade or gender. Moreover, we are able to rule out compositional changes in student cohorts across schools due to the policy, which allows us to test for student attrition. To study effects of TFN on middle-school students, we limit our sample to the population of Texas students in grades 6, 7, and/or 8 for the school years of the program (2007-2008 to 2010-2011).

Summary statistics for student characteristics and outcomes are presented in Table A1. Testing performance rates for reading and mathematics standardized tests, known as the Texas Assessment of Knowledge and Skills (TAKS) tests, are defined as whether a student met or exceeded the passing standard set by the state in the corresponding school year. Mean passing rates for math and reading TAKS tests range from 71-83 percent.

We categorize student-level disciplinary action into three main outcomes: total number of incidents, whether or not the student ever misbehaved, and total days of suspension in a given school year. On average, a student is reprimanded for one disciplinary incident per school year; however, only 27% of students misbehave in a given school year.

Attendance outcomes are based on mean student attendance rates for the entire school year. Student attendance rates are calculated by dividing the total number of days a student was present by the total number of eligible school days. As shown in Table A1, attendance in any given year is very high (above 95%).[13] That said, any measured changes in attendance rates are likely to be small.

3.2 Identification Strategy

Measuring the causal effects of Texas Fitness Now presents many challenges. For example, eligible schools self-select into TFN, and may additionally receive funding from other government programs, such as Title I, or the National School Lunch Program. Schools with the most motivated and ambitious faculty may therefore be those that choose to participate, and any estimated positive effects will overstate the benefits of the program.

To overcome such challenges and estimate effects of Texas Fitness Now, we use a regression discontinuity design (RDD). This strategy exploits the cutoff in program eligibility, the percent of economically disadvantaged students, to identify the causal effects of increased physical education requirements. Our approach is motivated by the idea that characteristics related to behaviors and outcomes of interest are likely to vary smoothly through this threshold. Thus, any discontinuity in fitness, test performance, discipline, or

[13]Notably, days suspended do not count as an absence.

attendance can be reasonably attributed to the change in the physical education curricula. We operationalize this identification strategy by estimating:

$$y_s = \theta EDcutoff + f(EDpct_s) + \lambda_t + \psi_g + \eta_s \tag{1}$$

where y_s is an average measure of student fitness, for school s or academic performance, attendance, and discipline outcomes for student s, f is a function of the percent of economically disadvantaged students for school s in school year 2006-2007, and $EDcutoff$ is a binary indicator for whether a school s meets the first year eligibility cutoff, as listed in Table 1. Because the program spans four years and multiple grade levels, we additionally include year fixed effects, λ_t in all specifications, and in some specifications we include grade fixed effects, ψ_g.[14] We control for the percent of economically disadvantaged students, normalized to zero, (running variable) linearly and allow it to vary on either side of the eligibility cutoff. Following Lee and Card (2008), we present standard errors that are clustered on the running variable, although we note that our estimates are not sensitive to this choice.[15]

Although, in practice, school eligibility for TFN was reevaluated each year, we use only the first year of eligibility criteria (the percent of economically disadvantaged students in 2006-2007) to define a school's position relative to the qualifying cutoffs in each year. In holding each school's eligibility constant across all years, we prevent the possibility of strategic schools manipulating their position across the threshold over time. Estimates based on this approach should yield more conservative estimates than those that depend on the yearly definition of treatment; however, we note that estimates for our preferred specifications are statistically similar for all outcomes using either approach.[16]

[14] As suggested by Lee and Card (2008) we use data on baseline covariates, including student race, ethnicity, gender, and economically disadvantaged status and school characteristics, such as student population, only to test the validity of the RD design, although below we additionally discuss results from models in which we add student-level demographic controls.

[15] We cluster on the running variable since the percent of economically disadvantaged students is rounded to the nearest tenth of a percentage point, although estimates are robust to clustering at the school level. Specifically, estimates on overall fitness levels and test scores remain statistically insignificant, while we estimate a statistically significant decrease in attendance rates and increase for all discipline outcomes at the full bandwidth of 15.

[16] We additionally note that the percent of economically disadvantaged students in a given school is highly correlated across years, and schools are unable to choose which students attend.

Our preferred specifications show estimates from all four school years, 2007-2008 to 2010-2011, using a bandwidth of 15 on either side of the cutoff. Given that the cutoff was expanded from 75 to 60 halfway through the program, we consider this bandwidth to be the largest possible range that exploits the variation in eligibility criteria, although we present estimates for a range of bandwidths in Figure A4.[17]

The identifying assumption for this research design is that characteristics related to outcomes of interest vary smoothly through the treatment threshold. Since eligibility for Texas Fitness Now is based on a school's previous year's percent of economically disadvantaged students, this feature helps to assuage concerns that the identification assumption may not hold.[18] Additionally, because schools likely do not have control over which students move into or out of their school district in any given year, manipulation of the running variable is unlikely. We test for this possibility, as suggested by McCrary (2008), in several ways. First, we confine the percent of economically disadvantaged students by school to be that of the first year program criteria, which eliminates the possibility for schools to move across the cutoff in subsequent years. In doing so, we estimate intent-to-treat effects, which will likely understate the true effects of the program.[19] Second, we test for discontinuities of several school characteristics, such as race, gender and ethnic composition as well as total number of students fitness tested across the eligibility threshold to address the possibility that unqualified schools close to the cutoff were

[17]A one-sided bandwidth greater than 15 would contain schools which may have been treated every year in the program. However, these schools, which contain a large proportion of economically disadvantaged students, may not be an appropriate comparison group for schools that fall just short of program participation after the expansion in eligibility. For example, a school with 80 percent economically disadvantaged students would be eligible for the program in the first two years, given the cutoff of 75, and we would effectively be comparing these students to those in schools with 70 percent economically disadvantaged students. However, when eligibility is expanded to 60, if we included this school in our analysis, we would estimate a local average treatment effect that effectively compares these same students to those in schools with 40 percent economically disadvantaged students. Therefore, we exploit the 15 percentage point expansion in program eligibility to estimate the local average treatment effects for students in schools that would not have already been treated prior to this criteria change.

[18]Schools are required to report the percent of economically disadvantaged students in October of the current school year. The Texas Comptroller announced original TFN eligibility cutoffs in June 2007, which suggests that schools were unaware of the threshold when reporting students statistics to the TEA in the previous year.

[19]Effects are similar when allowing for school eligibility status to vary across treatment years; we estimate no statistically significant effect on test scores, which corresponds to Columns 3 and 6 in Table 3, an increase in disciplinary incidents of 0.07, which corresponds to Column 3 in Table 4, and a decrease in attendance rates of 0.002, which corresponds to Column 3 in Table 5.

systematically different than those that barely qualified for funding. Third, we show that the percent of economically disadvantaged students does not exhibit a discontinuity at either the 60% or 75% cutoff, which provides some support for the notion that the State of Texas chose these cutoffs arbitrarily and schools were not able to manipulate around them. Fourth, we present evidence that student selection into or out of program-eligible schools is not driving our results by providing estimates of the number of schools that a student attends during the sample period. Fifth, we estimate Equation 1 for all outcome variables using pre-period data from 2006 to show that any estimated effects for 2007-2011 are a result of the program, and not an existing feature of the data.

With any education-based school reform, it is important to consider whether there are additional grants available for schools that meet this same cutoff, which may lead to additional treatments that affect academic outcomes but are unrelated to physical fitness. Indeed, Title I funding, which is set aside for schools with at least 40% of economically disadvantaged students is a major source of school funding and provides an average of \$630,000 dollars to Texas schools each year. We provide evidence in Figure A2 that this cutoff is not sharp, as many schools with small shares of economically disadvantaged students still receive Title I funds.[20]

Texas did initiate two performance incentives programs in 2006, the Governor's Educator Excellence Grant (GEEG) and the Texas Educator Excellence Grant (TEEG), as a way to increase quality of education through higher pay for school personnel and professional development. Although one component of eligibility for funds was based on the number of economically disadvantaged students, schools also were required to display acceptable testing performance. Due to this additional requirement, over 200 schools below our treatment threshold participated in these two grant programs, indicating that additional grant funding was continuous through the TFN eligibility cutoff. Furthermore,

[20]Similarly, there exist community standards for the National School Lunch Program, in which a school with many economically disadvantaged students are eligible for funds to provide lunch to all students. However, to participate in this program, schools receive funds on a phase-in rate, starting at the 42.5 percent economically disadvantaged student cutoff. Therefore, we would not observe a discontinuity of funds at the 60 or 75 percent cutoffs due to this program. Moreover, the Community Eligibility Provision was rolled out in Texas in 2013, which should mitigate any concern that discontinuities in school lunch funding is driving our results.

few middle schools participated in these two programs; over half of TEEG and GEEG funds went to Texas high schools, which are not included in our sample. Importantly, neither program used the 60 or 75 percent economically disadvantaged as a funding criteria.[21] Finally, we note that to the extent that these other school resources improve academic performance and/or attendance, any negative findings of TFN on performance would be understated.[22]

4 Main Results

4.1 Effects of Texas Fitness Now Eligibility on Funding

Figure 1 presents the estimate for the main measure of TFN participation: total grant money awarded. Here we present residual means plots (accounting for year and grade fixed effects) using 3 percentage point bins as well as the respective discontinuity estimates from Equation 1. In all figures the running variable (the percent of economically disadvantaged students) is normalized to zero due to the fact that this cutoff changed in 2009.

As shown in Figure 1, we estimate statistically significant discontinuities in funding based on the eligibility cutoff, although we note that this criteria is not sharp.[23] Specifically, schools that met the eligibility criteria received, on average, approximately \$10,600 in TFN funding, which corresponds to approximately \$15 per student. We note that, while \$15 per student does not seem like a large intervention, this represents about a 6% increase in total per pupil instructional spending and is 17 times the average Texas middle-school PE budget. Moreover, cost-benefit analyses of similar physical activity interventions estimate spending requirements of only approximately \$4 per student to increase physical activity to 30 minutes per day and subsequently decrease obesity by

[21] Additionally, we find no evidence of an existing discontinuity at the TFN eligibility threshold on total school funding and total operational expenses ($p > 0.8$).

[22] To our knowledge, there are no other grants that utilized the same economically disadvantaged cutoffs during our sample period.

[23] Although eligibility was intended to limit funding only to middle schools, eligibility was also extended to alternative schools with any grade level. About 6 percent of Texas schools that received funding did not contain students in 6th, 7th, or 8th grades. We do not include any of these schools in our analyses.

0.02 BMI units over two years (Barrett, Gortmaker, Long, Ward, Resch, Moodie, Carter, Sacks, Swinburn, Wang, and Cradock 2015). Furthermore, the $15 per student average includes some schools that did not receive any TFN funding; however, take-up of the program was high with approximately 88-95 percent of eligible Texas schools both applying and receiving the grant in a given year.

4.2 Effects of Texas Fitness Now on Fitness

Since the intent of TFN was to improve fitness outcomes and reduce obesity for middle-school students, in this section we present estimated discontinuities for body composition and physical fitness outcomes, including measured tests for BMI, body fat, aerobic activity, strength and flexibility. Importantly, these data are only available by school, grade, and gender, and are not obtainable at the individual level.

TFN participation stipulated that students attend PE class every day for at least 30 minutes. Since a majority of schools in Texas do not have requirements for the length of PE class, and many schools do not require students to attend PE for all three years of middle school, this was likely a noticeable change in curricula for many students (CDC, 2007). Indeed, a large majority of schools (82-87 percent) reported being able to restructure curriculum to meet this requirement (Texas Education Agency 2011).[24,25]

We first show effects of TFN on body composition. Importantly, the data do not include information on student-level BMI calculations; we only have information on the percent of students with a healthy BMI, students that are at-risk, or are overweight, and students that have high-risk, or are obsese. Figure 2 shows residuals means plots for the percent of students with a healthy body-mass index using 3 percentage point bins.[26] This figure shows some support for the notion that TFN was ineffective at reducing BMI for low-income students.

[24]In Texas middle schools that have a physical education requirement, there is no requirement for everyday physical activity. Students are required to attend PE class the equivalent of 225 minutes per two weeks or 30 minutes per day for four semesters, but may choose which semesters to participate.

[25]Since the TEA does not maintain records on block schedule schools, we are unable to test differences between students with an A/B class schedule and students with 7-8 class periods every day.

[26]Notably, Texas schools that use FITNESSGRAM© as a measure of physical fitness have flexibility to choose which measure of body composition to report- over 75 percent report BMI.

In Table 2, we display the corresponding point estimates. Each column is a separate regression, and each regression uses data for all 6th, 7th, and 8th graders in Texas from school years 2007-2008 to 2010-2011. In Column 1, we first estimate the optimal (bias-corrected) bandwidth and polynomial order, as suggested by Calonico, Cattaneo, Farrell, and Titiunik (2016). This procedure specifies one-sided optimal bandwidths ranging from 5.3 to 13.5 and first-order polynomials for all outcomes. In Column 2 we adopt a bandwidth of 12, for comparison, while in Column 3 we display estimates using a bandwidth of 15, which is the full bandwidth using the expansion in eligibility criteria in 2009.

