

IZA DP No. 8234

**Does Participation in 4-H Improve Schooling Outcomes?
Evidence from Florida**

Alfonso Flores-Lagunes
Troy Timko

June 2014

Does Participation in 4-H Improve Schooling Outcomes? Evidence from Florida

Alfonso Flores-Lagunes

*Binghamton University (SUNY)
and IZA*

Troy Timko

Independent Researcher

Discussion Paper No. 8234

June 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Does Participation in 4-H Improve Schooling Outcomes? Evidence from Florida¹

We examine the effect of participation in 4-H, the largest youth development program in the United States, on standardized test scores. We do this by utilizing grade-level longitudinal data on Florida's school districts from the Florida Department of Education combined with 4-H participation statistics from Florida 4-H. Specifically, we analyze the effect of the extent of 4-H participation for third through tenth grade on the mathematics and reading subtests of the Florida Comprehensive Assessment Test (FCAT). We use a difference-in-difference-in-differences (DDD) approach to control for potential confounders of the causal relationship at the level of school districts, grades, and years. Our results indicate that the extent of 4-H participation at the district-grade-year level is positively and significantly related to several measures of performance on the FCAT test.

JEL Classification: I21, I28, H52

Keywords: 4-H program, Florida, standardized test scores

Corresponding author:

Alfonso Flores-Lagunes
Department of Economics
Binghamton University (SUNY)
P.O. Box 6000
Binghamton, NY 13902-6000
USA
E-mail: aflores@binghamton.edu

¹ We thank Marilyn Norman, Nancy Johnson, Joy Jordan and Keith Diem from Florida 4-H for their help in obtaining the 4-H data and in understanding the 4-H organization. We also thank Kim Ward and Alice Raker, Program Specialists at the Florida Department of Education, for their help in acquiring and understanding the Florida Schools Indicator Report data. Useful comments have been provided by Vikesh Amin, Rick Ashley, Carmen Carrion-Flores, Damon Clark, Larry Kenny, Chris Parmeter, Sol Polacheck, Sue Stockly, participants at Binghamton University's labor group, seminar participants at Virginia Tech and The Ohio State University, and participants at the 2010 Southern Economics Association and the 2009 Agricultural and Applied Economics Association Meetings. Any errors or misinterpretations are our responsibility.

4-H is the largest youth development organization in the United States, with over six million young Americans participating in the various programs it offers, such as organized clubs, day camps, and school enrichment programs (4-H 2009).² 4-H has a presence in every county and state in the United States, partnering with universities within the Land Grant University System. This organization traces back its origins to 1902, when it provided a link between public school education and rural communities, and for this reason it is widely regarded as agriculturally-focused (4-H 2009). However, since the 1950s, it has spread into urban areas and its focus has shifted towards life skills development of youth. Today, “4-H’ers participate in fun, hands-on learning activities, supported by the latest research of land-grant universities, that are focused on three areas called mission mandates: science, engineering, and technology; healthy living; and citizenship” (4-H 2009).³ Examples of specific programs are Science Discovery, Public Speaking, Citizenship, Real Money Real World, Health Rocks! and Leadership. Given 4-H’s links to institutions of higher education and its remarkable reach and scope, it is likely that the program has a noticeable effect on America’s youth. Additionally, these same characteristics strategically position 4-H to be a key player in future youth development policies, and may also give the organization the potential of complementing the educational efforts of the public school system.

Perhaps surprisingly given the discussion above, economists have paid little attention to analyzing the 4-H program. A plausible explanation for this is the lack of readily available individual-level data containing 4-H participation and relevant outcomes. To our knowledge, evidence of the effects of the 4-H program has largely taken the form of surveys of opinions and perceptions provided by 4-H members and their families or 4-H volunteers (e.g., Guion 2002) or by comparing outcomes of 4-H

² The name 4-H stands for “Head, Heart, Hands, and Health”. The 4-H pledge reflects these words: “I pledge... my head to clearer thinking, my heart to greater loyalty, my hands to larger service and my health to better living for my club, my community, my country and my world”.

³ 4-H considers itself a Youth Development Program (YDP). In general, YDPs “...are programs that provide opportunities and support to help youth gain the competencies and knowledge they need to meet the increasing challenges they will face as they mature” (Roth et al., 1998). YDP’s goals lie at the heart of the positive youth development approach (see, e.g., Carnegie Council on Adolescent Development, 1989).

participants and non-participants (e.g., Goodwin, Carroll and Oliver 2005), typically finding positive effects on youth.⁴ However, common omissions in this literature are the use of objective measures of student performance and, more importantly, the use of proper controls for systematic differences between participants and non-participants that may confound the causal relationship between 4-H participation and youth's outcomes.⁵

Analyzing the causal link between 4-H participation and youth outcomes is important to assess its effectiveness. This is of interest for several reasons. First, the 4-H program can be seen as an additional input to the formal educational process (e.g., the “educational production function”), since the program is intended to have an effect on the youth development aspects that affect student performance (e.g., student health and behavior). As such, it is important to quantify its effect on the production of education, as measured by academic outcomes. Second, given its wide reach and scope, 4-H is strategically positioned to complement the improvement efforts of public school districts, potentially helping them succeed in the context of intense competition for resources and accountability provided by the No Child Left Behind (NCLB) Act of 2001. Within this background, it is of interest to analyze whether the 4-H program results in noticeable effects on standardized test scores, which is the main measure employed by the NCLB Act for school accountability. Finally, the 4-H program itself has faced challenges in justifying its existence to funders of the program (e.g., Farmer 2010; Murray 2010). Producing reliable estimates of the causal effects of 4-H on objective outcome measures is important in this policy issue.

For these reasons, we examine the effect of 4-H participation on standardized test scores in the state of Florida. Given the non-availability of suitable individual-level data, we collected grade-level longitudinal information on Florida's school districts from the Florida Department of Education and combined it with 4-H participation statistics from Florida 4-H. Specifically, we analyze the effect of the extent of 4-H participation—

⁴ Examples of findings reported in these studies are that 4-H participants are less likely to use drugs, have a better outlook on life, report that they get better grades, and have better leadership and oral skills.

⁵ A notable exception is the longitudinal 4-H Study of Positive Youth Development (e.g., Lerner et al. 2009), described in the next section.

defined as the proportion of students who participate in 4-H—for third through tenth grade on the average scores and passing rates in the mathematics and reading subtests of the Florida Comprehensive Assessment Test (FCAT). In order to control for the likely presence of unobserved confounding factors, we estimate the causal link between the extent of 4-H participation and FCAT scores employing a difference-in-difference-in-differences (DDD) approach. This method allows controlling for unobserved confounders through school district, grade level, and school year fixed effects, as well as interactions among them, thus focusing on the variability in the extent of 4-H participation in district-grade-year cells.

Although 4-H does not explicitly design its programs with the intention of increasing standardized test scores for its participants, there are several reasons to expect 4-H to influence test scores. First, 4-H programs are based upon its three mission mandates (science, engineering and technology; healthy living; and citizenship) and thus they aim to positively engage students in utilizing scientific reasoning, practicing healthy behaviors, and interacting in a positive manner with their peers and mentors (4-H 2009). All of these actions can have positive effects on academic performance. Second, the science, engineering and technology programs are designed with the experiential learning process in mind such that participants are encouraged to think, make choices, and reflect on what they have learned. This type of learning helps engage students in the learning process rather than passively absorbing knowledge (4-H 2007a), which increases the overall interest in academic subjects. Third, encouraging students to make healthy lifestyle choices likely enhances their school performance. Practicing healthy habits promotes healthy growth and helps students resist disease (Silliman 2007; Satcher 2001), likely leading to fewer missed school days. In turn, children who miss a large number of school days achieve poorer academic outcomes (Grossman and Kaestner 1997).⁶ Finally, positive behaviors related to good citizenship and non-cognitive skills in general have

⁶ The effects of healthy choices on test scores are perceivable even in the short-run. For instance, Figlio and Winicki (2005) document that simply changing school feeding programs to provide nutritional advantages to children on test days have effects on test performance.

also been related to better performance on standardized tests and schooling (e.g., Wentzel 1993; Heckman, Stixrud and Urzua 2006).

Our study relates to recent literature analyzing after-school programs (ASPs). There has been an increasing interest in the availability and the effects of ASPs in the United States, both in the policy arena (e.g., Carnegie Council on Adolescent Development 1992) and in academic research circles (e.g., Hollister 2003; Kane 2004; Roth et al. 1998). A widely cited cause for this is the concern for the amount of time that students spend on their own after regular school hours (e.g., Hollister 2003). For example, it has been documented that students without adult supervision after school are more likely to engage in undesirable activities (Azier 2004);⁷ and that shorter school days—resulting in more after school time—are related to higher juvenile property crime (Jacob and Lefgren 2003). Another important cause is the surge in test-based accountability at schools (e.g., Kane 2004) that coincides with the passage of the NCLB Act. Against this backdrop, after-school programs are seen as devices that can simultaneously keep students away from potentially risky behaviors and teach them useful skills. While not completely fitting the blueprint of a typical after-school program (such as those reviewed by Kane 2004), a share of 4-H programs take place after school hours and accomplish similar objectives.⁸

Our results indicate that the extent of 4-H participation at the district-grade-year level is positively and significantly related to several measures of performance on both the mathematics and reading subtests of the FCAT. In particular, in our preferred model, we find that the extent of 4-H participation in the same school year has a statistically significant positive effect on the average score on the reading subtest of the FCAT, as well as on passing rates (level 3 or above out of 5) on both the mathematics and reading subtests. The extent of participation in 4-H is also significantly related to a decrease in

⁷ The undesirable activities include skipping school, use of alcohol or marijuana, stealing, or hurting someone.

⁸ The ASPs reviewed by Kane (2004) are 21st Century Community Learning Centers, The After School Corporation, Extended-Service School Initiative, and San Francisco Beacons Initiative. These are all publicly funded programs and their scale is much smaller than 4-H.

the proportion of students who score in the lowest level of the mathematics subtest, and to an increase in the proportion who score in the highest level of the reading subtest. The effects we uncover appear of a reasonable magnitude as compared to estimates of the causal effect of other inputs into the educational production function (e.g., Krueger 1999). Finally, we also explore whether the effects of 4-H participation appear to accumulate over time, and how these estimated effects vary when considering different characteristics of the district-grade-year cells, such as their relative racial/ethnic composition and their rural or urban nature.

Background

We begin by providing relevant details about the 4-H program, selectively reviewing literature relevant to our study, and providing information about the Florida Comprehensive Assessment Test (FCAT), our outcome of interest.