As shown in Column 1, we estimate that TFN led to approximately a 2.2 percentage point *reduction* in the percent of students with a healthy BMI. This could be due to several reasons. For example, if students are working out more, they could be counteracting the effects of physical activity by eating more calories. Or, perhaps students are more tired and therefore less likely to play sports at home or participate in after-school activities. Another possibility is that students face bullying or hardship in the locker room and become discouraged or give up trying to lose weight. However, estimates in Columns 2 and 3 are statistically insignificant, indicating that the program likely had no effect on student BMI. Based on the estimates in Table 2, we can rule out effects of a 0.03 percentage point increase of students with a healthy BMI, or a 0.46% increase.[27]

We also note that there is heterogeneity across student preferences for physical fitness; therefore, it may be more informative to analyze the effects of daily PE classes on students that are overweight versus students that are obese. We present estimates and their corresponding 95% confidence intervals across a range of bandwidths for these respective groups of students in Figure 3. Across all bandwidths, we find that the number of obese students decreased as a result of the program, which implies that although the intervention was ineffective at helping students reach a healthy BMI overall, and may have *increased* weight for some students, such policies may be able to help the heaviest

[27] Another possibility is that, given the metrics of "healthy", "at-risk", and "high-risk", it is possible that TFN had an average, positive effect on BMI, but this effect was not large enough to move students into or out of the various categories.

individuals lose weight.

Importantly, although it may be difficult for school-mandated PE classes to affect BMI in an economically meaningful way, we may expect that overall physical fitness levels would improve. To observe effects of TFN on a broad measure of fitness, we construct a school-level variable for the average number of fitness tests passed and present these estimates in Columns 4-6 of Table 2. These tests include aerobic activity, strength and flexibility and do not include measures of BMI. Estimates shown in Table 2 are precise enough to rule out even small increases in the number of tests passed (1.9 percent), implying that TFN did not marginally increase fitness levels, on average.

Finally, we test for more specific indicators of physical fitness, as measured by the FITNESSGRAM© test and present results in Figure A3. These tests include aerobic capacity, strength, and flexibility.[28,29] Estimates for all fitness outcomes are statistically insignificant and indicate that TFN had little to no effect on aerobic capacity, strength, or flexibility among middle-school students. These estimates are similar in magnitude across all bandwidths.[30]

Contrary to von Hippel and Bradbury (2015), we find little evidence that TFN improved student fitness levels. However, we note that it is possible that TFN failed to encourage students that were already relatively healthy to marginally pass more fitness tests, but was able to target those students with the worst levels of physical fitness. We also note that, while, on average, TFN did not reduce the number of overweight students, if daily PE classes increase physical activity for sedentary adolescents, students may still gain other, unobserved, independent health benefits (Institute of Medicine 2012).

[28]FITNESSGRAM© provides opportunities for schools to test strength and flexibility in a variety of ways. These tests include curl-ups, trunk lift, 90 degree push-ups, pull-ups, flexed arm hang, sit and reach, and shoulder stretch. See `http://pyfp.org/doc/fitnessgram/fg-07-muscular.pdf` for a description the objectives, scoring, and instructions for each test.

[29]In testing aerobic activity, schools have the option to complete the pacer test or have students complete a mile run without stopping. Nearly 75% of schools opt for the pacer test over the mile run. The pacer test, also known as the progressive aerobic cardiovascular endurance run, is a multistage shuttle run designed to test endurance and aerobic capacity by requiring students to run across a 20-meter space at a specified and increasing pace, making the test increasingly more difficult as time progresses.

[30]While many reports have pointed to the positive outcomes for fitness, especially for young girls, we find no major differential effects of TFN on physical fitness outcomes by gender (TEA, 2011). See Table A2 for effects of TFN on body composition, aerobic capacity, and strength and flexibility for girls.

4.3 Effects of Texas Fitness Now on Test Scores

Given the potential for changes in PE curricula to affect student focus and achievement, we now examine the effects on academic outcomes. Specifically, the State of Texas measures academic performance for grades 3-12 based on passing rates for reading and mathematics on standardized tests, known as the Texas Assessment of Knowledge and Skills (TAKS) tests.[31] TAKS subject tests measure knowledge on the state-mandated curriculum objectives and consist of multiple-choice questions scored by a computer. Scores are scaled and the passing score levels change slightly from year to year depending on the test's level of difficulty. According to data from the Texas Education Agency on state testing, TAKS attendance and completion is 99-100% for all years during the sample period.

For students, the TAKS test represents a high-stakes test that they must sit for once a year in the spring. If a student does not pass either the math or reading exams at the end of the 8th grade year, they are not permitted to advance to high school. If a student fails either exam in the 6th or 7th grade, they may advance grades, but are required to take additional remedial courses to catch up to the knowledge level of their peers. We focus our analyses on exams that students must take every year, namely math and reading.[32]

In Figure 4 we present evidence that TFN did little to improve student performance, as measured by TAKS passing rates and raw test scores. Mirroring these findings, Table 3 displays estimates on passing rates for math and reading TAKS scores from a baseline specification derived from Equation 1, controlling for grade and year fixed effects.[33] Passing grades are determined by the Texas Education Agency, and are measured by the number of questions answered correctly compared to the passing standard set by the state in the corresponding year. We additionally show estimates for whether students

[31] From 2012-14 the TAKS test was phased out, as Texas switched to the State of Texas Assessments of Academic Readiness (STARR) test. Therefore, we do not analyze any longer-term effects of TFN on school years 2011-2012 or 2012-2013, after the program had ended due to concerns of comparability.

[32] While some middle-school students are required to additionally test for writing, social studies and science in some years, we limit the analysis to reading and mathematics TAKS scores, given that all students take these tests each year from 3rd-11th grade. When estimating effects for these alternative subject tests, we find no evidence that PE investments affect the percent of students that pass.

[33] In all specifications using individual-level data we control for year and grade fixed effects, although we note that our results are not sensitive to the inclusion of grade controls.

received a “commended” recognition, a distinction of high achievement that only 20-33% of students receive in a given year, and the number of questions the student answered correctly, i.e. the raw TAKS scores.[34]

Although the TEA reports that daily PE requirements have the potential to increase test scores, we find little evidence to support this finding (TEA, 2011). Estimates across all columns of Table 3 indicate statistically insignificant effects of TFN on both math and reading scores. These effects are consistent across specifications and are precise enough to rule out effects on math and reading passing rates of 0.56% and 0.048% percent or larger, respectively.[35] Therefore, our findings suggest that investments in physical education do not negatively (or positively) affect overall student performance, which is consistent with previous studies on adolescent physical activity.[36]

Since TFN was geared towards helping economically disadvantaged students, and since we may expect fitness interventions to affect students differently by gender, we additionally analyze how TFN affected test scores for students across these subgroups in Table A3.[37] Effects for females and economically disadvantaged students are positive and statistically similar to estimates of the overall sample, suggesting that these estimates are not driven by one particular group.[38]

[34]Specifically, the State of Texas designates a students' score to be “commended” if they score at least 2100 out of 2400 scaled points.

[35]These effects are relatively small when compared to effects found using first-order academic interventions. For comparison, assignment to smaller class sizes in the well-known Tennessee STAR experiment in grades K-3 increased student test scores in grades 6-8 by 3.6-6.0 percentile points (Schanzenbach 2007). Similarly, students in grades 4-8 lotteried into New York City charter schools gained 12 and 9 percent of a standard deviation each year on math and English test scores, respectively (Hoxby, Murarka, and Kang 2009). Our estimates suggest that students spending up to 2.5 hours more per week in PE gain less than 0.4 percentile points in math, with smaller effects for reading, or less than 0.9 percent of a standard deviation increase.

[36]For other studies that analyze the effects of physical education interventions on student performance, see Dills, Morgan, and Rotthoff (2011), and Cawley, Frisvold, and Meyerhoefer (2013), von Hippel and Bradbury (2015). In particular, Dills, Morgan, and Rotthoff (2011) estimates a value-added model and finds that weekly PE classes have no statistically significant or economically significant impact on test scores for elementary-aged children. Cawley, Frisvold, and Meyerhoefer (2013) uses the Early Childhood Longitudinal Study, Kindergarten Cohort and instruments for child PE time, according to state policies. They find no evidence of spillovers of PE on test scores for elementary school children. von Hippel and Bradbury (2015) uses school-level data to study TFN and finds no effect of the grant program on academic achievement.

[37]We also provide estimates for the effect of TFN on test scores by grade in Table A4. We find no statistically different effects of daily PE requirements by grade level.

[38]We could also measure the effects of TFN on test performance, discipline, and attendance by race and ethnicity, however, we do not include these subsamples in our main analyses for two reasons. First, we are unable to examine effects of TFN on fitness by race and ethnicity. Second, we estimate a small

4.4 Effects of Texas Fitness Now on Disciplinary Action

Although there is little evidence to suggest that mandatory PE classes affect student health and fitness, such initiatives may affect student behavior in a number of ways. First, it's possible that PE classes encourage restless students to expel nervous energy, allowing them to focus more on coursework, and be less disruptive throughout the day. However, if students become more tired throughout the day due to the increase in physical activity and/or have strong preferences against such classes, we would expect an increase in misbehavior. In Figure 5 and Table 4, we provide some evidence to suggest the latter.

Before discussing statistical evidence of TFN on disciplinary action, we first present visual evidence that mandatory PE requirements affect student behavior in the classroom. Figure 5 displays the effect of TFN on the total number of student disciplinary incidents, proportion of student offenders, and total days suspended. Each figure shows large, positive discontinuities at the eligibility cutoff. Overall, the set of results in Figure 5 indicate that daily PE requirements lead to more recorded instances of student misbehavior.[39]

Table 4 shows additional estimates from regressions with smaller bandwidths. Models with optimally-chosen bandwidths (Columns 1, 4, and 7) as well as models with a one-sided bandwidth of 12 (Columns 2, 5, and 8) yield statistically similar but insignificant estimates. Therefore, despite the proposition that PE classes incite student focus and good behavior, we show no evidence that TFN *reduced* classroom disruptions. However, we do present some evidence in Columns 3, 6, and 9 that daily compulsory PE requirements may actually *increase* instances of classroom misbehavior. Estimates from a model with the full bandwidth indicate that TFN resulted in a statistically significant

and statistically significant discontinuity at the 10% level for some outcomes one year before the program ($p \geq 0.09$), although we do not find such a discontinuity in aggregate outcomes. These effects yield some concerns that the RD model may be misspecified when looking at some subgroups, thus we omit any analysis by race and ethnicity throughout the paper.

[39] Arguably, we may expect schools that hired more staff to be able to report more disciplinary incidents due to increases in monitoring. Unfortunately, we do not have data on school-level expenditures from the TFN grant funding and are unable to speak to this mechanism directly. However, we acknowledge that the increases in disciplinary action that we observe in the data are not borne entirely by a small population of schools, which lends some evidence to the argument that these effects are at least partially student-driven. Moreover, only 7 percent of TFN schools added staff from 2008-2010, indicating that monitoring is unlikely to be responsible our results.

increase of 0.15 incidents for each student, on average, which corresponds to an increase in disciplinary action of about 15.6%, or 73 per school year.

Notably, this measure could represent an increase on either the intramargin or inframargin; that is, either students that were already likely to misbehave did so more frequently, or there were more instances of new offenders. We investigate the extent to which one of these effects is driving the total effect in Columns 4-6 of Table 4, which presents the proportion of total students that caused a classroom disruption. In Column 6, we estimate that TFN increased the proportion of misbehaving students by 0.02, or 7.4%. Therefore, we report suggestive evidence that daily PE classes for middle-school students may not only lead to more disciplinary action but also encourage more students to act out.

Finally, as a way to analyze the intensity of student misbehavior, we investigate how many days students were suspended as a result of disciplinary infractions and present results in Columns 7-9. Although estimates are not statistically significant across all bandwidths, estimates in Column 9 indicate that TFN increased the number of days suspended by 23.7 percent. In terms of class time, this corresponds to about 0.84 fewer days of traditional coursework for misbehaving students in TFN-eligible schools, as compared to students in the non-eligible middle schools.[40]

One explanation of these findings is that mandatory PE classes increase bullying in school. Although the ERC student-level data does not contain information on *where* the instances of disciplinary action occurred, it is possible that more frequent interaction in the locker room leads to more teasing and fighting throughout the school day. Given that nearly all cases of US school infractions occur in the classroom, (e.g. 60 percent of major offenses and over 70 percent of minor offenses (Gion, McIntosh, and Horner 2014)), it may also be possible that both classroom bully increases *and* locker room bullying increases as a result of more PE days, but the lack of visibility from teachers in gym escapes punishment. These implications are especially worrisome, given that such bullying can

[40]We additionally provide estimates of disciplinary action by grade in Table A4 as well as by gender and economically disadvantaged status in Table A3. While estimates are larger in magnitude for 8th graders, estimates are not statistically different at the 1% level across grades, gender, or economic status. Similar to findings in Table 4, estimates are less precise at smaller bandwidths.

be counterproductive to the goals of physical education programs, as children who are criticized for their physical skills or ostracized in gym class perform worse in school and experience a decrease in physical health and fitness in the long run (Jensen, Cushing, and Elledge 2013).