The 4-H Program

4-H serves over six million young Americans, making it the largest youth development organization in the United States.⁹ As a youth development program, 4-H's mission is to "empower youth to reach their full potential through working and learning in partnership with caring adults". It accomplishes this goal with the help of over 3,500 employees and 518,000 volunteers who deliver a wide array of programs, such as organized clubs, school enrichment programs, and day camps. 4-H programs operate in every county and state in the United States, as well as in more than 80 countries around the world, in partnership with 106 land-grant universities (4-H 2009). Although 4-H traces back its origins in 1902 to an agriculturally-focused program (it started as a link between public school education and rural communities), since the 1950s, it has spread into urban areas (4-H 2009). Today, 30 percent of students involved in 4-H reside in cities with more than 50,000 residents; 22 percent in towns between 10,000 and 50,000 residents; 36 percent in towns with under 10,000 residents; and 12 percent on farms (4-H 2007b).

⁹ As a comparison, 4-H (2007b) reports that the next largest youth development organizations are Boys & Girls Club of America (serving 4.8 million), Boy Scouts of America (2.8 million), Girl Scouts USA (2.6 million), Future Farmers of America (508 thousand), and Big Brothers Big Sisters (255 thousand).

4-H has three mission mandates that are at the center of every program it offers: science, engineering and technology; healthy living; and citizenship. Programs in 4-H take several forms such as school-based, after-school-based, camp settings, and clubs, and span a wide array of topics. Examples of topics include public speaking, leadership, environmental education, citizenship, arts, consumer and family science, foods and nutrition, health, technology, and biological and physical sciences. All of the programs adhere to one of the mission mandates and are centered around 4-H's "fundamental ideal of practical, 'learning by doing' experiences [that] encourage youth to experiment, innovate and think independently" (4-H 2009). In terms of organizational structure, "4-H is operated and supported [includes financially] by a shared leadership of public and private partners, including National 4-H Headquarters; United States Department of Agriculture (USDA) within the Cooperative State Research, Education and Extension Service; Cooperative Extension educators at land-grant universities; National 4-H Council; 4-H associations and foundations; and volunteers" (4-H 2009). The wide array and diversity of stakeholders underscores the importance of evaluations of the effectiveness of 4-H.

The state 4-H programs are independently operated and are associated with the land-grant universities in each state. In Florida, 4-H is part of the Florida Cooperative Extension Service within the University of Florida's Institute of Food and Agricultural Sciences (UF/IFAS) and Florida A&M University; and receives support from combined efforts of the county, state, and federal governments (Norman and Jordan 2006). We will discuss more about funding sources below in the context of our identification strategy. Florida 4-H offers programs in all of its 67 counties, which correspond exactly to the school districts in the state. Importantly, the 4-H county offices and school districts are independent organizations. During the 2007/08 school year, over 263,000 students were involved in Florida 4-H (Florida 4H 2008) at a cost of over \$21 million (IFAS Extension 2007 Annual Report). Of these members, approximately 47 percent live in suburbs and cities with populations greater than 50,000; 51 percent are female; 63.3 percent are in

elementary school; 22.7 percent are in middle school; and 9.9 percent are in high school (Florida 4H 2008).¹⁰

Most of the programs offered by Florida 4-H are delivered through school enrichment programs or organized clubs.¹¹ In school enrichment programs, which account for about 75 percent of programs, 4-H partners with local schools to offer 4-H programs (in-school or after-school). 4-H clubs, which account for about 15 percent of programs, are groups formed by members around a particular topic and are guided by one or more adults (employees or volunteers). Generally, the topics of programs in Florida—regardless of the delivery method—fall into one of ten areas: agricultural literacy, animal sciences, career development and workforce participation, citizenship and leadership education, communication sciences and expressive arts, environmental education and earth sciences, healthy lifestyle education, plant science, and science and technology. All programs offered by Florida 4-H are guided by the same 4-H principles discussed earlier.

Related Literature

Evidence on the effectiveness of the 4-H program in influencing the lives of its student participants has traditionally taken the form of surveys of opinions and perceptions provided by 4-H members and their families or 4-H volunteers, with no information about non-participants. The main purpose of these studies—some of them reviewed in Guion (2002)—is to measure participants' attitudinal and behavioral perceptions about different activities within 4-H. More recently, Goodwin, Carroll and Oliver (2005) provide evidence of the impact of 4-H on students by comparing both 4-H participants and non-participants based on student survey information about their perceptions of their own performance levels. Their study finds that 4-H members are less likely to use drugs, have a better outlook on life, and self-report obtaining better grades than non-4-H

¹⁰ About 4 percent of members are in Kindergarten.

¹¹ Other delivery methods include day camps and special interest clinics, school-age child care, and overnight camping. Together, these represent less than 10 percent of programs.

members. These studies likely fall short from establishing causal links given the absence of proper controls for systematic differences between participants and non-participants.

A study worthy of note is the 4-H Study of Positive Youth Development (see Lerner et al. 2009 and references therein), a longitudinal study that followed children over five years, beginning in 2002. The original sample included 1,719 fifth graders who were surveyed about their behaviors and participation in 4-H and other programs. To control for self-selection into programs, 4-H participants are matched to non-participants using a variety of covariates (Lerner et al. 2009). While the main focus of the study is on the effect of participation in 4-H on different aspects of the theory of Positive Youth Development (e.g., changes in a student’s character, confidence, caring, etc.), estimated effects on some schooling measures were reported. For instance, Lerner et al. (2009) report that “...4-H participants had better grades, were more behaviorally and emotionally engaged with school, and were more likely to see themselves going to college.” In contrast to this study, we focus on district-grade-year cells given our lack of access to individual-level data, and concentrate in the average cell score on a standardized test score (FCAT) representative of the type of objective performance measures employed by the NCLB Act.

A related body of literature pertains to the effectiveness of after-school programs (ASPs). These ASPs are very heterogeneous, varying in their primary objectives, costs of attendance, available methods of delivery, etc. Relative to 4-H, they are typically narrower in scope and scale. However, the findings reported about ASPs can help form our expectations about the potential effects of 4-H. Correspondingly, by examining the overall effect of a more general and widely available program, such as 4-H, we can shed additional light on the effectiveness of ASPs. Kane (2004) reviews and interprets the results of evaluations of four different ASPs.¹² He notes that none of the evaluations reported a statistically significant impact on test scores after one year of participation, and argues that this may be due to a lack of statistical power in those studies—which

¹² The four different ASPs are: 21st Century Community Learning Centers, The After-School Corporation, Extended-Service Schools Initiative, and San Francisco Beacons Initiative.

typically are based on small sample sizes—and to the low “intensity” of participation in the programs. In contrast, he reports that the evaluations find positive effects of ASPs on parental involvement in school and higher student homework completion. Hollister (2003) reviews ten evaluations of ASPs with “rigorous methodology” (defined as based at least partially on randomization), and concludes that there have been some effective programs with positive impacts on in-school outcomes such as test scores, and out-of-school outcomes such as drug and alcohol use or crime. Hollister (2003) finds that programs emphasizing mentoring and tutoring are more effective than those emphasizing remedial skill-building and parental involvement.¹³

Other studies analyze the relationship between after-school time and student outcomes. For instance, Azier (2004) analyzes the effects of having any kind of adult supervision after school on various outcomes of students from ten to fourteen years of age. Using fixed-effects methods to control for unobserved family characteristics, Azier finds that adult supervision has a negative and significant relationship with incidents of risky behavior, such as skipping school, using alcohol or marijuana, stealing, or hurting someone. Jacob and Lefgren (2003) employ exogenous variation in the school calendars provided by teacher-in-service days and information on criminal activity to document that shorter school days—resulting in more after-school time—are related to higher juvenile property crime. Interestingly, they also find that shorter school days are related to lower rates of violent juvenile crime, which they attribute to the decreased level of interaction among students that is provided by schools, which gives fewer opportunities for this type of crime to take place. The findings of these studies relate to 4-H since the program simultaneously provides a setting in which students are guided and mentored by adults and receive positive academic and behavioral skills.

The Florida Comprehensive Assessment Test (FCAT)

¹³ Examples of programs that emphasize mentoring and tutoring are: Big Brothers Big Sisters, Quantum of Opportunities, and Woodrock.

The use of standardized test scores as a means of student evaluation and school accountability has increased in the last two decades. In particular, the passage of the NCLB Act in 2001 required states to develop annual assessment instruments for students in grades three through eight. Even though standardized test scores are seen as imperfect assessment measures (e.g., Kane and Staiger 2002), they have been widely adopted to satisfy the requirement. The Florida Comprehensive Assessment Test (FCAT) is the state's explicitly-designed test to assess achievement of the Sunshine State Standards—a set of curricular standards approved by Florida's State Board of Education in 1996. Since 2001, the FCAT has been administered to students in grades three through ten. According to the Florida Department of Education, the two main purposes of the FCAT are to provide information to (i) parents about the level of mastery of the skills of their children and (ii) the public on the status of student education while allowing the public to hold schools and districts accountable for progress (Florida Department of Education 2009). Schools face rewards or sanctions depending on their performance, and there is evidence that the public pays attention to each individual school's performance on the FCAT (e.g., Figlio and Lucas 2004).

The data available to us on FCAT scores pertain to tests in reading comprehension (hereafter "Reading") and mathematics problem solving (hereafter "Mathematics") measuring individual student performance against national norms (Florida Department of Education 2009). The scores in these tests map into a 5-level scale, with a performance level of 3 or higher considered to be an "on grade" level. We consider this "on grade" level to be a "passing score" for the purpose of our analysis. Importantly, during the school years spanned by our data, there were no significant changes in the thresholds defining the 5-level FCAT scale.

As previously discussed, even though 4-H does not explicitly design its programs with the intention of increasing standardized test scores for its participants, there are specific reasons to expect a positive relationship between 4-H participation and test scores. Generally speaking, one can regard the 4-H program as an additional input into

the “education production function,” just as is the case with traditional school resources.¹⁴ Under this view, when existing alongside school instruction and other basic resources, 4-H promotes student learning that can be reflected in measures of academic performance such as standardized test scores. As such, 4-H can enhance the efforts of teachers and schools to improve student performance. Focusing on standardized test scores provides an objective measure of student performance, and also offers the advantage that they can directly be related to earnings later in life (e.g., Murnane, Willet and Levy 1995) and to economic growth rates (Hanushek and Kimko 2000), providing a way to relate estimated impacts of 4-H on test scores with expected economic benefits.

Data

In the absence of individual-level data containing both participation in 4-H and standardized test scores, we collected longitudinal data aggregated at the district-and-grade level for five school years in the state of Florida. These data contain the extent of participation in 4-H programs as well as student performance measures in the reading and mathematics subtests of the FCAT. Therefore, our main analysis relates the variation in district-grade-year cells in the extent of participation in 4-H to the variation at the same level in performance measures in the FCAT. We do this by controlling for numerous “fixed effects” at the coarser levels of districts, grades, years, and combinations of these three levels. In addition, we control explicitly for several variables such as demographic, economic, and school characteristics.