4.5 Effects of Texas Fitness Now on Attendance

If students' preferences for physical education differ from that of other school subjects, increasing PE requirements may affect incentives for student attendance. We test this hypothesis in Figure 6 and Table 5. In Table 5, Column 1 shows estimates from a model based on Equation 1 that uses a MSE-RD estimated optimal bandwidth. Estimates are similar across columns and indicate that TFN did not encourage students to attend school more frequently. Although the baseline attendance rates are high, for some bandwidths we observe a statistically significant decrease in attendance rates for students in TFN-eligible schools as a result of the program. Estimates across Columns 2-3 in Table 5 indicate that mandatory PE classes reduce attendance for all students by 0.30 percentage points, or 0.31 percent. These findings suggest that, at best, investments in physical education do not cause students to change their decision to come to school; at worst, daily PE mandates could discourage some students from attending class.[41]

In Tables A3 and A4, we additionally explore discontinuities in average attendance rates for different student subgroups, including gender, economic status, and grade across the cutoff. We find that effects on attendance are larger for economically disadvantaged students, although effects are not statistically different from the full sample. We find no differential effects by gender (Table A3) or grade (Table A4).

These findings suggest that in low-income schools, mandatory PE classes could potentially discourage student attendance. Four arguments support this idea: (i) overweight or unathletic students may fear being ostracized or face bullying in the locker room, and would rather skip school than face hardship, (ii) students may fear activities such as running and jumping are too difficult and prefer not to exercise at school, (iii) adolescents

[41]Notably, student suspensions do not factor into attendance as an absence. Therefore, it's not the case that the increase in disciplinary action is driving the reduction in attendance rates.

concerned for their appearance may not want to look sweaty or untidy during the school day, and/or (iv) middle-school students do not enjoy engaging in movement or physical activity.

It is well-documented that preferences for physical activities and recreation change as students mature. Accordingly, adolescents' overall level of physical activity decreases significantly in 7th and 8th grade at a critical time of physical and cognitive development, especially among girls, with only 17% meeting the daily activity guideline by age 15 (Nader, Bradley, and Houts 2008). Given that physical activity after elementary school progressively decreases, the drop in attendance could reflect taste-based preferences for sitting in a classroom over exercising at school (Butt, Weignberg, Breckon, and Claytor 2011). However, taken with the positive effects of disciplinary action reported in the previous section, TFN may have increased bullying enough to discourage some students to attend school.[42] In either case, to the extent that attendance is crucial for attaining knowledge, paramount for a student's academic success, or is beneficial for emotional or social growth, the effects discussed above are of considerable consequence.

4.6 Robustness Checks

As discussed in Section 3, we perform a number of robustness checks to provide additional support for the identification assumption. There may be some concerns that schools just above the eligibility cutoff are systematically different than schools just below the cutoff. For example, if schools that participate in TFN have a different composition of students, our findings may be picking up differential behavioral reactions to PE requirements across students. Moreover, if schools receiving TFN funding want to report improved fitness scores as a way to motivate future state funding opportunities, coaches may encourage the out-of-shape students to sit out of class on testing days (although this technically violates FITNESSGRAM© rules), which would overstate any positive

[42] We also acknowledge that one plausible alternative explanation is that injuries could result from increased physical exertion that also lead to more student absences. While we cannot directly address this issue using available data, according to conversations with PE coaches at various Texas high schools, injuries in class are not particularly common. Furthermore, the general policy is that injured students with a doctor's note would be allowed to sit on the sidelines and theoretically would not be expected to miss more than one class day due to an injury.

fitness results in schools just above the eligibility threshold. To test for randomness in the eligibility cutoff, we estimate effects of the percent of economically disadvantaged students in the 2006-2007 school year on the total number of students, the total number of students fitness tested, the percent of female students, the percent of black students, the percent of Hispanic students, and the percent of economically disadvantaged students in our sample and present these results in Figure 7. Across all outcomes these estimates are statistically insignificant at the 5% level, providing some support that schools on either side of the cutoff are similar on measurable characteristics.[43]

While we do estimate a statistically significant effect at the 10% level of nearly 4 percentage points for the proportion of black students at the cutoff ($t = 1.75$), we note that controlling for demographics yields estimates that are nearly identical. In particular, estimated effects from our preferred specification for attendance and all discipline outcomes are statistically similar at the 1% level when including controls for race and ethnicity. Moreover, we similarly estimate a statistically significant discontinuity of 3 percentage points in the proportion of black students prior to the program's initiation, but do not estimate significant effects for student outcomes in this period, implying that any changes observed in fitness, academic performance, attendance, and discipline after 2006 is a result of the intervention and not racial composition.

In Figure 8, we additionally test for the density of the running variable, the percent of economically disadvantaged students. To the extent that schools are aware of the eligibility cutoff and can manipulate this threshold, there will be a discontinuity in the number of schools in each bin. Estimates indicate that there is no discontinuity in the number of schools just above and just below the cutoff, suggesting schools did not manipulate the cutoff to receive TFN funding. Similarly, Figure A5 shows the average number of schools that a single student enrolled in during the four-year sample period to test for student attrition. Estimates are statistically insignificant across all bandwidths, indicating that students did not actively manipulate around the TFN eligibility cutoff.

Next, in Figure 9, we present evidence that any discontinuities in test scores, dis-

[43] We also note that, when replicating figures similar to Figure A4, estimates on all school characteristics are statistically insignificant at the 5% level across all possible bandwidths.

ciplinary incidents, and attendance are the result of the program, and not preexisting anomalies in the data. To this end, we replicate our findings from Equation 1, limiting our sample to the school year before the TFN program began, 2006-2007. We estimate no statistically significant discontinuities in test rates, disciplinary incidents, or attendance rates, which provides additional support for the notion that investments in PE programs, and not other factors, are driving our main results.

Finally, we note that since our procedure to determine the optimal bandwidth and polynomial order does not relax our assumption of linear fit and uniform weighting on either side of the cutoff, we have additionally analyzed how our estimates change under different functional forms and show these results in Table 6. When we impose second- and third-order polynomial fits into our main regression equation, we observe that, while the magnitude of the estimates remain consistent, the significance of the estimates decreases dramatically. Notably, since the choice of polynomial order seems to have a large impact on precision, this may indicate that using higher-order polynomials causes us to overfit the data. Similarly, using triangular kernel weights (Column 4) estimates yield similar effects for attendance as compared to the baseline results in Column 1, and estimates for discipline remain similar in magnitude.

5 Discussion and Conclusion

This paper analyzes the effects of increased physical education requirements on student health, fitness, academic performance, and student misbehavior. Using a regression discontinuity approach, we estimate that school-level interventions mandating daily PE classes do not lead to overall improvements in student fitness, including cardiovascular endurance, strength, and flexibility. In particular, although the goal of TFN was to reduce BMI, we show empirically that the program was ineffective at achieving this goal, on average, although we provide some evidence to indicate that TFN was effective at reducing BMI for the most at-risk students.

Moreover, we find that TFN did not lead to positive spillover effects in the classroom, including improvements in math and reading passing rates. However, we present some evidence that daily PE may be detrimental to student behavior, resulting in increases in disciplinary incidents and reductions in attendance. Given the current recommendations of daily compulsory PE by agencies such as the CDC as well as the US Surgeon General, these findings can better inform policymakers of the effectiveness and potential unintended consequences of such policies for adolescents.

Unfortunately, a limitation of the available data is the inability to accurately test for all the possible mechanisms that explain these results in the classroom. One potential explanation is that requiring students to spend more time in PE class only reduces time spent in other electives, like theater and choir. Alternatively, if students experience diminishing returns to learning, we may expect that as long as the time spent in PE class does not disproportionately take away from one particular academic subject, test performance should be unaffected. In either case, because students are not significantly reducing learning time in math and reading during the day, they perform similarly on standardized tests. Given that, in some cases, we estimate adverse consequences in attendance and disciplinary incidents, the null average effect for test scores seems surprising. One might expect disruptions in class or absences to lead to less learning overall. Although our results point to no effect on student learning, we acknowledge another possibility: athletically inclined students enjoy PE classes and perform better on exams, while those that are most negatively effected by the program perform worse. In this scenario, we would similarly estimate a zero effect on test scores, although we would expect the policy implications to vary based on the composition of students. However, we note that we do not find evidence of such heterogeneous effects across student subgroups of grade, gender, and economic status.

While these explanations are important to consider in terms of student achievement, they do not explain why we observe a decrease in attendance rates and an increase in disciplinary behavior for students at TFN-eligible schools. One mechanism that explains both negative student behaviors is the possibility that adolescents strongly dislike PE class due to social stigma. For example, overweight and obese children face strong social barriers and social isolation from their peers (Latner and Stunkard 2003; Janssen, Craig, Boyce, and Picket 2004). The physical demand of PE class along with the potential for increased teasing or bullying, either in the locker room or during class, may incentivize some students to act out or skip classes altogether. This is an especially important issue if interest in school and academic performance for affected students declines in the long run.

We conclude that despite the frequent and recent recommendations for more physical activity in schools, standard PE classes are not effective in improving students well-being and may even be detrimental. Given that the TFN program was the second-largest grant program in the United States at the time of its conception, our findings have important policy implications for school spending and time allocation. In terms of cost-effectiveness, we posit that the $37 million in funding would have likely been better spent on programs such as school-based health centers if the end goal is to improve student health (Guo, Wade, Pan, and Keller 2010), and/or Head Start or tutoring programs that have been proven to improve student performance and close the achievement gap for low-income students (Gibbs, Ludwig, and Miller 2011). Lastly, there is scope for more work to be done on testing potential mechanisms to determine why and how physical education classes might lead to negative outcomes for middle-school students.

References

American Academy of Pediatrics (2006): "Active Healthy Living: Prevention of Childhood Obesity through Increased Physical Activity," *Pediatrics*, 117, 1834–1842.

Barrett, J. L., S. L. Gortmaker, M. W. Long, Z. J. Ward, S. C. Resch, M. L. Moodie, R. Carter, G. Sacks, B. A. Swinburn, Y. C. Wang, and A. L. Cradock (2015): "Cost Effectiveness of an Elementary School Active Physical Education Policy," *American Journal of Preventative Medicine*, 49(1), 148–159.

Butt, J., R. Weignberg, J. Breckon, and R. Claytor (2011): "Adolescent Physical Activity Participation and Motivational Determinants Across Gender, Age, and Race," *Journal of Physical Activity and Health*, 8, 1074–1083.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2016): "rdrobust: Software for Regression Discontinuity Designs," Discussion paper, University of Michigan.

Carlson, S. A., J. E. Fulton, S. M. Lee, L. M. Maynard, D. R. Brown, H. W. K. III, and W. H. Dietz (2008): "Physical Education and Academic Achievement in Elementary School: Data from the Early Childhood Longitudinal Study," *American Journal of Public Health*, 98(4), 721–727.

Cawley, J., D. Frisvold, and C. Meyerhoefer (2013): "The Impact of Physical Education on Obesity Among Elementary School Children," *Journal of Health Economics*, 32(4), 743–755.

Cawley, J., C. Meyerhoefer, and D. Newhouse (2007): "The Impact of State Physical Education Requirements on Youth Physical Activity and Overweight," *Health Economics*, 16(12), 1287–1301.

Centers for Disease Control and Prevention (2007): "Physical Education and Physical Activity: Results from the School Health Policies and Programs Study 2006," .

——— (2010): "The Association Between School-Based Physical Activity, Including Physical Education, and Academic Performance," Accessed 13-November-2017 at `https://www.cdc.gov/healthyyouth/health_and_academics/pdf/pa-pe_paper.pdf`.

——— (2016): "Childhood Obesity Facts," Accessed 1-July-2016 at `www.cdc.gov/healthyschools/obesity/facts.htm`.

Cutler, D., E. Glaser, and J. Shapiro (2003): "Why Have Americans Become More Obese?," *Journal of Economic Perspectives*, 17(3), 93–117.

Datar, A., and R. Sturm (2004): "Physical Education in Elementary School and Body Mass Index: Evidence from the Early Childhood Longitudinal Study," *American Journal of Public Health*, 94(9), 1501–1506.

Dills, A. K., H. N. Morgan, and K. W. Rotthoff (2011): "Recess, Physical Education, and Elementary School Student Outcomes," *Economics of Education Review*, 30, 889–900.

Finkelstein, E., I. Fiebelkorn, and G. Wang (2003): "National Medical Spending Attributable to Overweight and Obesity: How Much, and Who's Paying," *Health Affairs*, W3, 219–226.

Finkelstein, E., J. Trogdon, J. Cohen, and W. Dietz (2009): "Annual Medical Spending Attributable to Obesity: Payer-and Service-Specific Estimates," *Health Affairs*, 28(5), w822–w831.

Gibbs, C., J. Ludwig, and D. Miller (2011): "Does Head Start Do Any Lasting Good?," NBER Working Paper No. 17452.

Gion, C. M., K. McIntosh, and R. Horner (2014): "Patterns of Minor Office Discipline Referrals in Schools Using SWIS," Accessed 29-November-2018 at `https://www.pbis.org/Common/Cms/files/pbisresources/EvalBrief_May2014.pdf`.

Guerra, P., M. C. Nobre, J. C. da Silverira, and J. de Aguiar Carrazedo Taddei (2013): "The Effect of School-Based Physical Activity Interventions on Body Mass Index: A Meta-Analysis of Randomized Trials," *Clinics (Sao Paolo)*, 68(9), 1263–1273.

Guo, J., T. Wade, W. Pan, and K. Keller (2010): "School-Based Health Centers: Cost-Benefit Analysis and Impact on Health Care Disparities," *American Journal of Public Health*, 100(9), 1617–1623.

Hoxby, C. M., S. Murarka, and J. Kang (2009): "How New York City's Charter Schools Affect Achivement, August 2009 Report," .

Institute of Medicine (2012): "Accelerating Progress in Obesity Prevention: Solving the Weight of the Nation. National Academies Press, Washington, DC.," Accessed 4-June-2018 at `https://www.ncbi.nlm.nih.gov/pubmed/24830053`.

——— (2013): "Educating the Student Body: Taking Physical Activity and Physical Education to School," Accessed 2-October-2017 at `https://www.ncbi.nlm.nih.gov/pubmed/24851299`.

Janssen, I., W. Craig, W. Boyce, and W. Picket (2004): "Associations Between Overweight and Obesity with Bullying Behaviors in School-Aged Children," *Pediatrics*, 113(5), 1187–1194.

Jensen, C., C. Cushing, and A. Elledge (2013): "Associations Between Teasing, Quality of Life, and Physical Activity Among Preadolescent Children," *Journal of Pediatric Psychology*, 39(1), 65–73.

Knaus, M., M. Lechner, and A. Reimers (2018): "For Better or Worse? The Effects of Physical Education on Child Development," IZA Discussion Paper No. 11268.

Latner, J., and A. Stunkard (2003): "Getting Worse: The Stigmatization of Obese Children," *Obesity Research*, 11(3), 452–456.

Lee, D., and D. Card (2008): "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 142, 655–674.

Let's Move (2016): "The Epidemic of Childhood Obesity," Accessed 02-July-2016 at http://www.letsmove.gov/learn-facts/epidemic-childhood-obesity.

Nader, P., R. Bradley, and R. Houts (2008): "Moderate-to-Vigorous Physical Activity From Ages 9 to 15 Years," *The Journal of the American Medical Association*, 300(3), 247–352.

National Center for Education Statistics (2018): "Indicator 7: Discipline Problems Reported by Public Schools," Accessed 29-November-2018 at https://nces.ed.gov/programs/crimeindicators/ind_07.asp.

Pate, R., and J. O'Neill (2008): "Summary of the American Heart Association Scientific Statement: Promoting Physical Activity in Children and Youth: A Leadership Role for Schools," *The Journal of Cardiovascular Nursing*, 23, 44–49.

Schanzenbach, D. W. (2007): "What Have Researchers Learned from Project STAR?," *Brookings Papers on Education Policy*, 9, 205–228.

Texas Education Agency (2011): "Evaluation of the Texas Fitness Now Grant Program: 2007-08 to 2009-10 School Years," Accessed 01-July-2016 at http://tea.texas.gov/WorkArea/DownloadAsset.aspx?id=2147496810.

Tomporowsk, P., C. Davis, P. Miller, and J. Naglieri (2008): "Exercise and Children's Intelligence, Cognition, and Academic Achievement," *Educational Psychology Review*, 20(2), 111–131.

Tremarche, P. V., E. M. Robinson, and L. B. Graham (2007): "Physical Education and its Effect on Elementary Testing Results," *The Physical Educator*, 64(2), 58–64.

Trudeau, F., and R. J. Shephard (2008): "Physical Education, School Physical Activity, School Sports and Academic Performance," *International Journal of Behavioral Nutrition and Physical Activity*, 5(10), 1–12.

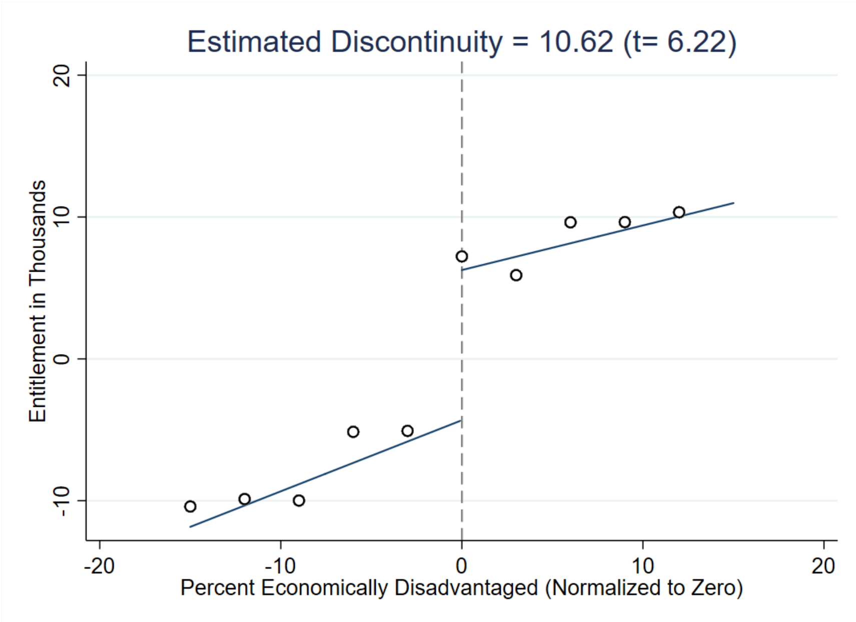
U.S. Department of Education (2013): "Carol M. White Physical Education Program," Accessed 23-November-2018 at https://www2.ed.gov/programs/whitephysed/funding.html.

von Hippel, P., and W. Bradbury (2015): "The Effects of School Physical Education Grants on Obesity, Fitness, and Academic Achievement," *Preventative Medicine*, 78, 44–51.

Wang, L., Q. Yang, R. Lowry, and H. Wechsler (2003): "Economic Analysis of a School-Based Obesity Prevention Program," *Obesity*, 11(11), 1313–1324.

Waters, E., A. de Silva-Sanigorski, B. Bedford, T. Brown, K. Campbell, Y. Gao, R. Armstrong, L. Prosser, and C. Summerbell (2011): "Interventions for Preventing Obesity in Children," *Cochrane Database of Systematic Reviews*, Accessed 2-October-2017 at http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001871.pub3/abstract.

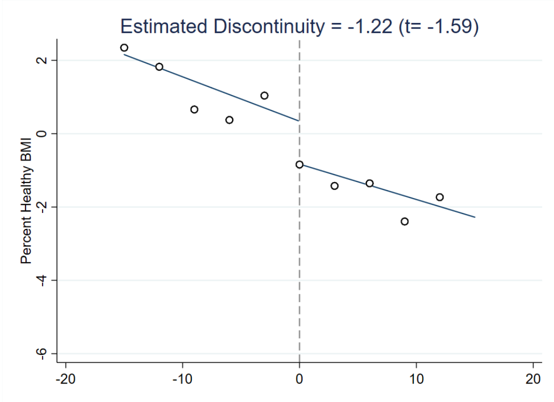
Figure 1: The Effect of Eligibility on Funding



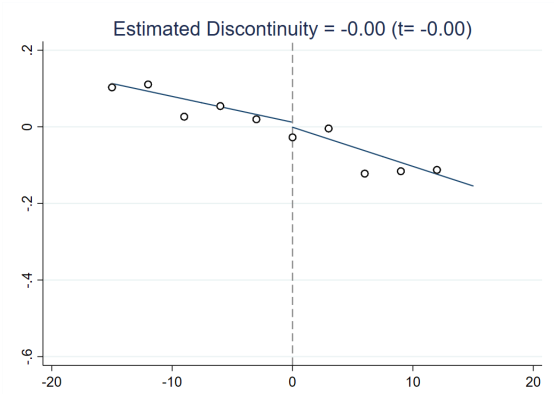
Notes: Funding data for the Texas Fitness Now (TFN) program from 2007-2011 is from the Texas Education Agency, grants division. Entitlement is calculated as the total grant allowance per school year. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes all Texas schools with students in grades 6, 7, and/or 8.

Figure 2: The Effect of Texas Fitness Now on Physical Fitness

Percent of Students with a Healthy BMI



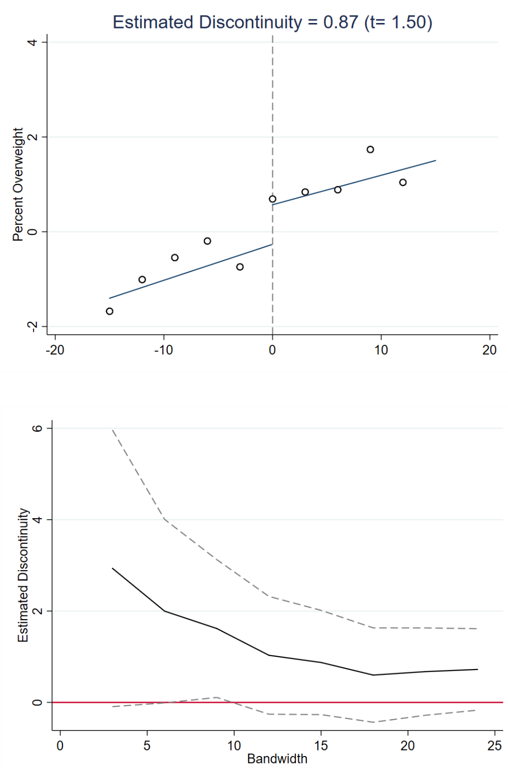
Average Number of Fitness Tests Passed



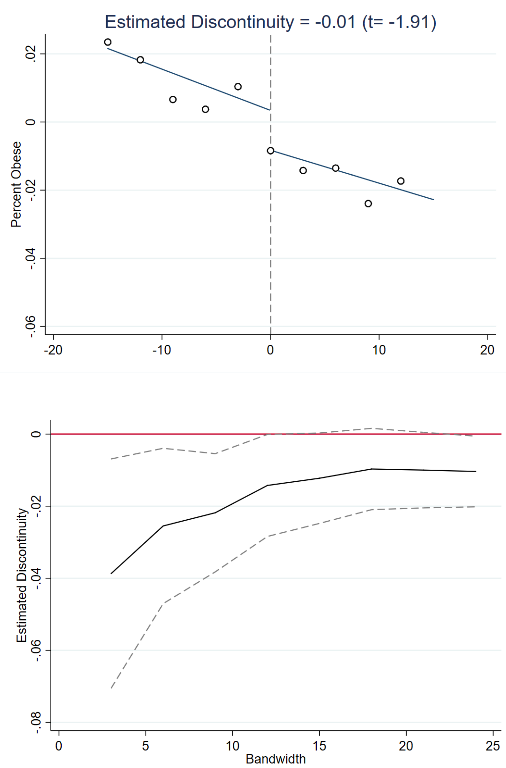
Notes: School-level data on fitness outcomes is from FITNESSGRAM© data provided by the Texas Education Agency (TEA). Each figure plots means of residuals (after differencing out year fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes students in Texas in grades 6, 7, and/or 8 from school years spanning 2007-2011.