Florida's school-related data for the 2002/03 through 2006/07 school years come from the Florida School Indicator Report (FSIR). FSIR provides numerous indicators of school status and performance on public elementary, middle, and high schools for each of Florida's 67 school districts, which correspond to the same number of counties in the state. The student performance outcomes available are the average score within each

¹⁴ Coleman et al. (1966) is an early influential study that analyzes the link between school resources or other measures of school quality and student outcomes. In the economics literature, there is a sizeable literature analyzing this link as well (e.g., Hanushek 1986; Krueger 1999).

school district and grade on the mathematics and reading subtests of the FCAT. Additionally, we know the percentage of students in each school district and grade that fall within each of the five achievement levels (from the lowest level of 1 to the highest level of 5) on each of the subtests of the FCAT.¹⁵ Information on the percentage of students tested on each FCAT subtest is also available to account for any atypical district-grade-year on this regard. The available grades are from third up to tenth grade. These data also contain school characteristics, including per pupil expenditures; percentage of instructional staff; average years of experience of teachers and percentage of them with advanced degrees; and percentage of students who are disabled, gifted, absent over 21 days from school, and English language learners. Data on each the proportion of students of each race and ethnicity, which were not readily available in the original FSIR data, were provided to us by the Florida Department of Education upon request. Finally, information on county unemployment rates during the relevant years was collected from the Bureau of Labor Statistics, Local Area Unemployment Statistics data.¹⁶

To analyze the impact of 4-H participation, we obtained data consisting of attendance numbers in 4-H clubs, school enrichment programs, and other community programs within each school district and grade. These data are regularly compiled by Florida 4-H from official 4-H forms filed by each of Florida's 67 county extension offices, which administrate the 4-H program in their counties and are independent of the school districts. The format of these 4-H participation data that was made available to us is "without duplication;" that is, if a student participates in more than one program offered by 4-H within that county, he/she is only counted once.¹⁷ Implicitly, since these

¹⁵ The five levels of the test are exogenously determined by the Florida Department of Education. To the best of our knowledge, these levels remain the same throughout the five school years spanned by our data.

¹⁶ An important variable previously employed in studies analyzing standardized test scores is the percentage of students eligible for free lunches. Unfortunately, we only have very sparse information on this variable for a small subset of the grade levels and years we analyze. As a result, we do not include this variable, but indirectly control for it through the numerous fixed effects we employ in our preferred specification.

¹⁷ The data "with duplication" could be interpreted as a measure of intensity of participation in the 4-H program by those students who participate (e.g., the intensive margin). Unfortunately, the data "with duplication" are only available at the district-year level (i.e., not broken down by grade level).

data simply aggregate participation over the different types of 4-H programs, they effectively weight the different programs equally. To construct the extent of participation in 4-H, we combined data on grade level enrollment for each school district and year with the data on 4-H participation by grade level within the same districts and years. The quotient of 4-H participation values and student enrollment is our variable “extent of participation in 4-H.”

There may be some potential issues with these data. First, since the participation data are collected by the county that administers the program, it is possible that some of the students cross into another county to attend 4-H programs. In this case, their participation will be attributed to a different school district from the one in which they attend school. Unfortunately, we do not have information on the number of these occurrences; however, we believe this number to be relatively small in comparison to the number of students who attend 4-H in their own school districts. Two factors drive our belief. The first is the fact that a considerable proportion of 4-H participation (about 75 percent) is through school enrichment programs that take place within the participant’s school. The second is the average size of counties in Florida (800 sq. miles), which implies that out-of-county 4-H participants would have to travel non-negligible distances. A second potential concern relates to the population of home-schooled students if this were to vary systematically with the proportion of 4-H participants. However, this does not appear to be a big concern given that only about 1.9 percent of students attending school in Florida are home schooled during the time period in our data. A third potential concern is that we do not currently have access to grade retention statistics. Thus, we rely on the numerous fixed effects and the set of other control variables to account for this potential source of confounding. Finally, a valid question is how representative our results may be of the national 4-H program. On the one hand, individual states have some flexibility in the delivery method and topics covered by their state-level 4-H programs. On the other hand, the core philosophies, structures, mandates, and general program

Nevertheless, the available data “without duplication” are better suited for our current analysis of participation in the 4-H program (e.g., the extensive margin).

curricula of state level 4-H programs are overseen at the national level. Therefore, it seems reasonable to think that the results presented below for Florida are indicative of the effectiveness of 4-H in other states, and nationally.

Table 1 provides descriptive statistics of the data. The first two columns include statistics for the full sample, consisting of 2,680 district-grade-year cells, while the subsequent columns provide statistics for each of the five school years spanned in our data (536 district-grade cells on each). The variable listed in the first row is the extent of participation in 4-H. The average participation across all years is 0.28, with the highest average participation being 0.33 and lowest being 0.22, while the standard deviation ranges from 0.45 to 0.52 across years. It is important that we have enough variation in this variable over each level (district, grade, years) to be able to control for potential confounders and to successfully identify the effects of interest. Figure 1 displays the standard deviations in our 4-H participation variable across districts for each year and grade level, which shows that it has a considerable amount of unconditional variation.

The second and third rows in Table 1 present the average FCAT scores for the mathematics and reading subtests. Perhaps more revealing than the average score are the next several rows showing the average percentage of students that fell in each of the FCAT levels (1 through 5) for each subtest. On average, only about 6.5 percent of students score in the highest level in mathematics and about 5.5 percent in reading. After these levels, there are two rows indicating the average passing rate, which corresponds to scoring in level 3 or above. The passing rates are just above one-half, with mathematics having a passing rate of about 57 percent and reading of about 53 percent. Two other relevant statistics are the percentage of students tested in each subtest since a concern can be that some district-grade-years may have an unusually high number of students missing the test in such a way as to artificially improve the scores. These variables are consistently around 0.95 across years, with a small standard deviation. The remaining variables in Table 1 correspond to demographic, economic, and school characteristics

available to us, many of which are commonly employed throughout the education production function literature (e.g., Hanushek 1986; Krueger 1999).¹⁸

Methodology

We start by motivating our methodology from an individual-level equation relating performance in the FCAT to the participation in 4-H and other factors:

$$(1) \quad FCAT_{idgy} = \alpha + \beta \cdot 4H_{idgy} + \mathbf{X}_{idgy} \delta + u_{idgy}$$

where i indexes individuals, d school districts, g grades, and y school years. $FCAT$ represents the average mathematics or reading score in the FCAT, $4H$ is an indicator of individual participation in 4-H, \mathbf{X} is a vector of covariates that impact performance in the FCAT, u represents unobservable factors that impact performance in the FCAT, and (α, β, δ) are parameters. We are interested in estimating β in a way that can be interpreted as the causal effect of participation in 4-H on FCAT performance. The main problem in using (1) to uncover a causal effect of 4-H participation on FCAT performance is that there are likely a variety of unobserved factors related to both of these variables that, if left uncontrolled, will confound the causal relationship of interest.

If individual-level data were readily available, β could be estimated using any of the methods commonly employed in observational studies, such as matching, instrumental variables, or selection models.¹⁹ Of course, each one of these methods relies on identifying assumptions that may be scrutinized. An alternative approach that we pursue here, since it does not require individual-level data, is to use the different levels in equation (1)—that is, school districts, grades, and years—to control for as many unobserved factors as possible that can potentially confound the causal relationship of

¹⁸ For some readers, the variable “% Disabled” may seem high. The reason is that “disable” is broadly interpreted in the United States’ public school system following the Individuals with Disabilities Education Act (IDEA) of 1975. For instance, it not only includes physically disabled children, but also children with other types of disabilities such as learning, speech, hearing, and emotional.

¹⁹ For a discussion of them, see Heckman, Lalonde, and Smith (1999); Imbens and Woldridge (2009); or Blundell and Costa Dias (2009).

interest. To do this, we aggregate individuals up to the level of cells defined by school district-grade-year. This aggregation is necessary for us to bypass the problem of unavailability of individual-level data. As a by-product of aggregation, by considering the causal relationship at the district-grade-year level, we estimate the full effect of 4-H that includes any potential spillover effects from 4-H participants to non-participants occurring within district-grade-year cells.

More formally, the relationship in (1) is aggregated by taking averages over individuals:

$$(2) \quad AvgFCAT_{dgy} = \alpha + \beta Part4H_{dgy} + \bar{\mathbf{X}}_{dgy} \delta + \bar{u}_{dgy}$$

where $AvgFCAT$ is the average FCAT score in mathematics or reading in the school district d , grade level g , and school year y ; $Part4H$ is the extent of participation in 4-H, and $\bar{\mathbf{X}}$ is a vector of observed control variables (those listed in Table 1). The parameter of interest is β , α is the corresponding intercept term in the aggregated model, δ is a conformable vector of coefficients, and \bar{u} represents the aggregated unobserved factors at the level of district-grade-year cells. In the unlikely case that only observed variables available in $\bar{\mathbf{X}}$ confound the relationship of interest—and their relationship with FCAT performance is linear—then the ordinary least squares (OLS) estimate of β in (2) identifies the causal effect of interest.

Unfortunately, it is likely that \bar{u} contains unobserved confounders of the causal relationship of interest, in which case the OLS estimate of β in (2) will not identify a causal effect. To begin exploiting the structure of our data to control for potential unobserved confounders, we include fixed effects for school districts ($sdistrict$), grade level ($grade$), and school year ($syear$). Their inclusion in the aggregated model allows controlling for unobserved differences in FCAT performance in school districts (e.g., due to permanent differences in county resources that may be devoted to schools), grade levels (e.g., permanent grade-level differences in test performance), and school years (e.g., trends or time-idiosyncratic shocks affecting test scores in all of Florida). All of those unobserved factors may be simultaneously related to the extent of 4-H

participation, and thus confound the causal effect of interest. While this is a vast improvement over the model in (2), we can account for a greater amount of potential unobserved confounders with a richer specification that includes interactions among the previous three sets of fixed effects, giving rise to our DDD model:

$$(3) \quad \begin{aligned} AvgFCAT_{dgy} = & \alpha + \beta Part4H_{dgy} + \bar{X}_{dgy} \delta + sdistrict_d + grade_g + syear_y \\ & + (sdistrict_d \times grade_g) + (sdistrict_d \times syear_y) + (grade_g \times syear_y) + \bar{\epsilon}_{dgy}. \end{aligned}$$

In this specification, $(sdistrict_d \times grade_g)$ controls for systematic factors across grade levels *and* school districts that impact FCAT scores and do not change over the five years of the data, such as specific district policies that apply to particular grades (e.g., earmarked resources for specific grade levels or class size restrictions). If these unobserved factors are simultaneously related to 4-H participation and were left uncontrolled for, they would confound the causal effect of interest. These are the most extensive set of interactions that we include, consisting of 536 effects (67 districts times 8 grades). The school district-year effects $(sdistrict_d \times syear_y)$ are the second largest set of interactions (335 of them) that control for factors that change over time for different school districts but that are universal to all grade levels. Clear examples of these effects in Florida are hurricanes or tropical storms that cause disruptions in particular school districts affecting test performance. These interactions also account for changes over time in the population within the school districts such as the effects of new sources of employment (e.g., hospitals or factories) that may change the demographic composition of the population (e.g., more educated parents). Alternatively, they could account for changes over time in relative test preparation intensities across different school districts, perhaps influenced by past FCAT performance.²⁰ The last type of interaction consists of

²⁰ Variations in these intensity levels might also affect the extent of 4-H participation, turning these factors into confounders (if they were uncontrolled for). Consider a school district under a high degree of scrutiny due to poor past FCAT performance of its students. The school district might actively discourage participation in extracurricular activities such as 4-H if it is perceived to take away time from extra test preparation. In addition, increases in test preparation assignments might simply limit the time students have available to spend on other activities like 4-H.