Figure 3: Analyzing Changes in BMI for Overweight and Obese Students

Percent of Overweight Students



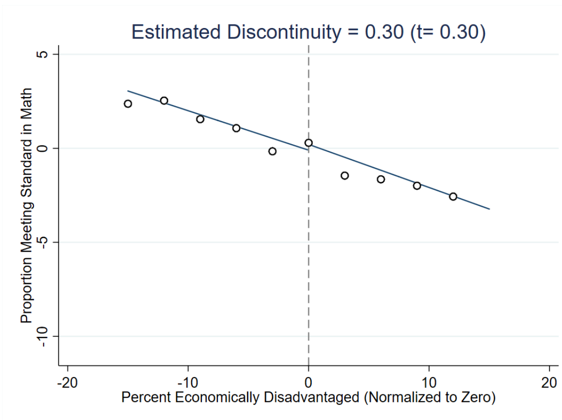
Percent of Obese Students



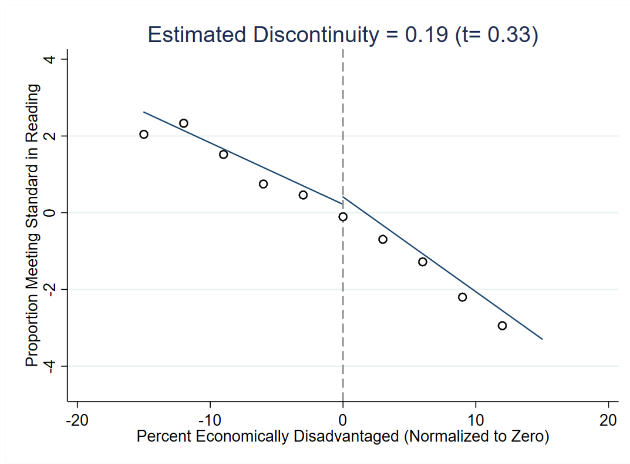
Notes: School-level data on fitness outcomes is from FITNESSGRAM© data provided by the Texas Education Agency (TEA). Each figure plots means of residuals (after differencing out year fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes students in Texas in grades 6, 7, and/or 8 from school years spanning 2007-2011.

Figure 4: The Effect of Texas Fitness Now on Test Scores

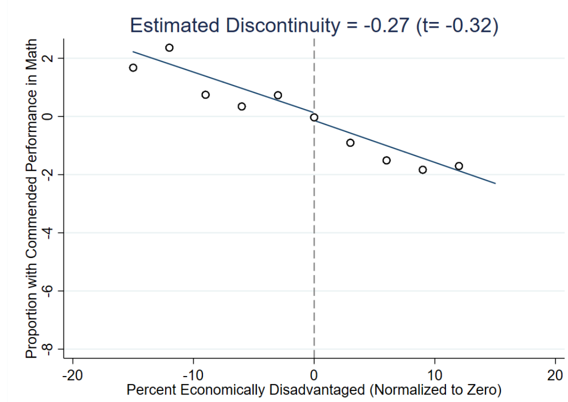
Pass Rate Math TAKS Test



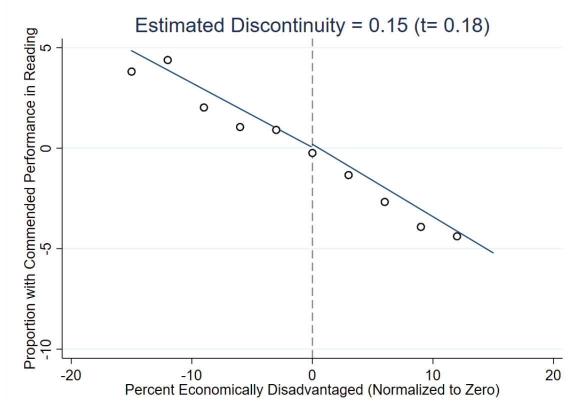
Pass Rate Reading TAKS Test



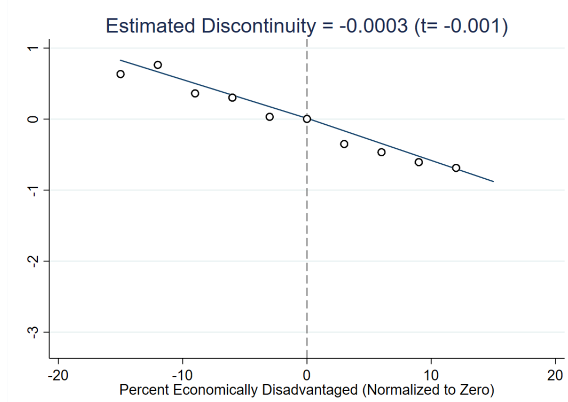
Commended Performance Math TAKS Test



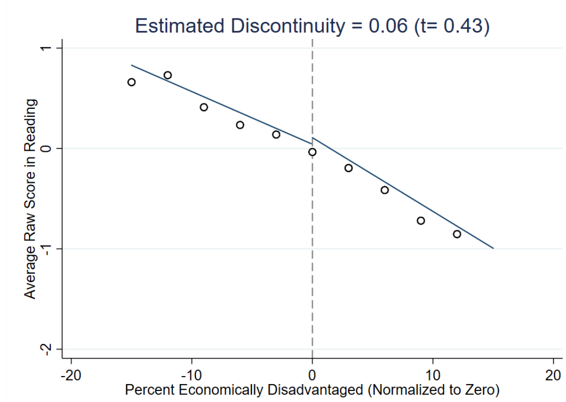
Commended Performance Reading TAKS Test



Raw Score Math TAKS Test



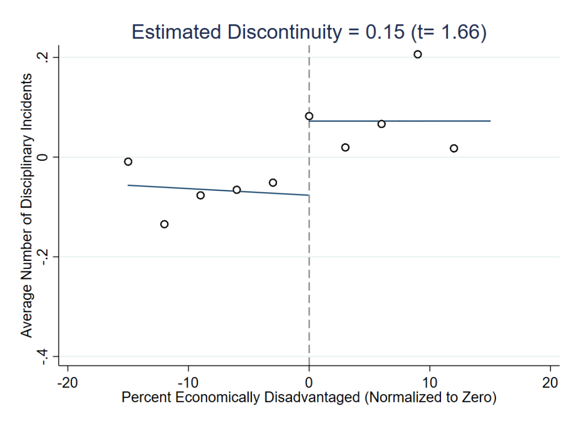
Raw Score Reading TAKS Test



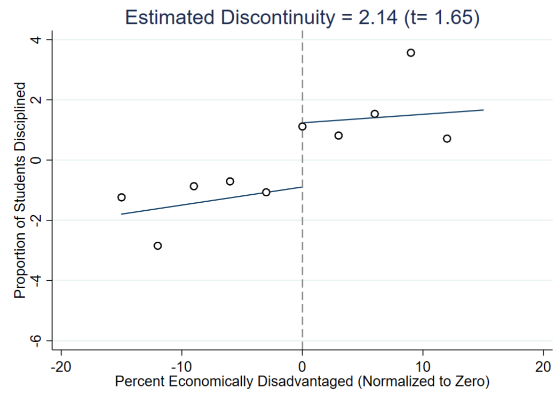
Notes: Student-level data on test scores is from the Education Research Center at UT-Austin. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and 8 from school years spanning 2007-2011.

Figure 5: The Effect of Texas Fitness Now on Disciplinary Action

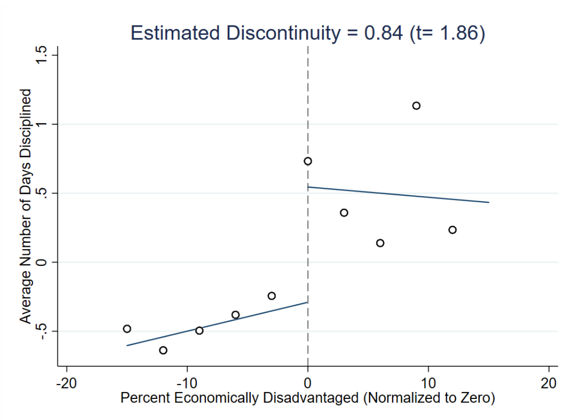
Total Disciplinary Actions



Proportion of Students with Disciplinary Action



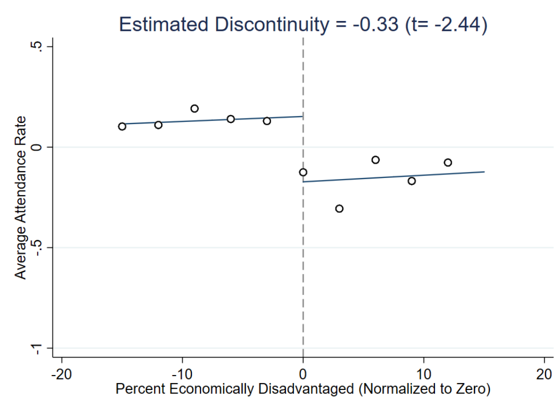
Days Suspended



Notes: Student-level data on disciplinary outcomes is from the Education Research Center at UT-Austin. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and 8 from school years spanning 2007-2011.

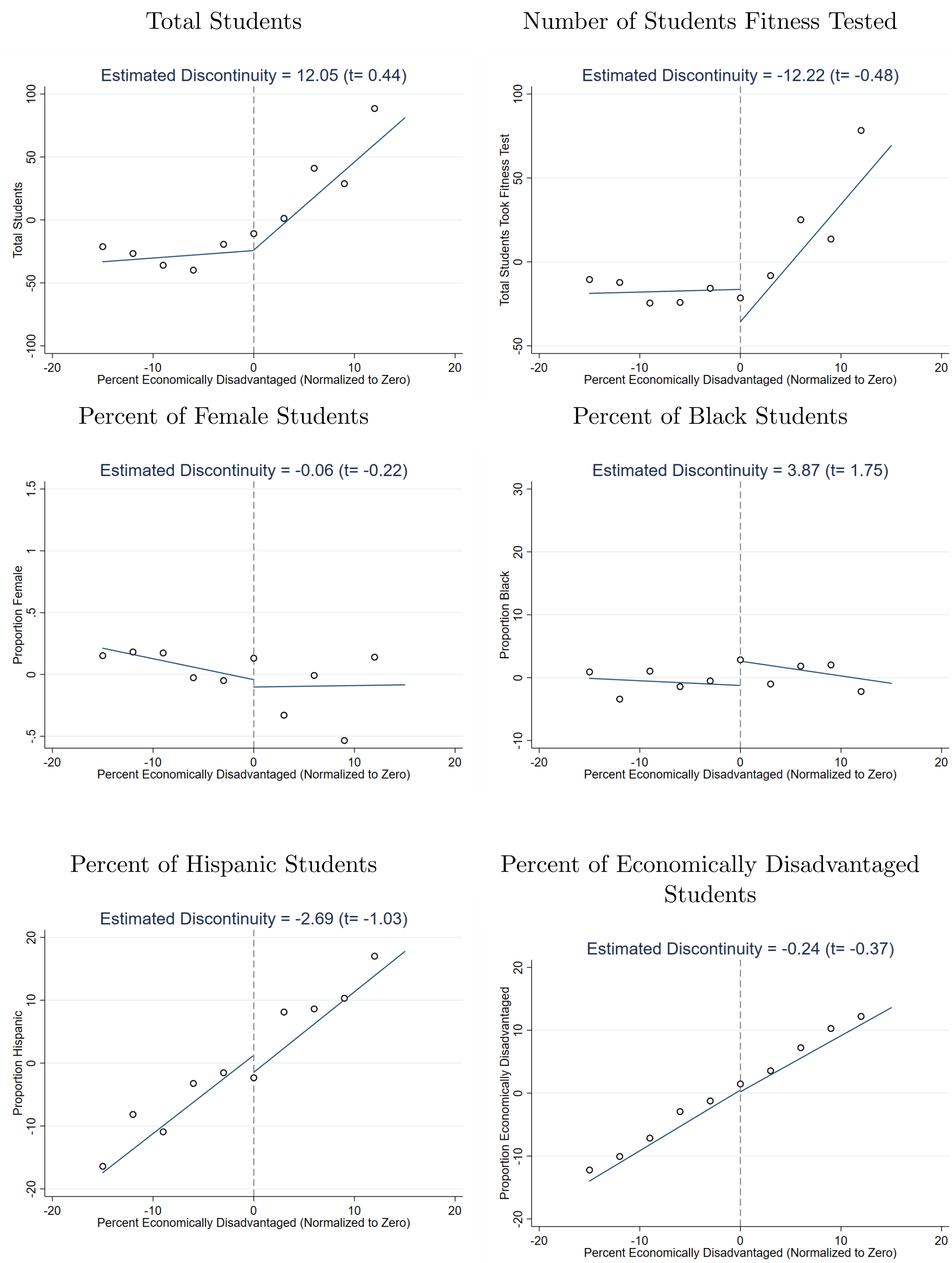
Figure 6: The Effect of Texas Fitness Now on Attendance

Attendance Rate for All Students



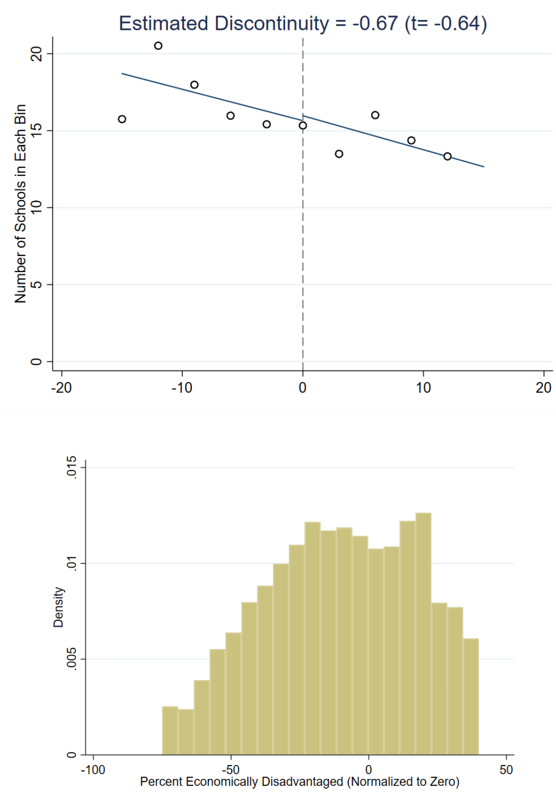
Notes: Student-level data on attendance is from the Education Research Center at UT-Austin. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and 8 from school years spanning 2007-2011.

Figure 7: Testing Discontinuity of School Composition



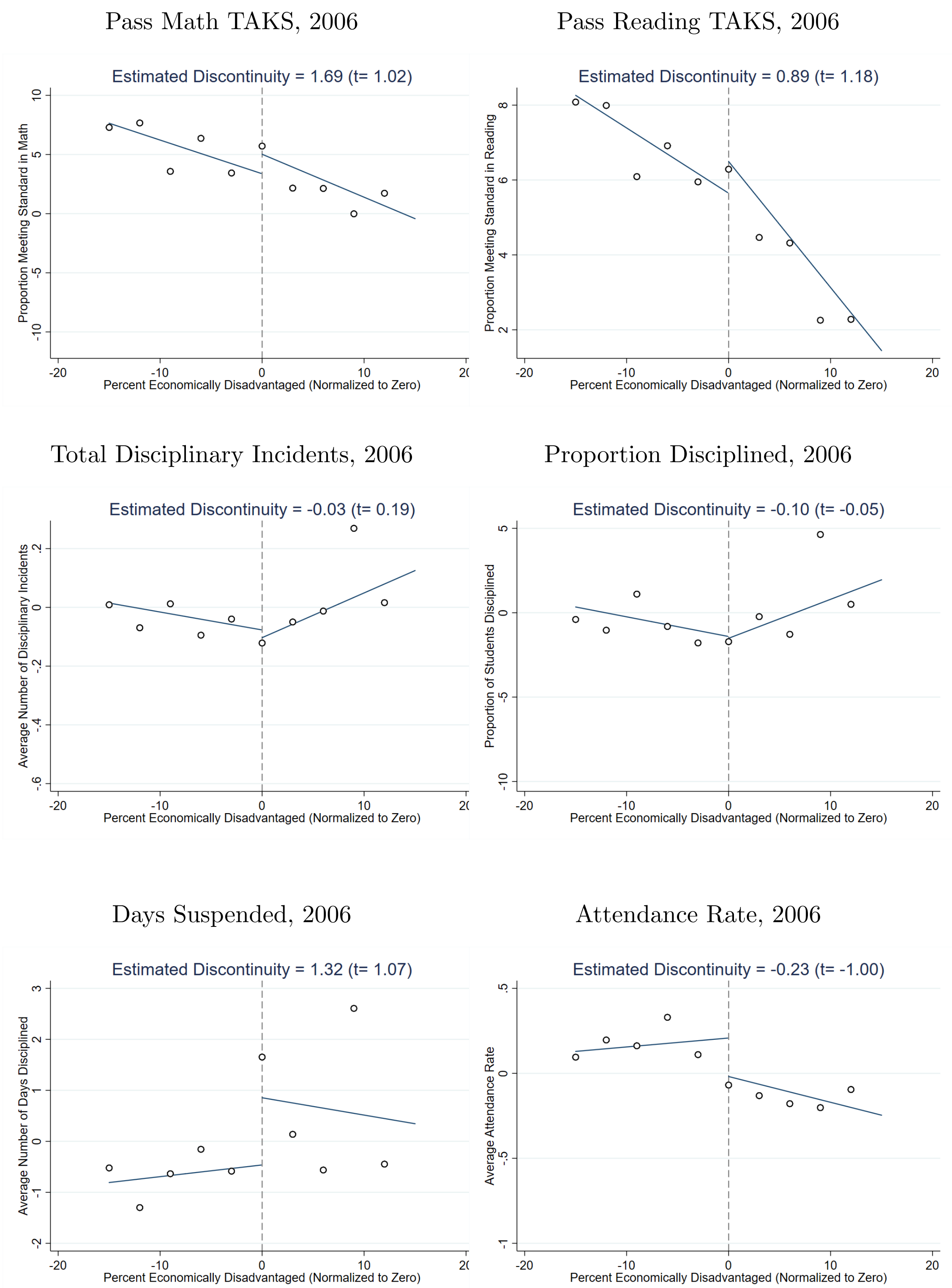
Notes: Data on school characteristics is from the Education Research Center at UT-Austin. Data on the total number of students fitness tested is from the Texas Education Agency's Academic Excellence Indicator System. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and 8 from school years spanning 2007-2011.

Figure 8: Testing the Density of Number of Bins



Notes: Data on student characteristics is from the Education Research Center at UT-Austin. Data on school characteristics is from the Texas Education Agency's Academic Excellence Indicator System. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and/or 8 from school years spanning 2007-2011.

Figure 9: Testing Discontinuities in the Pre-Period



Notes: Student-level data on disciplinary action, attendance rates, and TAKS scores is from the Education Research Center at UT-Austin. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcomes listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes all students in Texas schools in grades 6, 7, and 8 from the 2006-2007 school year.

Table 1: Texas Fitness Now Funding Schedule

School Year	ED Cutoff	Schools Eligible	Amount Granted
2007-2008	75%	605	$10,000,000
2008-2009	75%	575	$9,378,914
2009-2010	60%	981	$8,875,670
2010-2011	60%	1125	$8,500,000

Notes: Data on TFN funding and grantee awards is from the Texas Education Agency, Grants Division. ED cutoff represents the percent of economically disadvantaged students required in the previous year to be eligible for TFN funding. Total funding is approximately $37 million and average funding per school is $11,000.

Table 2: Effects of Texas Fitness Now on Physical Fitness

	Healthy BMI			Number of Tests Passed		
	(1)	(2)	(3)	(4)	(5)	(6)
%ED > Cutoff	-2.19*	-1.42	-1.22	0.00	0.01	-0.00
	(1.19)	(0.88)	(0.77)	(0.07)	(0.05)	(0.04)
Bandwidth	6.9	12	15	5.8	12	15
Observations	1591	2769	3473	1378	2840	3555

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. School-by-grade data from the FITNESSGRAM© test for school years spanning 2007-2011 is from the Texas Education Agency. Each coefficient is generated by a separate regression of Equation 1 using the listed fitness outcome as the dependent variable, controlling for year fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction. The sample includes Texas students in grades 6, 7, or 8.

Table 3: Effects of Texas Fitness Now on Standardized Test Scores

	Math TAKS			Reading TAKS		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Pass Test						
%ED > Cutoff	0.005	0.006	0.004	0.011	0.005	0.0004
	(0.008)	(0.011)	(0.010)	(0.012)	(0.008)	(0.007)
Bandwidth	8.9	12	15	11.2	12	15
Observations	737,503	1,002,403	1,289,442	923,137	338,327	431,677
Panel B. Commended Performance						
%ED > Cutoff	0.008	0.001	-0.003	0.0003	0.002	0.002
	(0.009)	(0.010)	(0.008)	(0.010)	(0.010)	(0.008)
Bandwidth	10.9	12	15	8.2	12	15
Observations	893,230	1,002,403	1,289,442	674,118	1,002,373	1,289,364
Panel C. Raw Score						
%ED > Cutoff	0.152	0.090	0.109	0.216	0.172	0.093
	(0.248)	(0.315)	(0.278)	(0.326)	(0.206)	(0.178)
Bandwidth	8.0	12	15	11.0	12	15
Observations	663,142	999,023	1,285,172	905,681	998,993	1,285,094

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Student-level testing data for school years spanning 2007-2011 is from Education Research Center at UT-Austin. Each coefficient is generated by a separate regression of Equation 1 using the listed academic performance outcome as the dependent variable, controlling for year and grade fixed effects. A student passes an exam if they meet the standards for the test for that year. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction. The sample includes all students in Texas students in grades 6, 7, or 8.

Table 4: Effects of Texas Fitness Now on Disciplinary Action

	Total Disciplinary Incidents			Proportion of Students Disciplined			Total Days Suspended		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
%ED > Cutoff	0.080	0.075	0.149*	0.005	0.012	0.021*	0.703	0.616	0.836*
	(0.133)	(0.101)	(0.090)	(0.019)	(0.015)	(0.013)	(0.580)	(0.541)	(0.451)
Bandwidth	7.8	12	15	7.4	12	15	10.1	12	15
Observations	656,604	1,010,648	1,299,744	624,046	1,010,648	1,299,744	832,261	1,010,648	1,299,744

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Student-level data for school years spanning 2007-2011 is from Education Research Center at UT-Austin. Each coefficient is generated by a separate regression of Equation 1 using the listed discipline outcome as the dependent variable, controlling for year and grade fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in a given year. The sample includes Texas students in grades 6, 7, or 8.

Table 5: Effects of Texas Fitness Now on Attendance

	(1)	(2)	(3)
%ED > Cutoff	-0.002	-0.003**	-0.003**
	(0.002)	(0.002)	(0.001)
Bandwidth	8.9	12	15
Observations	750,912	1,008,485	1,297,023

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Student-level data for school years spanning 2007-2011 is from Education Research Center at UT-Austin. Each coefficient is generated by a separate regression of Equation 1, controlling for year and grade fixed effects. Student-level attendance rates are calculated by dividing the total number of days students were present by the total number of school days. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in a given year. The sample includes Texas students in grades 6, 7, or 8.

Table 6: Testing Alternative Specifications

	Linear Fit (1)	Quad Fit (2)	Cubic Fit (3)	Triangular Kernel (4)
Panel A. Pass Math TAKS				
%ED > Cutoff	0.003	0.005	0.009	0.004
	(0.006)	(0.010)	(0.012)	(0.007)
Observations	1,289,364	1,289,364	1,289,364	1,289,364
Panel B. Pass Reading TAKS				
%ED > Cutoff	0.004	0.013	0.027	0.008
	(0.005)	(0.009)	(0.010)	(0.001)
Observations	1,289,442	1,289,442	1,289,442	1,289,442
Panel C. Total Disciplinary Incidents				
%ED > Cutoff	0.149*	0.012	0.099	0.092
	(0.069)	(0.127)	(0.148)	(0.010)
Observations	1,299,744	1,299,744	1,299,744	791,258
Panel D. Proportion Disciplined				
%ED > Cutoff	0.0214*	0.002	0.012	0.013
	(0.013)	(0.021)	(0.028)	(0.015)
Observations	1,299,744	1,299,744	1,299,744	1,299,744
Panel E. Days Suspended				
%ED > Cutoff	0.836*	0.412	0.752	0.666
	(0.451)	(0.670)	(0.746)	(0.512)
Observations	1,299,744	1,299,744	1,299,744	1,299,744
Panel F. Attendance Rate				
%ED > Cutoff	-0.003**	-0.002	0.001	-0.003*
	(0.001)	(0.002)	(0.002)	(0.001)
Observations	1,297,023	1,297,023	1,297,023	1,297,023

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Individual-level data on Texas middle school students from 2007-2011 is from the Education Research Center at UT-Austin. Each coefficient is generated by a separate regression of Equation 1 using the listed outcome as the dependent variable. Each regression includes year and grade fixed effects and reports results from a full one-sided bandwidth of 15. Column 1 replicates the baseline results for comparison. Columns 2 and 3 allow for the days from the cutoff to vary quadratically and cubically (in addition to on either side of the threshold), respectively. Column 4 fits the model using a triangular kernel instead of uniform kernel. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction.