40 ($grade_g \times year_y$) effects that control for factors that change across years and grades yet remain constant across school districts. Examples of these are statewide policies affecting particular grades or year-to-year differences in the level of difficulty of the FCAT by grade caused, for instance, by the introduction of new questions for each grade. Again, these unobserved factors might be confounders if they are related to 4-H participation and they were left uncontrolled for.

Given that the model in (3) controls for the largest possible number of potential confounding factors in our data, it is our preferred model specification. Within this aggregated model, we are able to identify the causal effect of the extent of participation in 4-H on measures of FCAT performance as long as any remaining unobserved factors impacting FCAT scores not captured by our variables and large set of fixed effects are uncorrelated with the extent of participation in 4-H. In other words, that $\bar{\varepsilon}_{dgy}$ in (3)—which represents unobserved factors that vary simultaneously at all three levels of school district, grades, and years—does not contain any confounders of the causal relationship of the extent of participation in 4-H and FCAT performance. This is our identification assumption. Importantly, for our model to fail to identify the causal effect, it is required that uncontrolled confounders vary *simultaneously* at the level of school districts, grades, and years, as variation at single levels or pairwise combinations of levels are accounted for by the large set of fixed effects included.

We refer to the model in (3) as a DDD model since the causal effect is identified by relating the triple differencing of the FCAT performance measure to the corresponding triple differencing of the extent of participation in 4-H (over the three levels of school districts, grades, and school years). Traditional DDD models (e.g., Katz 1996; Ravallion et al. 2005) employ this same strategy on a binary treatment or policy variable that results from the interaction of three indicator variables: one for the time-period indicator in which the policy takes place, and two other indicators, each of which relates to two different groups where only one of these groups is eligible for the policy. Here, the effect is identified by the interaction of each of the three indicator variables, which has the

interpretation of a triple differencing. Our DDD model in (3) can be seen as an extension of the traditional DDD model to the case of a continuous treatment variable (extent of 4-H participation), in which different groups (that define our cells) are exposed to different levels of the treatment variable. Similar DDD approaches to ours have been used in other settings involving schooling outcomes, such as Figlio (2006), Currie et al. (2009), Levine and Schanzenbach (2009), among others.

We close this section with a discussion of the sources of unconditional and conditional (on the available variables and fixed effects) variation in the extent of 4-H enrollment. The latter type of variability, at the level of district-grade-year cells, is the crucial one employed to identify β . Perhaps the main driver of the unconditional variability is the flow of funds to the (local) county 4-H offices.²¹ As mentioned earlier, the sources of funds are diverse. The federal and state funds that flow to the local offices are not based on the perceived academic performance of children in each county (let alone the performance on the FCAT), but rather on the number of children in the counties. Regarding the conditional variability employed to identify the causal effect of interest, one source is due to factors related to the provision of 4-H programs, such as unplanned changes in year-to-year 4-H staff in particular counties that affect the availability of 4-H programs for particular grades, and changes in the availability of 4-H volunteers at the district level that happen to impact particular grades.²² Another potential source is due to factors related to the provision of “competing” programs for children such as the Boys and Girls Club and the YMCA that may affect participation in 4-H. Changes in the provision of these competing programs can arise out of similar factors as those discussed above in the context of provision of 4-H programs. Importantly, we argue that these sources of conditional variability are likely uncorrelated to the residual variation in FCAT performance measures at the district-grade-year level. Indeed, a key factor aiding us in this regard is that the 4-H program (as well as most of the other

²¹ The main sources of information for the ensuing discussion are a set of conversations with Florida 4-H administrators at the University of Florida’s Institute of Food and Agricultural Sciences (UF/IFAS).

²² These factors can result, for instance, due to the regular labor force mobility of 4-H staff or volunteers in and out of the county. Those migration flows are likely unrelated to the FCAT scores.

competing programs) does not have an explicit goal of impacting standardized test scores.²³

Results and Implications

The estimated effects of the extent of 4-H participation on FCAT performance on the mathematics and reading subtests are reported in Table 2. For simplicity, only the coefficient of interest (β) is presented, while the set of estimated coefficients for the variables contained in \bar{X} for our preferred model are relegated to the supplementary appendix online. The column headings indicate the model being estimated. We measure FCAT performance for each of the two subtests in three ways, presented in the horizontal panels of the table: average scores, the percentage of students passing the test (level 3 or above), and the percentage of students that fall into each of the 5 levels of the test.

As motivated in previous sections, the expectation is that a higher extent of participation in 4-H is related to better FCAT performance. If this is the case, we expect a positive relationship with average scores and passing rates. Regarding the percentage of students in each of the five levels of the test, the expected sign is less clear-cut: we would expect a positive relationship with the percentage of students performing in the highest level of the FCAT, a negative one with the percentage in the lowest level, but no apriori expectation with respect to the intermediate performance levels due to students moving simultaneously up and into and up and out of these levels, as performance gains are made.

The first model, presented for comparison purposes, is a regression of the different FCAT performance measures on the extent of 4-H participation and a constant term. By failing to account for any potential confounders of the relationship, this model

²³ An interesting point raised by one of our Referees is whether there may be strategic or compensatory funding decisions between 4-H and the other “competing” groups. In conversations with Florida 4-H administrators, they appeared skeptical about this possibility given their own experience with 4-H and their perception that the local funding to these programs is typically in short supply, making it difficult to respond strategically in the short run.

estimates a simple association between the two variables. Interestingly, for most of the performance measures, the estimated effect is of the opposite expected sign (for those that the expected direction is unambiguous) and is often statistically significant. For example, the estimated effect is negative for the average FCAT scores in each of the mathematics and reading subtests. Leaving all potential confounders uncontrolled appears to result in a negative association between the extent of participation in 4-H and the measures of performance in the FCAT.

The second model shown in Table 2 is an OLS regression that includes the list of variables in $\bar{\mathbf{X}}$ that are available to us.²⁴ By controlling for these observed factors, many of the counterintuitive signs on the estimated effects in the previous model are reversed and their magnitudes are considerably decreased. For example, the effect of the extent of 4-H participation on the average score in the FCAT is now positive in both mathematics and reading, with the latter showing a statistically significant relationship. Similarly, the passing rate and percentage in levels 1 and 5 in reading have the expected signs and are statistically significant. On the contrary, the passing rate and the percentage in levels 1 and 5 in mathematics have opposite expected signs and are statistically significant. Since this model does not control for any *unobserved* confounding factors, it likely comes up short from uncovering a causal effect.

Model 3 in Table 2 introduces a first set of fixed effects for school districts (*sdistrict*), grade levels (*grade*), and school years (*syear*), along with the control variables. By not including the full set of fixed effects available to us, there are likely unobserved factors left unaccounted for that can confound the causal relationship of interest. Nevertheless, in this model most of the estimated effects have the expected sign. Among the statistically significant effects (at the 10 percent level) with the expected sign is the passing rate and the percentage of students in level 5 in the reading subtest of the FCAT.

²⁴ The models that control for observed variables ($\bar{\mathbf{X}}$) include those listed in Table 1 plus square terms of the proportion of students that fall in the different racial and ethnic categories, as well as interactions of these proportions with the county unemployment rates.

Model 4 in Table 2, which corresponds to equation (3), presents our preferred model's estimates of the relationship between 4-H participation and FCAT performance. This model controls for the most extensive set of factors possible with our data, including observed ones (\bar{X}) and unobserved ones through the full set of fixed effects. Therefore, as previously discussed, this model identifies the causal relationship of interest by employing variation at the district-grade-year level. All estimated coefficients in this model conform to their expected sign. Importantly, several estimated effects are statistically significant at the 10 percent level, such as both of the FCAT passing rates, the average score for reading, and the proportions of students in level 1 of mathematics and level 5 of reading.^{25,26}

We now discuss the interpretation of the magnitude of the results in the context of our preferred model specification. The estimated effect when the outcome is the average test score implies that a one percentage point increase in the extent of participation in 4-H results in a statistically significant increase of 1.02 points in the average reading scores of the FCAT. To put this effect in perspective, it is helpful to translate it into standard deviations: given the sample standard deviation in FCAT average reading scores of 14.15, the estimated effect represents an increase of 0.07 standard deviations. The magnitude of this effect seems to be in line with impacts found elsewhere in the literature for other educational inputs. For instance, it has been found that a reduction in class size of about 30 percent leads to an increase in 0.22 standard deviations in standardized test scores (e.g., Krueger 1999). Taking the mean extent of participation in 4-H of 28 percent (Table 1), a 30 percent increase in 4-H participation corresponds to an increase in 4-H enrollment of about 8 percentage points (or about 80,000 students). Thus, assuming linear

²⁵ Note that Model 4, which corresponds to equation (3), estimates a large number of parameters (1,017)—to isolate the variation at the district-grade-year level—with a sample size of 2,680; thus, it is not surprising that the precision of the estimates declines considerably relative to some of the previous models. Interestingly, Model 4 often yields more precise estimates relative to Model 3 (which omits the interactions among fixed effects).

²⁶ In looking at Figure 1, one potential concern is that the evidently lower variability on the extent of 4-H participation for grades 9 and 10 could result in different estimated effects. To assess this possibility, we estimated models using only grades 3 to 8 and adding an interaction of the extent of 4-H participation with an indicator variable for those two grades. In both models we find very similar results to those reported for Model 4, and the interactions in the second model are largely statistically insignificant.

effects, this 30 percent increase in 4-H participation is related to an increase of 0.56 standard deviations in FCAT average reading scores.

The coefficient of interest when the outcome under consideration is the percentage of students achieving a passing level in the corresponding subtests of the FCAT indicates that a one percentage point increase in the extent of 4-H participation results in a statistically significant increase of 1.1 percentage points in the passing rate of the mathematics subtest of the FCAT. The corresponding effect on the passing rate in the reading subtest is slightly lower at 0.8 percentage points, and it is statistically significant as well. These effects are also important, but not as large as the effects of other inputs. For instance, Papke (2005) reports that a 10 percent increase in school-level spending is related to increases in passing rates of about 3.7 percentage points in Michigan (for her econometric model that controls for the largest amount of potential confounders).

The estimated effects of the extent of 4-H participation on the proportion of students scoring on different levels of the FCAT indicate that a one percentage point increase in 4-H participation significantly decreases the proportion of students scoring in the bottom level of the mathematics subtest by 0.88 percentage points, while increasing the proportion scoring in the top level of the reading subtest by 0.37 percentage points. Another effect that is statistically significant is the proportion of students scoring in level 4 of the mathematics subtest, which is increased by 0.6 percentage points. The estimated effects on the remaining levels of the subtests are not statistically significant at the 10 percent level. In summary, the findings from our preferred model provide evidence that the extent of participation in 4-H has a positive and statistically significant effect on several measures of FCAT performance, suggesting that the programs delivered by 4-H in Florida are beneficial to its student participants from an academic standpoint. To end, we note that most of the estimated coefficients (in our preferred model) for the control variables employed (reported in the supplementary appendix online) have the expected signs and that a number of them are statistically significant.²⁷

²⁷ Many of these control variables are not statistically significant. This is likely due to the fact that our preferred model employs variation at the level of district-grade-year cells, and some of those variables may have a relatively small amount of variability at that level.

Heterogeneity in the Estimated Effects, Cumulative Effects, and Robustness Check

Table 3 reports two exercises to analyze the heterogeneity of the estimated effects with respect to differences in district-grade-year cells along two important dimensions: race/ethnicity and rural/urban locations. Column (1) of Table 3 reproduces the DDD results from Table 2 for reference (our preferred model's set of results). Columns (2) to (4) in Table 3 present estimated effects of the interaction of the extent of 4-H participation variable with the proportion of students in the district-grade-year cells that belong to the White, Black, or Hispanic race and ethnicities.²⁸ The main feature of these results is that most of the statistically significant effects occur for the interaction with the proportion of Blacks, and all are of the expected sign. Moreover, these estimated effects are of a large magnitude relative to our preferred model's set of results that average across race and ethnicities. For example, the estimated effect on the passing rates on the two FCAT subtests of the interaction term between the 4-H participation rate and the proportion of Black students in the district-grade-year is 6.62 and 5.06 for mathematics and reading, respectively. These estimated effects are significantly larger than the estimated overall effects of the 4-H participation rate of 1.1 and 0.8 for mathematics and reading, respectively. For the other two interactions (with the proportion of White and Hispanic students) the estimated effects are mostly statistically insignificant. Column (5) reports the p-value of the test for equality of effects across the three proportions shown (White, Black, and Hispanic). Only two of the tests are able to reject the equality of the three effects. Thus, while the estimates suggest that the 4-H program in Florida has a significant impact on standardized test scores predominantly in district-grade-year cells with a high proportion of Black students, the estimates are not precise enough to be conclusive about it.

²⁸ While other race and ethnicities are reported in our data (Asian, Native American, and multi-racial) and their interactions with the extent of 4-H are included in the estimated models, we focus here on the three larger groups since the others account for relatively small proportions. Note that the extent of 4-H participation does not enter by itself in this model since it is fully interacted with all race/ethnicity proportions available in the data.

Columns (6) and (7) of Table 3 present estimated effects of 4-H breaking down counties into predominantly “rural” or “urban.” We follow the categorization of counties given by the 2009 Florida Statutes (Section 288.0656).²⁹ According to this criterion, 48 percent of Florida counties are deemed rural. The reported estimated effects from this exercise are obtained by interacting the 4-H participation rate with an indicator of whether the county is rural or urban (recall that counties correspond to school districts in Florida). Looking at the results, perhaps surprisingly given the perception of 4-H as an agricultural program, all of the statistically significant effects of 4-H occur in counties that are urban. They are all of the expected sign, and they are larger than our preferred model’s results that average across type of county. Regarding the rural counties, while all their estimated effects are statistically insignificant, they are of the expected sign (with the only exception being the percentage of students in level 5 of the mathematics subtest). Column (8) reports a test of equality between the effects estimated by rural/urban status. While the tests indicate that we are not able to reject the equality of effects across these two types of counties for most of the estimated effects, two notable exceptions are the average FCAT score and the percentage of students in level 1 in the mathematics subtest.

Table 4 presents the estimated coefficients from two different models. The first model investigates the possibility that the effect of the extent of 4-H participation accumulates over time. To do this, we include one lag of the extent of 4-H participation variable along with the concurrent value. In particular, our lag is a year-and-grade lag, which ensures that we measure the lagged effect of the same grade level over time. That is, we look at, e.g., the (lagged) effect of participation in 4-H in second grade on test scores in third grade (for the same school district). As a result of using the year-and-grade lag, we lose one grade *and* one year of data, reducing our sample size to 1,876 cells (from 2,680). For this reason, it is difficult to include more lags given the short time dimension in our data.

²⁹ The main characteristics that classify a county as rural according to this Statute are (1) having a population of 75,000 or less, or (2) having a population of 125,000 or fewer and be contiguous to a county with a population of 75,000 or fewer.

The results support the notion that the effects of the extent of participation in 4-H accumulate over time, and are presented in column (1) of Table 4. The inclusion of a lagged 4-H participation variable increases the magnitude of the concurrent value of 4-H participation by between 17 and 50 percent (among those that are statistically significant). All estimated coefficients preserve their expected signs, and one estimate becomes statistically significant (average mathematics score), while two lose statistical significance (% Level 4 in mathematics and % Level 5 in reading). Furthermore, a handful of estimated lagged effects are statistically significant and tend to have larger magnitudes. In particular, the estimate for the mathematics passing rate is a significant 1.82, compared to the concurrent 4-H effect coefficient of 1.48 and the preferred model's estimate of 1.1. These estimated coefficients imply that the 2-year cumulative effect of 4-H participation on mathematics passing rate is 3.3 ($=1.48+1.82$). We interpret the results of this model as suggestive evidence that the effects of participating in 4-H accumulates over time, although we believe that more years of data would be necessary to convincingly estimate the cumulative effects of (aggregate) 4-H participation.

To end, we present one robustness check to our model, in which we add to our preferred specification one year-and-grade *lead* of the extent of participation in 4-H. This reduces again our sample size to 1,876. There is no reason to expect that the lead of 4-H participation be related to current FCAT scores, so finding significance on this variable in any of our models would suggest misspecification.³⁰ A similar robustness check is employed in Currie et al. (2009). Column (2) indicates that all of the coefficients on the lead of the extent of participation in 4-H are of a small magnitude (compare to our preferred model's results), and are all highly statistically insignificant (p-values ranging from 0.16 to 0.97). Importantly, the coefficients on the concurrent extent of participation in 4-H that are statistically significant are of a very similar magnitude relative to our preferred model's results. We conclude that this exercise does not yield evidence detrimental to the plausibility of our identifying assumption.

³⁰ One conceivable source of misspecification could be “feedback” from the idiosyncratic error term to future values of the extent of participation in 4-H (e.g., if negative shocks to test scores in a given period impact future values of the extent of participation in 4-H).

Conclusions

The 4-H program can be considered the largest youth development program in the world. Yet, economists have paid little attention to measuring its effects on youth (student) outcomes. We partially fill this void by examining the effect of 4-H participation on standardized test score outcomes utilizing grade-level longitudinal data on Florida's school districts from the Florida Department of Education combined with 4-H participation statistics from Florida 4-H. Employing a difference-in-difference-in-differences (DDD) approach to control for potential confounders of the causal relationship at the level of school districts, grade levels, and school years, our results indicate that the extent of 4-H participation at the district-grade-year level is positively related to various measures of performance in the mathematics and reading subtests of Florida's Comprehensive Assessment Test (FCAT).

More specifically, our preferred model specification indicates that the extent of 4-H participation has a positive and statistically significant effect on the average score on the reading subtest of the FCAT, as well as on passing rates (level 3 or above out of 5) on both the mathematics and reading subtests. The extent of participation in 4-H is also significantly related to a decrease in the proportion of students that score in the lowest level of the mathematics subtest, and to an increase in the proportion that score in the highest level of the reading subtest. Two exercises to explore the heterogeneity of the effects show that 4-H appears to be more effective in district-grade-year cells with a higher proportion of Black students and in those cells that are urban, although the precision of these estimates prevents us from being conclusive. We also find evidence indicating that the effects of 4-H participation accumulate over time since the estimated effects are higher when including one (year-and-grade) lag of the extent of participation in 4-H.

Evaluation of the causal impacts of youth development programs such as 4-H is important to various stakeholders, including policy makers, 4-H administrators,

volunteers, and the parents of participating students. Our results can be employed to approximate the economic gains brought about by 4-H. As an illustration, consider the present value of lifetime earnings for a fourth grader from an increase of one standard deviation in standardized test scores calculated by Kane and Staiger (2002) at between \$90,000 and \$210,000.³¹ Using a simple back-of-the-envelope calculation, these numbers imply that the present value of lifetime earnings from 4-H participation, according to our preferred model's results for the FCAT reading subtest, is between \$6,300 and \$14,700 on average. A second illustration relates test scores and economic growth: Hanushek and Kimko (2000) report that an increase of one standard deviation in their measure of standardized test scores is related to a boost in the growth rate of real GDP per capita of 1.4 percentage points per year.³² Taking their calculation at face value, our preferred model's results for FCAT reading average scores suggest that a one percentage point increase in 4-H participation could boost Florida's per capita real GDP by roughly 0.1 percentage points per year.

Several things must be kept in mind, however, when translating these impacts on test scores to economic impacts for any kind of cost-benefit analysis. First, the 4-H program has not been designed with the purpose of improving test scores. These improvements are therefore most likely by-products of positive youth development. For example, improvements in test scores could result from increased student interest in academically related topics, improved general motivation for school, or more productive interaction with instructors. Thus, improvements in FCAT outcomes might represent only a partial effect of the program's overall impact on participants. Second, the large reach and capabilities of 4-H can complement the efforts by teachers, schools, and communities to improve their schools in the face of the higher accountability brought about by the NCLB Act. This synergy can provide valuable positive externalities to schools and communities beyond the private benefits to participants in the program. Third, an

³¹ Kane and Staiger's calculations are based on recent economic studies presenting estimates of the relationship between test performance and earnings of young adults, such as those in Murnane et al. (1995).

³² Hanushek and Kimko (2000) utilize measures of cognitive skills of individuals from six international tests on mathematics and science conducted over a three-decade period.

important cost factor to consider is that 4-H makes use of a vast pool of volunteers in the delivery of its programs, whose value of donated time is generally not accounted for when reporting the actual cost of the program. All of these factors, among others, are important to keep in mind when considering the net benefits of the 4-H program.

References

- 4-H. 2007a. California 4-H Project Leader Digest. University of California Cooperative Extension. Available at: <http://www.ca4h.org/4hresource/forms/manuals/Project-Leader-Digest.pdf>. Accessed May 5, 2009.
- 4-H. 2007b. 4-H Youth Development Statistics.
- 4-H. 2009. The 4-H Story. Available at <http://4-h.org/4hstory.html>. Accessed May 5, 2009.
- Aizer, A. 2004. Home Alone: Supervision After School and Child Behavior. *Journal of Public Economics* 88: 1835-1848.
- Blundell, R. and M. Costa Dias. 2009. Alternative Approaches to Evaluation in Empirical Microeconomics. *Journal of Human Resources* 44 (3): 565-640.
- Carnegie Council on Adolescent Development. 1992. A Matter of Time: Risk and Opportunity in the Nonschool Hour. Wahington, DC.
- . 1989. Turning Points: Preparing American Youth for the 21st Century. Wahington, DC.
- Coleman, J.S., E.Q. Campbell, C.J. Hobson, J. McPartland, A.M. Mood, F.D. Weinfeld, and R.L. York. 1966. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office.

- Currie, J., E.A. Hanushek, E.M. Kahn, M. Neidell, and S.G. Rivkin. 2009. Does Pollution Increase School Absences? *The Review of Economics and Statistics* 91(4): 682-694.
- Farmer, S. 2010 (April 14). Proposed Budget Before Georgia Legislature Calls for Elimination of 4-H and Half of Extension. *Farm and Ranch News*.
- Figlio, D.N. and J. Winicki. 2005. Food for Thought: The Effects of School Accountability on School Nutrition. *Journal of Public Economics* 89: 381-394.
- Figlio, D.N. and M.E. Lucas. 2004. What's in a Grade? School Report Cards and the Housing Market. *American Economic Review* 94: 591-604.
- Figlio, D.N. 2006. Testing, Crime and Punishment. *Journal of Public Economics* 90: 837-851.
- Florida 4-H. 2008. 2007-2008 Florida 4-H Statistical Snapshot.
- Florida Department of Education. 2009. Florida Comprehensive Assessment Test . Available at: <http://fcats.fldoe.org/fcatspub3.asp>. Accessed May 5, 2009.
- Goodwin, J., J.B. Carroll and M. Oliver. 2005. Public School Students' Out of School Time Study. Measuring the Impact of Public School Students' Out of School Time Study. Research Report, Colorado State University.
- Grossman, M. and R. Kaestner. 1997. The Effects of Education on Health. In: *The Social Benefits of Education*, by Jehre Behrman and Nancy Stacy (Eds.), pp. 69-125. Ann Arbor: University of Michigan Press.
- Guion, L. 2002. Communicating the Outcomes and Impacts of 4-H. Florida Cooperative Extension Service EDIS PE059.
- Hanushek, E.A. 1986. The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature* 24: 1141-1177.

- Hanushek, E.A., D.D. Kimko. 2000. Schooling, labor force quality, and the growth of nations. *American Economic Review* 90: 1184-1208.
- Heckman, J.J., R. LaLonde, and J. Smith. 1999. The Economics and Econometrics of Active Labor Market Programs. In: Handbook of Labor Economics vol 3A, by O. Ashenfelter and D. Card, pp. 1865-2097. Amsterdam: Elsevier Science North Holland.
- Heckman, J.J., J. Stixrud, and S. Urzua. 2006. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics* 24(3): 411-482.
- Hollister, R. 2003. The Growth in After-School Programs and Their Impact. Paper Comissioned by the Brookings Roundtable on Children.
- Imbens, G.W. and J. Wooldridge. 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature* 47(1): 5-86.
- Jacob, B.A. and L. Lefgren. 2003. Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime. *American Economic Review* 93(5): 1560-1577.
- Kane, T.J. and D. Staiger. 2002. The Promise and Pitfalls of Using Imprecise School Accountability Measures. *Journal of Economic Perspectives* 16(4): 91-114.
- Kane, T.J. 2004. The Impact of After-School Programs: Interpreting the Results of Four Recent Evaluations. Working paper of the William T. Grant Foundation.
- Katz, L.F. 1996. Wage Subsidies for the Disadvantaged. NBER Working paper 5679.
- Krueger, A.B. 1999. Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics* 114: 497-532.
- Lerner, R., J. Lerner, E. Phelps, and Colleagues. 2009. Waves of the Future: The first five years of the 4-H study of positive youth development. Medford, MA: Institute for

- Applied Research in Youth Development, Tufts University and National 4-H Council.
- Levine, P.B. and D. Schanzenbach. 2009. The Impact of Children's Public Health Insurance Expansions on Educational Outcomes. *Forum for Health Economics & Policy* (12): 1-26.
- Murnane, R.J., J.B. Willet, and F. Levy. 1995. The Growing Importance of Cognitive Skills in Wage Determination. *Review of Economics and Statistics* 77(2): 251-266.
- Murray, K. 2010 (April 23). Hundreds Gather at the [Florida] Capitol to Lobby for Continued 4-H Funding. Tallahassee News.
- Norman, M. and J. Jordan. 2006. Discovering the Organizational Structure. Florida Cooperative Extension Service EDIS 4-HS FS101.3.
- Papke, L. 2005. The Effects of Spending on Test Pass Rates: Evidence from Michigan *Journal of Public Economics* 89: 821-839.
- Ravallion, M., E. Galasso, T. Lazo, and E. Philipp. 2005. What Can Ex-Participants Reveal about a Program's Impact? *Journal of Human Resources* XL(1): 208-230.
- Roth, J., J. Brooks-Gunn, L. Murray, and W. Foster. 1998. Promoting Healthy Adolescents: Synthesis of Youth Development Program Evaluations. *Journal of Research on Adolescence* 8(4): 423-459.
- Satcher, D. 2001. The Surgeon General's Call To Action To Prevent and Decrease Overweight and Obesity. Available at: <http://www.surgeongeneral.gov/topics/obesity/calltoaction/toc.htm>. Accessed December 13, 2009.
- Silliman, B. 2007. Critical Indicators of Youth Development Outcomes for 4-H National Mission Mandates. North Carolina State University.

Wentzel, K.R. 1993. Does Being Good Make the Grade? Social Behavior and Academic Competence in Middle School. *Journal of Educational Psychology* 85: 357-364.

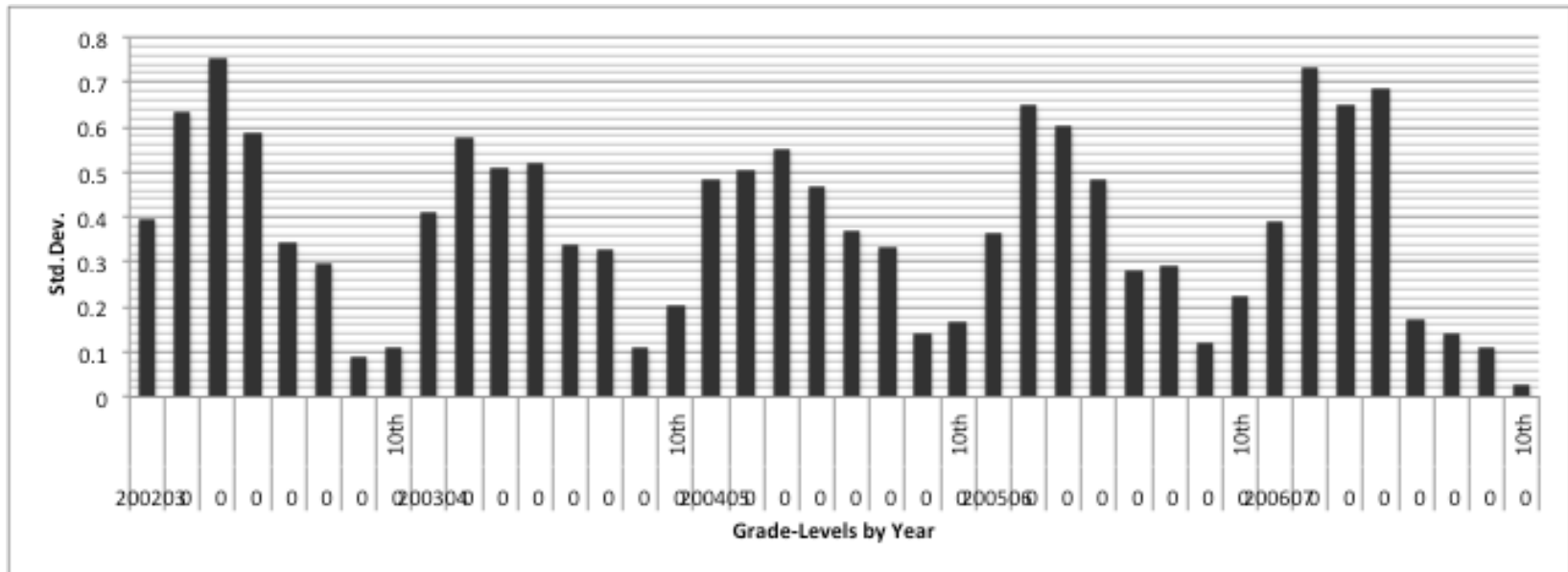


Figure 1: Standard deviation of the extent of 4-H participation over school districts

Table 1. Descriptive Statistics

School Years	Full Sample		2002 to 2003		2003 to 2004		2004 to 2005		2005 to 2006		2006 to 2007	
Variable	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Extent of Participation in 4-H	0.28	0.48	0.33	0.52	0.28	0.45	0.29	0.45	0.29	0.46	0.22	0.49
FCAT Math Mean	310.27	16.13	304.67	16.02	307.30	15.91	310.25	15.42	313.41	15.23	315.70	15.59
FCAT Reading Mean	302.59	14.15	298.07	13.66	300.10	14.77	301.95	14.16	305.65	13.24	307.16	12.84
FCAT Math % Level 1	20.50	9.16	23.64	9.23	22.34	9.72	20.44	8.82	18.80	8.29	17.28	8.21
FCAT Math % Level 2	22.26	5.04	23.00	4.37	22.54	4.49	22.29	5.05	21.94	5.38	21.54	5.67
FCAT Math % Level 3	30.98	5.73	29.83	6.15	30.00	5.96	31.35	5.30	31.43	5.34	32.31	5.49
FCAT Math % Level 4	19.69	7.23	17.82	6.67	18.63	6.86	19.53	7.04	20.99	7.48	21.47	7.44
FCAT Math % Level 5	6.57	3.90	5.73	3.47	6.43	3.76	6.42	3.86	6.88	3.91	7.38	4.27
FCAT Reading % Level 1	25.02	10.90	27.83	10.05	27.18	11.12	25.60	10.93	22.70	10.75	21.79	10.33
FCAT Reading % Level 2	21.64	7.41	21.91	6.63	21.22	6.69	21.67	7.67	21.81	7.99	21.62	7.96
FCAT Reading % Level 3	29.51	7.52	28.38	6.67	28.17	7.54	29.08	7.45	30.77	7.61	31.17	7.78
FCAT Reading % Level 4	18.33	9.18	17.04	7.82	17.64	8.66	18.10	9.41	19.53	10.12	19.34	9.50
FCAT Reading % Level 5	5.50	3.05	4.86	2.61	5.79	2.98	5.59	3.01	5.20	2.93	6.05	3.53
FCAT Math Passing %	57.24	12.25	53.38	11.48	55.06	11.99	57.30	11.91	59.30	11.99	61.16	12.30
FCAT Reading Passing %	53.34	16.38	50.28	14.26	51.61	16.30	52.77	16.93	55.50	17.03	56.56	16.40
FCAT Math % Tested	94.49	6.62	93.97	6.89	94.58	6.72	94.64	6.76	94.45	6.41	94.83	6.26
FCAT Reading % Tested	94.58	6.56	94.09	6.83	94.67	6.60	94.74	6.66	94.53	6.31	94.86	6.37
% Teachers w/adv degr.	30.77	7.26	31.31	7.58	31.34	7.16	31.00	6.94	30.33	7.22	29.87	7.32
% Instructional Staff	65.03	5.37	64.20	5.49	64.55	5.70	65.06	5.28	65.57	4.86	65.75	5.32
% English Learners	3.69	4.21	3.24	4.09	3.62	4.16	3.65	4.14	3.87	4.22	4.05	4.42
Unemployment Rate	4.17	1.01	5.08	0.88	4.55	0.89	3.75	0.71	3.36	0.66	4.10	0.89
% White	64.76	19.59	66.38	19.12	65.71	19.33	64.91	19.62	63.85	19.78	62.95	19.95
% Black	19.82	14.87	20.37	14.94	20.15	14.97	19.73	14.95	19.54	14.84	19.32	14.67
% Hispanic	11.95	13.15	10.54	12.40	11.12	12.68	11.91	13.14	12.70	13.46	13.48	13.84
% Asian	1.23	1.02	1.12	0.93	1.16	0.96	1.22	1.02	1.29	1.05	1.33	1.09
% Native American	0.33	0.42	0.30	0.36	0.30	0.38	0.33	0.41	0.36	0.44	0.36	0.48
% Multi-racial	1.91	1.46	1.29	1.11	1.56	1.24	1.89	1.40	2.24	1.52	2.56	1.60
% Gifted	3.83	2.74	3.82	2.68	3.86	2.70	3.80	2.61	3.76	2.81	3.91	2.89
% Disabled	16.62	3.73	17.07	3.70	17.00	3.79	16.69	3.75	16.31	3.79	16.01	3.52
Per Pupil Expenditures	5257.59	827.66	4517.42	363.62	4831.43	454.63	5173.08	548.97	5586.46	678.12	6179.54	795.08
Teacher Avg Yrs Exper.	13.13	1.99	13.81	2.02	13.48	2.06	12.91	1.78	12.87	2.29	12.60	1.47
% Absent over 21 days	11.40	5.73	11.33	5.44	11.19	5.61	11.76	6.07	11.77	5.88	10.96	5.62
Grade-District Student Enrollment	3087.60	5142.73	3043.77	5198.67	3093.61	5252.07	3102.42	5175.66	3113.76	5113.84	3084.43	4988.39
Number of Observations	2680		536		536		536		536		536	

Table 2: Estimated Effects of the Extent of 4-H Participation on Average Performance on the Mathematics and Reading Subtests of the FCAT

	Model 1: OLS		Model 2: OLS with controls		Model 3: Grade, district, and year fixed effects with controls		Model 4: DDD with controls (preferred model)		
	β	p-value	β	p-value	β	p-value	β	p-value	
Average FCAT score									
Mathematics	-2.528 ***	0.000	0.224	0.635	0.473	0.468	1.021	0.159	
Reading	-0.757	0.188	1.334 ***	0.003	0.377	0.289	1.020 *	0.061	
Percentage Passing FCAT (Level 3 or above)									
Mathematics	-4.853 ***	0.000	-2.763 ***	0.000	0.242	0.624	1.102 *	0.060	
Reading	7.716 ***	0.000	4.797 ***	0.000	0.497 *	0.071	0.803 *	0.065	
FCAT Mathematics Subtest									
%Level 5	-2.154 ***	0.000	-0.892 ***	0.000	0.273	0.158	0.025	0.905	
%Level 4	-2.783 ***	0.000	-1.162 ***	0.000	-0.067	0.801	0.596 *	0.098	
%Level 3	0.085	0.717	-0.710 **	0.023	0.036	0.902	0.481	0.293	
%Level 2	2.501 ***	0.000	1.870 ***	0.000	0.097	0.675	-0.287	0.424	
%Level 1	2.279 ***	0.000	0.846 **	0.026	-0.362	0.395	-0.879 *	0.054	
FCAT Reading Subtest									
%Level 5	-0.587 ***	0.000	0.299 ***	0.007	0.222 *	0.056	0.374 *	0.086	
%Level 4	4.188 ***	0.000	2.487 ***	0.000	-0.062	0.812	0.411	0.130	
%Level 3	4.115 ***	0.000	2.012 ***	0.000	0.338	0.201	0.018	0.959	
%Level 2	-4.233 ***	0.000	-2.505 ***	0.000	-0.102	0.633	-0.149	0.610	
%Level 1	-3.501 ***	0.000	-2.327 ***	0.000	-0.421	0.138	-0.612	0.126	

Note: Sample size in all models is 2,680 cells. All estimated standard errors are clustered at the district-grade-year level. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.

Table 3. DDD Estimation Employing Lagged 4-H Participation Rate and Interacting the 4-H Participation Rate with Different Variables of Interest

	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)
	DDD, Model 4 (preferred model)		Proportion White		Proportion Black		Proportion Hispanic		Test for equality of effects in columns (2), (3) and (4)		Rural County		Urban County		Test for equality of effects in columns (5) and (6)
	β	p-value	β	p-value	β	p-value	β	p-value	p-value	β	p-value	β	p-value	β	p-value
Average FCAT score															
Mathematics	1.02	0.16	-1.37	0.39	6.89 *	0.06	0.35	0.96	0.21	0.34	0.69	3.04 ***	0.01	0.06 *	
Reading	1.02 *	0.06	-0.25	0.85	4.98 *	0.09	-1.11	0.80	0.36	0.83	0.20	1.58	0.11	0.53	
Percentage Passing FCAT															
Mathematics	1.10 *	0.06	-1.24	0.32	6.62 **	0.04	4.56	0.45	0.10 *	0.78	0.25	2.06 *	0.08	0.34	
Reading	0.80 *	0.07	-0.44	0.67	5.06 **	0.02	0.57	0.84	0.13	0.56	0.29	1.52 **	0.04	0.29	
FCAT Mathematics Subtest															
%Level 5	0.02	0.91	-0.26	0.59	-0.15	0.87	-1.24	0.35	0.78	-0.06	0.81	0.28	0.44	0.44	
%Level 4	0.60 *	0.10	-0.42	0.60	1.55	0.39	2.56	0.44	0.56	0.41	0.33	1.14 *	0.10	0.36	
%Level 3	0.48	0.29	-0.55	0.55	5.21 *	0.09	3.24	0.44	0.19	0.43	0.34	0.64	0.59	0.87	
%Level 2	-0.29	0.42	0.75	0.37	-2.11	0.34	-4.93 *	0.10	0.05 **	-0.47	0.23	0.27	0.73	0.39	
%Level 1	-0.88 *	0.05	0.47	0.65	-4.75 *	0.04	0.27	0.95	0.21	-0.45	0.38	-2.14 **	0.01	0.08 *	
FCAT Reading Subtest															
%Level 5	0.37 *	0.09	0.78 *	0.10	-0.89	0.28	-0.02	0.99	0.29	0.42	0.12	0.24	0.44	0.65	
%Level 4	0.41	0.13	-1.11	0.11	2.01	0.11	-0.63	0.74	0.18	0.29	0.37	0.76 *	0.10	0.41	
%Level 3	0.02	0.96	-0.12	0.88	3.94 **	0.03	1.22	0.63	0.21	-0.15	0.71	0.52	0.46	0.41	
%Level 2	-0.15	0.61	0.39	0.60	-0.60	0.72	-1.02	0.66	0.79	-0.10	0.77	-0.29	0.58	0.77	
%Level 1	-0.61	0.13	0.07	0.94	-4.42 *	0.06	1.32	0.74	0.28	-0.41	0.37	-1.20	0.12	0.38	

Notes: Sample size in all models is 2,680 cells. The effects for each group in columns (2)-(4) and (6)-(7) represent the interaction between the 4-H participation variable and the variable defining the subgroup. In both models the 4-H participation variable is fully interacted with the categories. The race and ethnicity variables represent the proportion of students of that race/ethnicity that attend a given district-grade-year cell. The rural and urban variables refer to the Florida counties, as defined by The 2009 Florida Statutes. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.

Table 4. DDD Models Employing Concurrent, Lag and Lead of 4-H Participation Rate

	(1)				(2)			
	Model with Concurrent and one Lag of the extent of 4-H participation				Model with Concurrent and one Lead of the extent of 4-H participation			
	Concur.	p-value	Lag	p-value	Concur.	p-value	Lead	p-value
Average FCAT score								
Mathematics	1.53 *	0.07	1.04	0.29	0.64	0.53	0.73	0.41
Reading	1.22 *	0.08	1.05	0.21	1.25 *	0.10	-0.05	0.95
Percentage Passing FCAT								
Mathematics	1.48 **	0.04	1.82 **	0.05	1.00	0.20	0.34	0.59
Reading	0.98 *	0.10	1.09	0.11	0.95 *	0.10	0.02	0.97
FCAT Mathematics Subtest								
%Level 5	0.28	0.16	0.10	0.65	0.20	0.54	0.41	0.25
%Level 4	0.70	0.12	1.35 **	0.01	0.69	0.13	0.45	0.25
%Level 3	0.49	0.38	0.38	0.52	0.11	0.85	-0.52	0.27
%Level 2	-0.31	0.46	-1.35 ***	0.01	-0.78	0.11	-0.62	0.16
%Level 1	-1.16 **	0.04	-0.49	0.45	-0.37	0.60	0.34	0.56
FCAT Reading Subtest								
%Level 5	0.45	0.16	0.34 *	0.10	0.51 **	0.02	0.02	0.93
%Level 4	0.54	0.17	-0.01	0.98	0.51	0.18	0.27	0.51
%Level 3	-0.01	0.97	0.77 *	0.09	-0.07	0.89	-0.26	0.55
%Level 2	-0.27	0.50	-0.63	0.16	-0.46	0.25	-0.39	0.26
%Level 1	0.68	0.13	-0.51	0.40	-0.45	0.43	0.24	0.60

Notes: The sample size for the two models is 1,876 due to the use of one lag (or lead), equivalent to losing one time period and one grade level in the data. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.

Appendix Table A1: Complete DDD Results (Model 4) FCAT Mathematics Subtest

Variable	FCAT Mean		% Passing		% Level 5		% Level 4		% Level 3		% Level 2		% Level 1	
	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
Extent of Participation in 4-H	1.02	0.16	1.10 *	0.06	0.02	0.91	0.60 *	0.10	0.48	0.29	-0.29	0.42	-0.88 *	0.05
FCAT Math % Tested	-14.57 ***	0.01	-9.30 **	0.05	-2.63 **	0.05	-4.75	0.14	-1.91	0.51	-0.19	0.95	9.44 **	0.01
% English Learners	-0.44 *	0.06	-0.40 **	0.04	-0.11 *	0.07	-0.26 *	0.08	-0.03	0.82	0.32 **	0.02	0.06	0.67
% Gifted	-0.05	0.78	0.09	0.59	0.13 **	0.02	0.05	0.70	-0.08	0.34	-0.11	0.32	0.02	0.85
Teacher Avg Yrs Exper.	-0.15	0.42	-0.15	0.31	-0.07	0.12	-0.15	0.12	0.07	0.61	0.13	0.17	0.01	0.95
Per Pupil Expenditures (x100)	0.12 *	0.10	0.10	0.10	0.04 **	0.03	0.07 *	0.09	-0.02	0.71	-0.06	0.11	-0.04	0.30
% Disabled	-0.57 ***	0.00	-0.40 ***	0.00	-0.08 *	0.06	-0.27 ***	0.00	-0.05	0.59	0.12	0.18	0.28 **	0.01
% Absent Over 21 days	-0.06	0.43	-0.08	0.21	0.00	0.87	0.00	0.94	-0.08 *	0.06	0.03	0.50	0.05	0.29
% Teachers w/adv degr.	0.14	0.36	0.10	0.41	0.09 **	0.04	0.10	0.26	-0.09	0.25	-0.14 *	0.08	0.05	0.58
% Instructional Staff	0.21 **	0.04	0.07	0.37	0.03	0.27	0.07	0.27	-0.02	0.73	0.12 **	0.01	-0.19 ***	0.00
% Black	0.20	0.61	-0.20	0.53	-0.02	0.88	-0.13	0.60	-0.06	0.78	0.25	0.20	-0.08	0.74
% Hispanic	0.08	0.85	0.11	0.76	0.02	0.89	-0.37	0.19	0.46 *	0.08	0.15	0.54	-0.28	0.33
% Asian	0.49	0.77	0.91	0.55	-0.54	0.37	-0.14	0.89	1.59	0.15	-1.06	0.31	0.00	0.99
% Native American	-0.67	0.78	-0.02	0.99	-0.17	0.80	1.12	0.35	-0.97	0.43	-0.03	0.98	0.02	0.99
% Multi-racial	-1.40	0.25	-1.65 *	0.10	-0.24	0.44	-0.26	0.71	-1.15 *	0.06	0.90	0.18	0.61	0.41
% Black Squared (x100)	-0.04	0.85	0.03	0.88	0.07	0.11	0.20 **	0.04	-0.25 **	0.03	-0.19 *	0.06	0.18	0.22
% Hispanic Squared (x100)	-0.01	0.98	-0.03	0.91	-0.12	0.26	0.30	0.18	-0.21	0.36	0.13	0.55	-0.07	0.74
% Asian Squared	0.07	0.66	-0.18	0.20	0.07	0.26	-0.04	0.70	-0.21 *	0.06	0.14	0.13	0.01	0.96
% Native American Squared	-0.14	0.76	0.13	0.72	-0.16 *	0.05	0.18	0.47	0.11	0.52	-0.17	0.52	0.00	0.98
% Multi-racial Squared	0.05	0.58	0.10	0.18	0.03	0.31	-0.02	0.67	0.09 *	0.06	-0.07	0.20	-0.04	0.44
% Black*Unemp Rate	-0.17 **	0.04	-0.07	0.35	-0.03	0.22	-0.05	0.33	0.01	0.83	-0.01	0.77	0.08	0.13
% Hispanic*Unemp Rate	-0.10	0.27	-0.09	0.24	0.00	0.85	-0.01	0.86	-0.08	0.25	-0.02	0.70	0.11 *	0.06
% Asian*Unemp Rate	-0.06	0.86	-0.07	0.82	0.07	0.56	0.11	0.58	-0.25	0.25	0.15	0.47	-0.03	0.91
% Native Amer*Unemp Rate	0.06	0.91	-0.27	0.53	0.08	0.58	-0.41	0.13	0.06	0.85	0.15	0.67	0.13	0.67
% Multi-racial*Unemp Rate	0.28	0.22	0.30	0.11	0.04	0.53	0.08	0.54	0.18	0.15	-0.15	0.25	-0.11	0.44
Constant	346.46 ***	0.00	87.71 ***	0.00	14.49 ***	0.00	33.75 ***	0.00	39.82 ***	0.00	13.57 *	0.07	-0.04	0.99

Note: All estimated standard errors are clustered at the district-grade-year level. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.

Appendix Table A2: Complete DDD Results (Model 4) FCAT Reading Subtest

Variable	FCAT Mean		% Passing		% Level 5		% Level 4		% Level 3		% Level 2		% Level 1	
	Coef	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
Extent of Participation in 4-H	1.02 *	0.06	0.80 *	0.07	0.37 *	0.09	0.41	0.13	0.02	0.96	-0.15	0.61	-0.61	0.13
FCAT Reading % Tested	-17.56 ***	0.00	-8.88 ***	0.01	-3.01 *	0.07	-1.65	0.46	-4.22 *	0.09	-1.03	0.64	9.69 ***	0.01
% English Learners	-0.11	0.58	-0.24	0.12	0.01	0.92	0.03	0.81	-0.27 **	0.03	0.11	0.36	0.14	0.33
% Gifted	0.21	0.27	0.25 *	0.07	0.11 **	0.05	0.04	0.62	0.10	0.25	-0.07	0.43	-0.19	0.18
Teacher Avg Yrs Exper.	0.37 ***	0.01	0.18	0.21	0.10 **	0.03	0.08	0.25	-0.01	0.94	0.08	0.49	-0.22 ***	0.01
Per Pupil Expenditures (x100)	-0.04	0.48	-0.03	0.52	-0.02	0.28	-0.01	0.84	0.00	0.93	0.02	0.56	0.00	0.93
% Disabled	-0.63 ***	0.00	-0.42 ***	0.00	-0.08 **	0.04	-0.15 *	0.06	-0.19 **	0.02	0.07	0.36	0.37 ***	0.00
% Absent Over 21 days	-0.05	0.40	-0.04	0.32	0.00	0.79	-0.02	0.38	-0.02	0.62	-0.04	0.18	0.07	0.13
% Teachers w/adv degr.	0.08	0.53	0.01	0.94	0.01	0.77	0.06	0.34	-0.06	0.43	0.00	0.95	-0.01	0.90
% Instructional Staff	0.17 *	0.06	0.08	0.22	0.04	0.20	0.01	0.90	0.04	0.50	0.04	0.42	-0.14 **	0.03
% Black	0.00	0.99	-0.11	0.70	-0.14	0.23	-0.07	0.64	0.11	0.60	0.12	0.51	-0.02	0.92
% Hispanic	0.24	0.52	-0.02	0.94	-0.14	0.13	-0.07	0.69	0.18	0.41	0.40 **	0.03	-0.41	0.16
% Asian	0.83	0.55	0.12	0.91	0.00	0.99	0.42	0.56	-0.30	0.73	0.31	0.71	-0.33	0.76
% Native American	3.33 *	0.08	2.23	0.13	0.54	0.36	2.45 **	0.02	-0.77	0.51	-1.90 *	0.09	-0.02	0.99
% Multi-racial	-0.25	0.76	-0.28	0.67	0.52	0.10	0.40	0.39	-1.19 **	0.01	-1.42 ***	0.01	1.70 ***	0.01
% Black Squared (x100)	0.12	0.46	-0.08	0.46	0.03	0.52	0.19 **	0.01	-0.30 ***	0.00	0.06	0.59	0.02	0.86
% Hispanic Squared (x100)	0.33	0.34	0.43	0.08	0.03	0.73	-0.20	0.22	0.60	0.01	-0.08	0.63	-0.38	0.12
% Asian Squared	-0.12	0.44	-0.08	0.48	0.02	0.70	-0.11	0.20	0.01	0.93	0.00	0.99	0.08	0.45
% Native American Squared	-0.02	0.94	0.11	0.55	0.04	0.52	-0.03	0.83	0.10	0.53	-0.07	0.47	-0.01	0.95
% Multi-racial Squared	-0.02	0.72	-0.07	0.19	-0.03	0.34	-0.04	0.50	0.00	0.96	0.13 ***	0.00	-0.04	0.39
% Black*Unemp Rate	-0.16 **	0.03	-0.06	0.29	0.01	0.67	-0.06 *	0.09	-0.01	0.83	-0.03	0.49	0.10	0.11
% Hispanic*Unemp Rate	-0.14 *	0.07	-0.08	0.17	0.02	0.29	-0.01	0.82	-0.09 **	0.05	-0.09 *	0.06	0.17 ***	0.00
% Asian*Unemp Rate	-0.14	0.64	-0.03	0.89	-0.02	0.78	-0.05	0.77	0.04	0.83	-0.02	0.92	0.03	0.89
% Native Amer*Unemp Rate	-0.84 **	0.05	-0.51	0.13	-0.18	0.16	-0.52 **	0.03	0.19	0.49	0.34	0.15	0.09	0.77
% Multi-racial*Unemp Rate	0.15	0.37	0.13	0.34	-0.06	0.27	-0.05	0.62	0.24 **	0.01	0.23 *	0.05	-0.39 ***	0.01
Constant	319.51 ***	0.00	62.35 ***	0.00	6.23 *	0.10	18.86 ***	0.00	37.53 ***	0.00	26.64 ***	0.00	12.38	0.16

Note: All estimated standard errors are at the district-grade-year level. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.