Appendix

Figure A1: Healthy Fitness Zone Standards

BOYS

	Aerobic Capacity VO_{2max} (ml/kg/min) PACER, One Mile Run & Walk Test			Percent Body Fat				Body Mass Index			
	NI-Health Risk	NI	HFZ	Very Lean	HFZ	NI	NI-Health Risk	Very Lean	HFZ	NI	NI-Health Risk
5	*Completion of test. Lap count or time standards not recommended.*			≤8.8	8.9-18.8	18.9	≥27.0	≤13.8	13.9-16.7	16.8	≥17.5
6				≤8.4	8.5-18.8	18.9	≥27.0	≤13.7	13.8-16.9	17.0	≥17.8
7				≤8.2	8.3-18.8	18.9	≥27.0	≤13.7	13.8-17.3	17.4	≥18.3
8				≤8.3	8.4-18.8	18.9	≥27.0	≤13.8	13.9-17.8	17.9	≥19.0
9				≤8.6	8.7-20.6	20.7	≥30.1	≤14.0	14.1-18.5	18.6	≥19.9
10	≤37.3	37.4-40.1	≥40.2	≤8.8	8.9-22.4	22.5	≥33.2	≤14.2	14.3-18.9	19.0	≥20.8
11	≤37.3	37.4-40.1	≥40.2	≤8.7	8.8-23.6	23.7	≥35.4	≤14.5	14.6-19.7	19.8	≥21.8
12	≤37.6	37.7-40.2	≥40.3	≤8.3	8.4-23.6	23.7	≥35.9	≤15.0	15.1-20.5	20.6	≥22.7
13	≤38.6	38.7-41.0	≥41.1	≤7.7	7.8-22.8	22.9	≥35.0	≤15.4	15.5-21.3	21.4	≥23.6
14	≤39.6	39.7-42.4	≥42.5	≤7.0	7.1-21.3	21.4	≥33.2	≤16.0	16.1-22.1	22.2	≥24.5
15	≤40.6	40.7-43.5	≥43.6	≤6.5	6.6-20.1	20.2	≥31.5	≤16.5	16.6-22.9	23.0	≥25.3
16	≤41.0	41.1-44.0	≥44.1	≤6.4	6.5-20.1	20.2	≥31.6	≤17.1	17.2-23.7	23.8	≥26.0
17	≤41.2	41.3-44.1	≥44.2	≤6.6	6.7-20.9	21.0	≥33.0	≤17.7	17.8-24.4	24.5	≥26.7
>17	≤41.2	41.3-44.2	≥44.3	≤6.9	7.0-22.2	22.3	≥35.1	≤18.2	18.3-25.1	25.2	≥27.5

	Curl-up *# completed*	Trunk Lift *inches*		Push-up # completed	Modified Pull-up # completed	Flexed Arm Arm Hang seconds	Back Saver Sit & Reach** *inches*	Shoulder Stretch
5	≥2	6	12	≥3	≥2	≥2	8	Healthy Fitness Zone = Touching fingertips together behind the back on both right and left sides
6	≥2	6	12	≥3	≥2	≥2	8	
7	≥4	6	12	≥4	≥3	≥3	8	
8	≥6	6	12	≥5	≥4	≥3	8	
9	≥9	6	12	≥6	≥5	≥4	8	
10	≥12	9	12	≥7	≥5	≥4	8	
11	≥15	9	12	≥8	≥6	≥6	8	
12	≥18	9	12	≥10	≥7	≥10	8	
13	≥21	9	12	≥12	≥8	≥12	8	
14	≥24	9	12	≥14	≥9	≥15	8	
15	≥24	9	12	≥16	≥10	≥15	8	
16	≥24	9	12	≥18	≥12	≥15	8	
17	≥24	9	12	≥18	≥14	≥15	8	
17+	≥24	9	12	≥18	≥14	≥15	8	

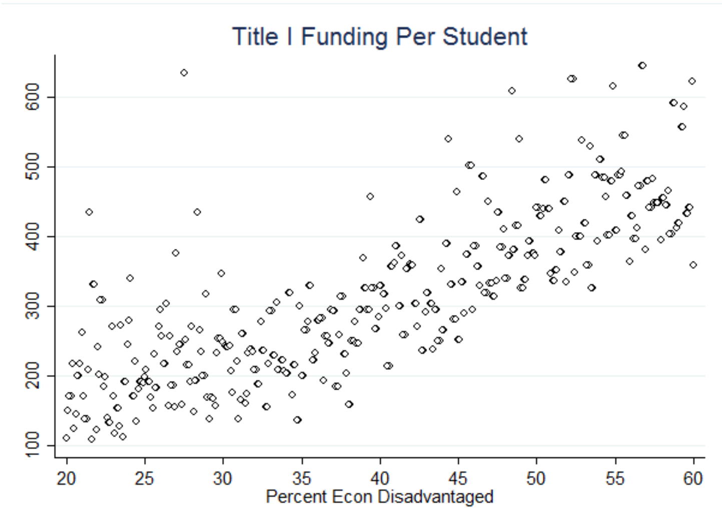
GIRLS

	Aerobic Capacity VO_{2max} (ml/kg/min) PACER, One Mile Run & Walk Test			Percent Body Fat				Body Mass Index			
	NI-Health Risk	NI	HFZ	Very Lean	HFZ	NI	NI-Health Risk	Very Lean	HFZ	NI	NI-Health Risk
5	*Completion of test. Lap count or time standards not recommended.*			≤9.7	9.8-20.8	20.9	≥28.4	≤13.5	13.6-16.7	16.8	≥17.3
6				≤9.8	9.9-20.8	20.9	≥28.4	≤13.4	13.5-17.0	17.1	≥17.7
7				≤10.0	10.1-20.8	20.9	≥28.4	≤13.4	13.5-17.5	17.6	≥18.3
8				≤10.4	10.5-20.8	20.9	≥28.4	≤13.5	13.6-18.2	18.3	≥19.1
9				≤10.9	11.0-22.6	22.7	≥30.8	≤13.7	13.8-18.9	19.0	≥20.0
10	≤37.3	37.4-40.1	≥40.2	≤11.5	11.6-24.3	24.4	≥33.0	≤14.0	14.1-19.5	19.6	≥21.0
11	≤37.3	37.4-40.1	≥40.2	≤12.1	12.2-25.7	25.8	≥34.5	≤14.4	14.5-20.4	20.5	≥21.9
12	≤37.0	37.1-40.0	≥40.1	≤12.6	12.7-26.7	26.8	≥35.5	≤14.8	14.9-21.2	21.3	≥22.9
13	≤36.6	36.7-39.6	≥39.7	≤13.3	13.4-27.7	27.8	≥36.3	≤15.3	15.4-22.0	22.1	≥23.8
14	≤36.3	36.4-39.3	≥39.4	≤13.9	14.0-28.5	28.6	≥36.8	≤15.8	15.9-22.8	22.9	≥24.6
15	≤36.0	36.1-39.0	≥39.1	≤14.5	14.6-29.1	29.2	≥37.1	≤16.3	16.4-23.5	23.6	≥25.4
16	≤35.8	35.9-38.8	≥38.9	≤15.2	15.3-29.7	29.8	≥37.4	≤16.8	16.9-24.1	24.2	≥26.1
17	≤35.7	35.8-38.7	≥38.8	≤15.8	15.9-30.4	30.5	≥37.9	≤17.2	17.3-24.6	24.7	≥26.7
>17	≤35.3	35.4-38.5	≥38.6	≤16.4	16.5-31.3	31.4	≥38.6	≤17.5	17.6-25.1	25.2	≥27.2

	Curl-up *# completed*	Trunk Lift *inches*		90^0 Push-up # completed	Modified Pull-up # completed	Flexed Arm Arm Hang seconds	Back Saver Sit & Reach** *inches*	Shoulder Stretch
5	≥2	6	12	≥3	≥2	≥2	9	Healthy Fitness Zone = Touching fingertips together behind the back on both right and left sides
6	≥2	6	12	≥3	≥2	≥2	9	
7	≥4	6	12	≥4	≥3	≥3	9	
8	≥6	6	12	≥5	≥4	≥3	9	
9	≥9	6	12	≥6	≥4	≥4	9	
10	≥12	9	12	≥7	≥4	≥4	9	
11	≥15	9	12	≥7	≥4	≥6	10	
12	≥18	9	12	≥7	≥4	≥7	10	
13	≥18	9	12	≥7	≥4	≥8	10	
14	≥18	9	12	≥7	≥4	≥8	10	
15	≥18	9	12	≥7	≥4	≥8	12	
16	≥18	9	12	≥7	≥4	≥8	12	
17	≥18	9	12	≥7	≥4	≥8	12	
17+	≥18	9	12	≥7	≥4	≥8	12	

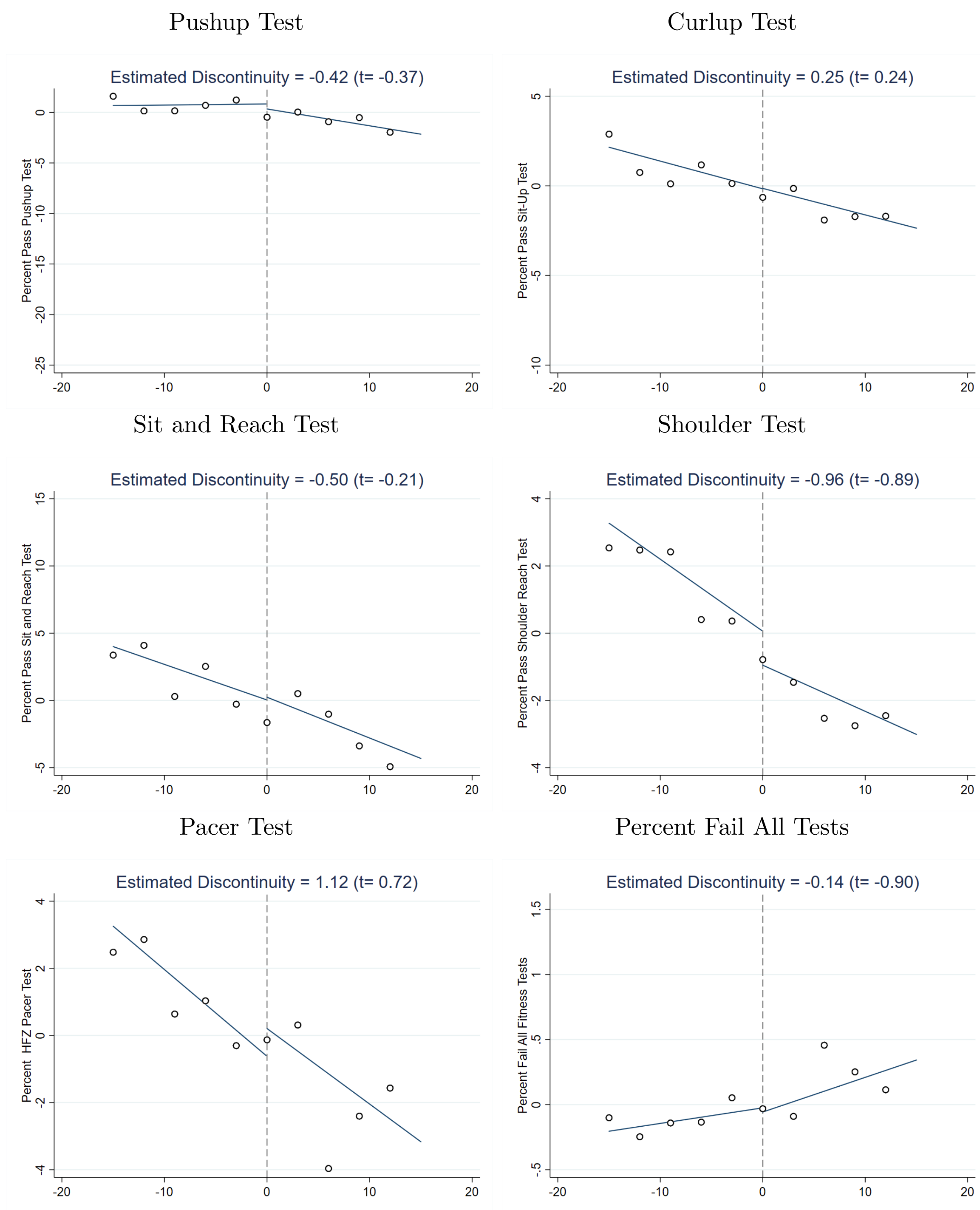
Notes: Data on FITNESSGRAM© standards for Healthy Fitness Zone are from the Cooper Institute. See `http://www.cooperinstitute.org/healthyfitnesszone` for more information.

Figure A2: Title 1 Funding by Percent Economically Disadvantaged



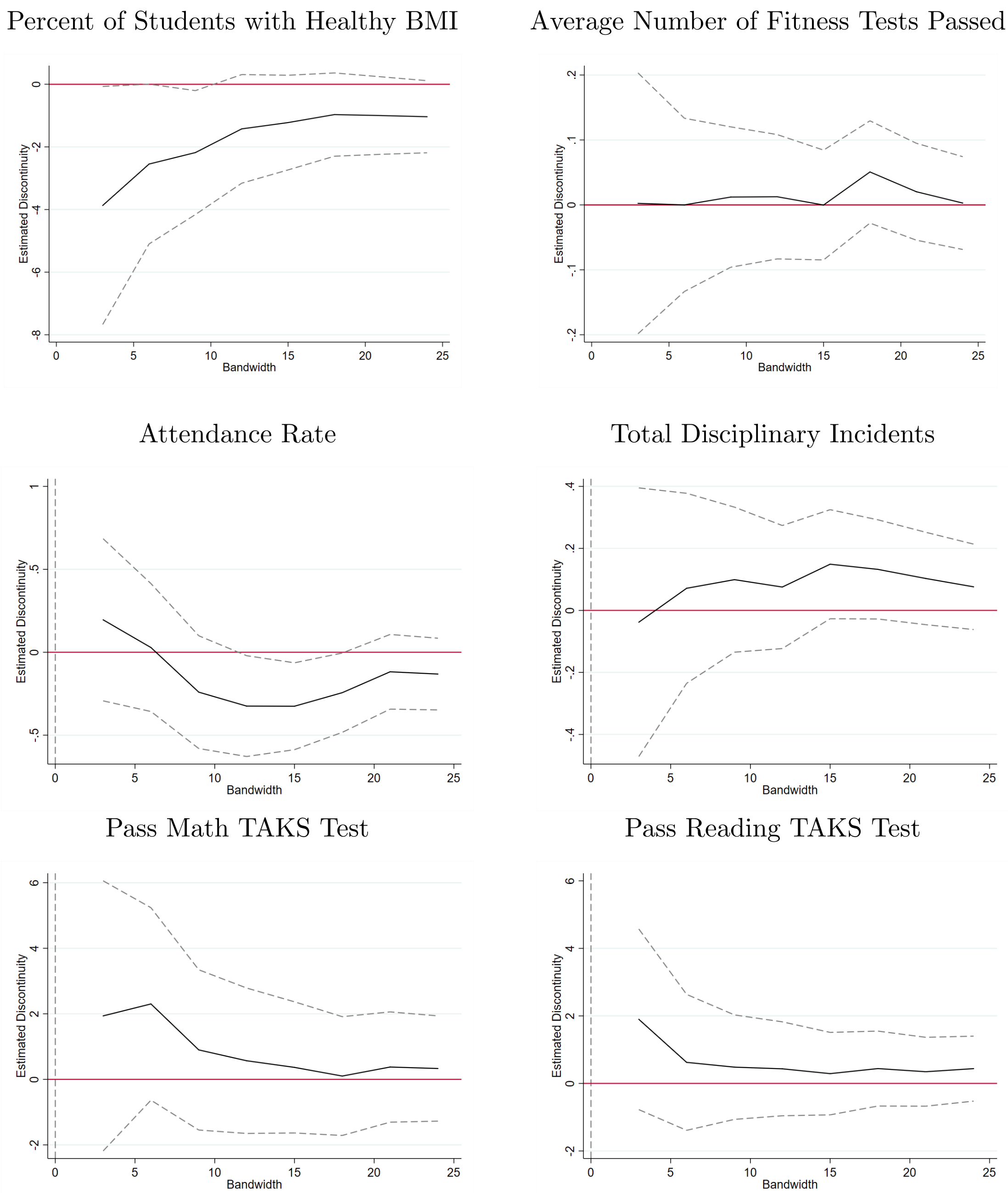
Notes: Data on district-level Title 1 funding by funding source are from the Texas Education Agency Public Education Information Management Systems Reports. Title I funds are aimed at schools with at least 40% economically disadvantaged students.

Figure A3: The Effect of Texas Fitness Now on Physical Fitness



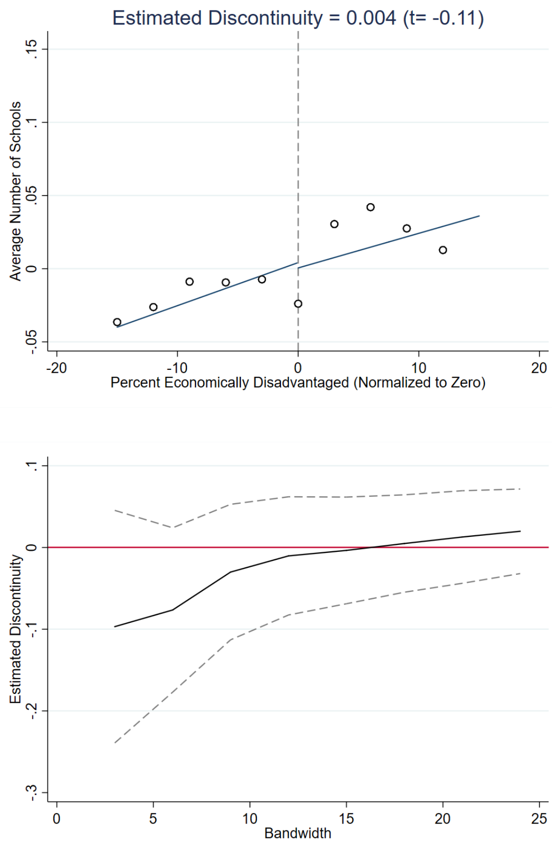
Notes: School-level data on fitness outcomes is from FITNESSGRAM© data provided by the Texas Education Agency (TEA). Each figure plots means of residuals (after differencing out year fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes students in Texas in grades 6, 7, and/or 8 from school years spanning 2007-2011.

Figure A4: Effect of Varying Bandwidth on Estimates



Notes: School-level data on BMI and physical fitness is from FITNESSGRAM© data provided by the Texas Education Agency (TEA). Individual-level data on test scores, discipline, and attendance is from the Education Research Center at UT-Austin. Each panel reports estimates and their corresponding 95% confidence intervals from linear regressions, using uniform kernel weights and allowing the slopes to vary on each side of the threshold, for a range of different bandwidths. The sample includes Texas students in grades 6, 7, or 8 from school years spanning 2007-2011.

Figure A5: Testing Student Attrition



Notes: Individual-level data on school enrollment is from the Education Research Center at UT-Austin. The sample includes Texas students in grades 6, 7, and 8 from school years 2007-2008 to 2010-2011.

Table A1: Summary Statistics

	Mean	St. Dev.	Min	Max
School Characteristics				
Total Number of Students Enrolled	515	337	1	1,816
Amount Entitled by Texas Fitness Now Grant	5,024.02	9,706.64	0	62,442
Percent Economically Disadvantaged	70.6	45.6	0	100
Percent Female	48.6	50.0	0	100
Percent White	22.4	26.18	0	100
Percent Black	17.0	37.6	0	100
Percent Hispanic	57.4	49.5	0	100
Charter School	0.04	0.19	0	1
Health and Fitness Outcomes				
Percent Healthy BMI	63.41	12.06	0	100
Percent Healthy Body Fat	73.77	21.50	0	100
Percent Pass Pacer Test	58.01	22.62	0	100
Percent Complete Mile Run	60.82	23.61	0	100
Percent Pass Push-Up Test	73.57	16.67	0	100
Percent Pass Curl up Test	79.49	16.68	0	100
Percent Pass Sit and Reach Test	64.56	25.00	0	100
Percent Pass Shoulder Test	72.44	13.66	0	100
Percent Pass All Fitness Tests	22.87	15.77	0	88
Percent Fail All Fitness Tests	1.08	2.52	0	71
Academic Outcomes				
Math TAKS Passing Rate	0.71	0.45	0	1
Reading TAKS Passing Rate	0.83	0.38	0	1
Math TAKS Commended Rate	0.20	0.40	0	1
Reading TAKS Commended Rate	0.33	0.47	0	1
Math TAKS Raw Score	31.70	11.91	0	50
Reading TAKS Raw Score	35.24	11.74	0	48
Total Disciplinary Incidents	0.96	0.44	0	97
Proportion of Students Disciplined	0.27	2.65	0	1
Total Days Suspended	3.54	15.59	0	910
Attendance Rate	0.96	0.05	0.01	1

Notes: Individual-level data on student characteristics and academic outcomes, including economically disadvantaged status, race, ethnicity, test scores, discipline, and attendance are from the Education Research Center at UT-Austin. Data on fitness outcomes are from the standardized fitness testing program, FITNESSGRAM©, are from the Texas Education Agency (TEA). Texas Fitness Now grant entitlements data are from the publicly-available list of grantee awards provided by the TEA. Entitlements per student for each school are calculated using the total amount of funding divided by enrollment. The sample includes Texas students in grades 6, 7, or 8 from school years spanning 2007-2011.

Table A2: Effects of Funding Cuts on Physical Fitness- Females

	Healthy BMI			Number of Tests Passed		
	(1)	(2)	(3)	(4)	(5)	(6)
%ED > Cutoff	-0.92	-0.59	-0.31	-0.28	-0.12	0.05
	(1.56)	(1.14)	(1.03)	(0.20)	(0.17)	(0.15)
Bandwidth	7.6	12	15	7.7	12	15
Observations	1455	2762	3454	1774	2762	3454
	Pacer Test			Mile Run		
%ED > Cutoff	1.49	1.61	1.36	3.62	2.16	1.60
	(2.01)	(1.91)	(1.72)	(3.76)	(2.73)	(2.45)
Bandwidth	11.0	12	15	6.9	12	15
Observations	1617	1783	2265	689	1278	1578
	Push-Up Test			Curl Up Test		
%ED > Cutoff	-0.56	0.47	1.07	0.01	-0.59	0.03
	(2.10)	(1.54)	(1.40)	(1.93)	(1.32)	(1.21)
Bandwidth	6.5	12	15	5.8	12	15
Observations	1369	2560	3193	1302	2753	3443
	Sit and Reach			Shoulder Stretch		
%ED > Cutoff	-1.64	-1.10	-0.44	0.96	0.54	0.11
	(3.34)	(2.87)	(2.59)	(1.31)	(1.20)	(1.08)
Bandwidth	9.8	12	15	10.2	12	15
Observations	1120	1419	1780	1540	1811	2265

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. School-by-grade data from the FITNESSGRAM© test for school years spanning 2007-2011 is from the Texas Education Agency. Each coefficient is generated by a separate regression of Equation 1 using the listed fitness outcome as the dependent variable, controlling for year fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction. The sample includes all female Texas students in grades 6, 7, or 8.

Table A3: Effects of Texas Fitness Now on Academic Outcomes, by Subgroup

	All (1)	Female (2)	Economically Disadvantaged (3)
Panel A. Pass Math TAKS			
%ED > Cutoff	0.004	0.004	0.003
	(0.010)	(0.011)	(0.011)
Observations	1,289,442	627,388	890,987
Panel B. Pass Reading TAKS			
%ED > Cutoff	0.003	0.003	0.004
	(0.006)	(0.006)	(0.007)
Observations	1,289,364	627,394	891,311
Panel C. Total Disciplinary Incidents			
%ED > Cutoff	0.149*	0.106	0.178*
	(0.090)	(0.069)	(0.103)
Observations	534,882	631,916	918,294
Panel D. Proportion of Students Disciplined			
%ED > Cutoff	0.0214*	0.0170	0.0241*
	(0.013)	(0.012)	(0.016)
Observations	1,299,744	631,916	918,294
Panel E. Number of Days Suspended			
%ED > Cutoff	0.836*	0.512**	1.007*
	(0.451)	(0.257)	(0.516)
Observations	1,299,744	631,916	918,294
Panel F. Attendance Rate			
%ED > Cutoff	-0.003*	-0.003*	-0.004*
	(0.001)	(0.001)	(0.002))
Observations	1,297,023	630,652	916,543
Bandwidth	15	15	15

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Data on test scores, disciplinary action, and attendance rates for 6th, 7th, and 8th graders is from the Education Research Center at UT-Austin for school years spanning 2007-2011. Each coefficient is generated by a separate regression of Equation 1 using the listed outcome as the dependent variable, controlling for year and grade fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction.

Table A4: Effects of Texas Fitness Now on Academic Outcomes, by Grade

	All Grades (1)	6th Grade (2)	7th Grade (3)	8th Grade (4)
Panel A. Pass Math TAKS				
%ED > Cutoff	0.004	0.006	0.002	0.002
	(0.010)	(0.012)	(0.012)	(0.012)
Observations	1,289,442	431,679	433,930	423,833
Panel B. Pass Reading TAKS				
%ED > Cutoff	0.003	0.000	0.006	0.001
	(0.0062)	(0.007)	(0.007)	(0.006)
Observations	1,289,364	431,677	433,926	423,761
Panel C. Total Disciplinary Incidents				
%ED > Cutoff	0.149*	0.113	0.128	0.209*
	(0.090)	(0.075)	(0.110)	(0.118)
Observations	1,299,744	433,046	435,958	430,760
Panel D. Proportion of Students Disciplined				
%ED > Cutoff	0.021*	0.017	0.020	0.027
	(0.013)	(0.012)	(0.016)	(0.015)
Observations	1,299,744	433,046	435,938	430,760
Panel E. Number of Days Suspended				
%ED > Cutoff	0.836*	0.440	0.897	1.202*
	(0.451)	(0.280)	(0.548)	(0.615)
Observations	1,299,744	433,046	435,938	430,760
Panel F. Attendance Rate				
%ED > Cutoff	-0.003**	-0.002**	-0.003**	-0.004**
	(0.001)	(0.001)	(0.002)	(0.002)
Observations	1,297,023	432,117	435,077	429,829
Bandwidth	15	15	15	15

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Data on test scores, disciplinary action, and attendance rates for 6th, 7th, and 8th graders is from the Education Research Center at UT-Austin for school years spanning 2007-2011. Each coefficient is generated by a separate regression of Equation 1 using the listed outcome as the dependent variable, controlling for year fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction.