

ORIGINAL ARTICLE

Open Access



# The time-varying role of the family in student time use and achievement

Marie C. Hull<sup>1,2</sup>

Correspondence:

mchull2@uncg.edu

<sup>1</sup>Department of Economics,  
University of North Carolina–  
Greensboro, PO Box 26170,  
Greensboro, NC 27402-6170, USA  
<sup>2</sup>IZA, Bonn, Germany

## Abstract

In this paper, I use a unique dataset linking administrative school data with birth records to quantify the importance of time-varying family factors for child achievement and time use. Specifically, I take a model of academic achievement commonly used in the test score literature, and I augment it to include a family-year effect. Identification comes from the large number of sibling pairs observed in the same year. While prior literature has focused on specific shocks, such as job loss, I capture the full set of innovations that are shared across siblings in a given year. The distributions of fixed effects reveal that annual family innovations, relative to what was expected based on the previous year, are more important than teacher assignment for student achievement and also play a substantial role in the time students spend on homework, free reading, and television.

**JEL Classification:** I21, J13, J24

**Keywords:** Human capital, Educational achievement, Time use, Family inputs

## 1 Introduction

Across disciplines, there is a long literature on the relationship between family background and educational outcomes. Despite the widely held value of equality of opportunity, past research has documented that the family into which a child is born has a large impact on the course of his life. However, the family environment should also be understood as a dynamic environment: Changes in family life can also have a substantial influence on educational outcomes. While many analyses focus on permanent family characteristics, there is also a literature that examines the impact of various transitory ones. In this paper, I formulate a new method to quantify the time-varying role of the family using a within-sibling-year measure, and then, I use this method to estimate its importance for student time use and achievement.

Studies of permanent family characteristics have examined both the share of the variation in an outcome that can be attributed to the family as well as the impact of specific family characteristics on outcomes. In the literature on sibling correlations, researchers often estimate a linear model with a family effect and then analyze estimates of the variance components. Results indicate that at least half of the variation in educational attainment and adult IQ is due to factors shared by siblings (Björklund and Salvanes 2011; Björklund and Jäntti 2012). Although this fraction is sizable, these shared factors include the family environment in addition to shared genes, neighborhoods, and schools. Importantly for this study, sibling correlations for adult outcomes fail to isolate the role

of time-dependent family factors. Other strands of the literature on permanent family characteristics have focused on the effect of specific characteristics, like parent's educational attainment or race, and find that many of these play a substantial role in children's outcomes.<sup>1</sup>

What role is there for the time-varying family environment? While many family-level traits are fixed at birth, others change as a child ages. A number of studies have used fixed effects strategies or exogenous variation to analyze the causal impact of specific factors. For example, Stevens and Schaller (2011) employ child fixed effects to determine that parental job loss increases the probability of grade retention. Ananat et al. (2011) use plant closings as an exogenous source of job loss, and Lyle (2006) uses military deployments and job assignments during the Gulf War to study the impact of parental absences and household relocations. Milligan and Stabile (2011) and Dahl and Lochner (2012) exploit changes to government benefits to estimate that a \$1000 increase in income increases test scores by about 0.06 standard deviations for children from low-income families. Another related paper is Carneiro and Ginja (2016), which studies how permanent and transitory income shocks affect parental investments in children. Many more transitory family factors could affect student learning, but we may not observe them or even be able to measure them. For example, a negative health shock to a sibling or grandparent could affect student learning and time use, but these are often not observed.<sup>2</sup> An increase in parental oversight could impact student time use and achievement, but the various dimensions of parental oversight are difficult, if not impossible, to measure. Studies about job loss or income shocks help us understand the relationship between specific family-level changes and student achievement, but they do not provide a higher-level understanding of the time-varying role of the family. This paper takes a step toward the bigger picture by treating these transitory traits as a bundle and quantifying their importance.

In this paper, I begin with a model of student outcomes that is commonly found in the test score literature, though I apply the model to both test scores and time use outcomes. I then decompose the error from the model into a family-year-specific component and the rest of the error. I call this family-year fixed effect the "family innovation," but this is merely shorthand for a whole bundle of time-varying family factors, which could include planned inputs or shocks that are not captured by the other controls in the regression. A unique data set from North Carolina that merges administrative school data with birth records makes this analysis possible.<sup>3</sup> I identify the family innovations from sibling pairs observed in the same year. To quantify the time-varying role of the family, I calculate the standard deviation of these family innovation parameters for two test scores and four measures of student time use.<sup>4</sup> I also analyze the correlations between family innovations to different student outcomes, both for the whole sample and by socioeconomic status. These correlations tell us whether time-varying family circumstances that affect one outcome also tend to affect others.

I find that the time-varying family environment plays an important role in student time use and test scores. A one standard deviation move in the distribution of family innovation parameters is associated with a 0.17 standard deviation change in math score and a 0.19 standard deviation change in reading score. These results indicate that the time-dependent family environment is more important than teacher assignment for student achievement. For example, recent estimates from Chetty et al. (2014) show that a one standard deviation change in teacher quality moves math scores by 0.14 standard deviations

and English scores by 0.1 standard deviations.<sup>5</sup> The time use results imply that a one standard deviation increase in the distribution of family innovations increases homework time by 18 h, free reading time by 28 h, and TV watching by 102 h over the course of the school year. These changes are 20–22% of the average time use for each activity. The results vary by the characteristics of the sibling pair, and they are robust to models that allow different students to be on different learning trajectories. When I analyze the correlations between family innovations for different outcomes, I find that family innovations to achievement are most strongly (and positively) related to family innovations to educational uses of time (i.e., homework and free reading). However, these correlations vary by the socioeconomic status of the family. The results suggest that programs targeted to increase education-related time inputs at home could produce gains to student achievement, especially among low-SES students.

The analysis in this paper is descriptive in nature: I do not attempt to ascribe a causal interpretation to the family innovation parameter.<sup>6</sup> My goal is simply to apply a fixed effects strategy to evaluate the relative importance of time-dependent family circumstances for student outcomes. The family innovation parameter captures a bundle of factors that make a set of sibling outcomes deviate from their expected growth. In this sense, it only captures aspects of the bundle that lead sibling outcomes to move together. Correlated shocks across siblings could result from parental behavior, from social interaction effects, from local neighborhood shocks, or just from unobserved characteristics that are shared by siblings and affect changes in test scores and other outcomes. Conceiving of the family innovation as a bundle of factors is consistent with the literature on sibling correlations.<sup>7</sup> Whereas past studies have focused on specific inputs and events affecting the family, the results presented here suggest a larger role for the time-varying family environment than perhaps previously thought.

In the next section, I detail my method for estimating the family innovation parameters. I describe the matched education and birth data sets in Section 3. Section 4 contains the main results, and Section 5 discusses heterogeneity in the results by child and family characteristics. Section 6 concludes.

## 2 Method

I begin by taking a value-added version of the education production function:

$$y_{ifst}^k = \gamma^k y_{it-1}^k + x_{if} \beta^k + \delta_{st}^k + \varepsilon_{ifst}^k, \quad (1)$$

where  $y_{ifst}^k$  is the outcome  $k$  for student  $i$  from family  $f$  attending the school-grade pair  $s$  in year  $t$ . I include the superscript  $k$  to index outcome since I apply the model to six different outcomes (two test scores and four uses of time). This year's outcome  $k$  is a function of last year's outcome  $y_{it-1}^k$ ; student observables  $x_{if}$ , which include permanent family characteristics; a school-grade-year fixed effect  $\delta_{st}^k$ ; and an error term  $\varepsilon_{ifst}^k$ .

The goal of this paper is quantify the time-varying role of the family for student time use and achievement. To this end, I decompose the error term from Eq. (1) into a family-year specific component and the rest of the error:

$$\varepsilon_{ifst}^k = \xi_{ft}^k + v_{ifst}^k, \quad (2)$$

where  $\xi_{ft}^k$  and  $v_{ifst}^k$  are independent. The parameter of interest is then the family-year fixed effect  $\xi_{ft}^k$  since it captures the effect of time-varying circumstances for each family-year

combination. Throughout the paper, I refer to the family-year effect as the “family innovation.” To understand the role of family innovations, I analyze (i) the dispersion in  $\xi_{ft}^k$  for each outcome and (ii) the correlations between the family innovation parameters for different outcomes. Below, I define a model primitive to provide a framework to discuss assumptions and the interpretation of the family innovation; I then return the focus to the estimated equation.

For interpretation, it is helpful to think about underlying changes in the family environment separately from the impact they have on a specific child within the family. Define  $\theta_{ft}$  as the underlying family innovation. This could be a change in family structure or composition (via marriage, divorce, birth, or death), a change in economic conditions (such as entry into the labor market, job loss, inheritance), a change in parental supervision or parenting technique, or some other time-dependent factor that a set of siblings experiences. Although each sibling is exposed to the same underlying family innovation, the impact of  $\theta_{ft}$  may not be the same for each sibling and likely differs for each outcome. Denote  $\alpha_{ift}^k$  as the impact of an underlying family innovation on a particular child. For a given child  $i$ , the family-year effect can be written as  $\alpha_{ift}^k \theta_{ft}$ . In this framework, the impact of the family innovation  $\alpha_{ift}^k$  may vary by child characteristics, like gender or age, or permanent family characteristics, like parent’s education.

While I do not estimate  $\theta_{ft}$  and  $\alpha_{ift}^k$  separately, understanding  $\xi_{ft}^k$  as the product of these two terms helps us interpret results based on  $\xi_{ft}^k$ . When analyzing the standard deviation of  $\xi_{ft}^k$  for an outcome, or correlations between  $\xi_{ft}^k$  for different outcomes, the underlying innovation  $\theta_{ft}$  or its associated impact  $\alpha_{ift}^k$  could drive the results. First, note that  $\xi_{ft}^k$  is a family average of  $\alpha_{ift}^k \theta_{ft}$ .

$$\xi_{ft}^k = \left( \frac{1}{n_f} \sum_{i \in f} \alpha_{ift}^k \right) \theta_{ft} \quad (3)$$

If  $\alpha_{ift}^k$  is higher for boys, the estimate of  $\xi_{ft}^k$  will be higher for a set of brothers than a set of sisters, even if the underlying innovation  $\theta_{ft}$  for these two families is the same. Furthermore, the impact of the family innovation  $\alpha_{ift}^k$  might vary by permanent family characteristics. If  $\alpha_{ift}^k$  is larger for children of high school dropout mothers, the estimate of  $\xi_{ft}^k$  will be larger, even if the underlying innovation is the same. We can also think of birth spacing as a permanent family characteristic. For example,  $\alpha_{ift}^k$  might be larger for a two-child family where the siblings are close in age. Close spacing could then lead  $\xi_{ft}^k$  to be larger for the same underlying innovation. In analyzing the time-varying role of the family, it is important to recognize that variation can be driven both by the underlying changes that occur at the family level as well as how the children in a family react to those changes. Though I do not estimate  $\alpha_{ift}^k$ , I do calculate the variation in the family innovation  $\xi_{ft}^k$  separately for characteristics of sibling sets. For example, I calculate the dispersion in family innovations for female siblings, male siblings, and mixed-gender siblings, and I report how the dispersion varies by grade difference, which approximates age difference. These results reveal how the family innovation estimates depend on the similarity of siblings.

Equation (1) contains several important components. In the achievement regressions, the lagged test score controls for previous inputs and persisting innate ability;<sup>8</sup> in the time use regressions, the lagged outcome controls for habit persistence. This could include persistent family factors that are unobserved, like a family tradition of watching some

television program. In the teacher quality literature, this specification is often called a value-added model since controlling for the the previous year's score isolates the teacher's contribution given a student's baseline achievement. Here, we can interpret the family innovation as relative to the student's expected growth. An important assumption is that unobserved permanent family characteristics are captured by last year's outcome and are thus uncorrelated with the family innovation. While this assumption is admittedly strong, the teacher quality literature analogously assumes that teacher assignment is uncorrelated with unobserved permanent family characteristics.<sup>9</sup> The school-grade-year fixed effect is also important as it prevents me from confounding the time-varying school and neighborhood environment with the time-varying family environment. Depending on the outcome under study, these could include changes in school quality, neighborhood characteristics, local labor market conditions, and geographic and temporal variations in technology. I do not include a teacher effect in the model due to data limitations.<sup>10</sup>

Modeling the family innovation as a fixed effect makes my approach analogous to models with family fixed effects or teacher fixed effects. Some family innovations may be true shocks, like the untimely death of a grandparent or involuntary job loss, and in this sense, they are orthogonal to the other controls. However, other family innovations may be correlated with permanent family characteristics or time-varying school characteristics, such as parents' willingness to adapt discipline strategy or parental responses to changes in school inputs. It is important to allow for this correlation in my estimation strategy. Estimating a model with two high-dimensional fixed effects (in this case, the family-year effect and the school-grade-year fixed effect) is not trivial, and to this end, I utilize the estimator proposed by Correia (2014).

A key threat to interpretation is the potential for students from different families to be on different learning trajectories. In the above model, I assume that outcomes decay at the same rate for all students. However, high-SES families may be more effective at building off of initially high achievement levels. For example, evidence has shown that low-income students lose more knowledge over the summer relative to their peers (Cooper et al. 1996; Downey et al. 2004). If this is the case, the family innovation parameter would capture differential trends as well as the effect of time-varying family factors. To investigate whether differential trends affect the results, I estimate two alternative models that allow outcomes to decay more flexibly. Specifically, I estimate a version that includes a third-order polynomial of  $y_{it-1}^k$  and a version that allows the effect of  $y_{it-1}^k$  to vary by mother's education.

Another important consideration for interpretation is the relative importance of various time-varying factors. An innovation in parental support for learning could affect achievement and time use; this sort of innovation is important broadly. On the other hand, some siblings in every year may randomly be sick on the same day of the test, which would drive down their test scores in a correlated way. This sort of shock would lead sibling test scores to move together by construction, but it has no broader importance. The analogue to this example in the peer effects literature is the so-called barking dog.<sup>11</sup>

## 2.1 Calculating the (co)variation in family innovations

One way to calculate the dispersion in the family-year fixed effects is to save the estimates of the family innovations and then compute their variance. However, since the estimate of each family innovation parameter is based on a small number of observations, the

estimated variance would contain a substantial amount of estimation error and thus be biased upward. To avoid estimation error, I create a sample of sibling pairs and compute the covariance of the error terms (composite errors) for siblings in the same year.<sup>12</sup> Let students  $i$  and  $i'$  be siblings from family  $f$  in year  $t$ . Student  $i'$  attends school  $s'$ , which may be may or may not be same school as his sibling. The object I compute to find the variance of family innovations is then

$$\text{Cov}\left(\varepsilon_{ifst}^k, \varepsilon_{i'fs't}^k\right) \quad \text{for } i \neq i'. \tag{4}$$

I report the variation in family innovations as a standard deviation for ease of interpretation.

The variance of the family innovation is equal to the covariance of the composite errors as long as two conditions hold. First, the family innovation must be uncorrelated with the idiosyncratic error, i.e.,  $\text{Cov}\left(\xi_{ft}^k, v_{ifst}^k\right) = 0$ . This condition holds by construction since I model the family innovation as a fixed effect. Second, the idiosyncratic errors for siblings in the same year must be unrelated, i.e.,  $\text{Cov}\left(v_{ifst}^k, v_{i'fs't}^k\right) = 0$  for  $i \neq i'$ . This condition rules out the possibility that the sibling idiosyncratic errors are related for a reason beyond the siblings coming from the same family. For example, both siblings might be affected by a neighborhood block shock that is more local than the school level but does not occur at the family level. While this study conceives of the family innovation as primarily originating from the family unit, I do allow for this possibility in my interpretation. Strictly speaking, the family innovation captures time-varying factors shared by siblings that are not captured by the other controls in the regression.

To analyze the correlation between family innovations to different outcomes, I also use sibling pairs. Let  $k$  and  $\ell$  represent two different outcomes. Normally, the formula to find the correlation between family innovations for these two different outcomes would be

$$\text{Corr}\left(\xi_{ft}^k, \xi_{ft}^\ell\right) = \frac{\text{Cov}\left(\xi_{ft}^k, \xi_{ft}^\ell\right)}{\sqrt{\text{Var}\left(\xi_{ft}^k\right) \text{Var}\left(\xi_{ft}^\ell\right)}}. \tag{5}$$

I replace both variances in the denominator with the covariance of sibling composite errors, as explained above. I replace the numerator with the covariance of sibling composite errors for different outcomes, i.e.,  $\text{Cov}\left(\varepsilon_{ifst}^k, \varepsilon_{i'fs't}^\ell\right)$  for  $i \neq i'$ . For this substitution to be valid, the family innovation to one outcome must be uncorrelated with the idiosyncratic error for the other outcome, i.e.,  $\text{Cov}\left(\xi_{ft}^k, v_{ifst}^\ell\right) = 0$ , and the siblings' idiosyncratic errors for the two different outcomes must also be unrelated, i.e.,  $\text{Cov}\left(v_{ifst}^k, v_{i'fs't}^\ell\right) = 0$  for  $i \neq i'$ . These restrictions concern the relationship between the production processes of different outcomes. While violations are possible, it is difficult to ascertain how violations on net would affect the correlation estimates. I present the correlation results for descriptive purposes and take them as suggestive evidence only.

### 3 Data and descriptive statistics

The education data for this study is provided by the North Carolina Education Research Data Center (NCDERC) and the raw birth record files come from the North Carolina State Center for Health Statistics. I present summary statistics for these data in Table 1. The data on outcomes are from the EOG (End of Grade) files. These files contain math and reading test scores as well as student responses to a brief questionnaire on their time

**Table 1** Descriptive statistics

Sample	Full		Matched		Sibling-pair	
Student-year observations	5,896,480		3,781,981		1,241,126	
Unique students	1,832,309		1,088,628		524,792	
Unique schools	2144		2144		2141	
Match rate to birth data	0.641					
Unique mothers			696,632		226,562	
Math score (std devs)	0	(1)	−0.012	(0.988)	−0.026	(1.008)
Reading score (std devs)	0	(1)	−0.014	(0.989)	−0.061	(1.008)
Homework (std devs)	0	(1)	−0.013	(0.989)	−0.028	(0.988)
Free reading (std devs)	0	(1)	−0.048	(0.969)	−0.058	(0.968)
Computer use (std devs)	0	(1)	−0.027	(0.974)	−0.036	(0.965)
TV watching (std devs)	0	(1)	0.023	(1.003)	0.009	(1.015)
Homework (h/week)	2.41	(2.37)	2.39	(2.35)	2.35	(2.34)
Free reading (h/day)	0.80	(0.63)	0.77	(0.61)	0.77	(0.60)
Computer use (days/month)	3.01	(5.90)	2.84	(5.73)	2.80	(5.69)
TV watching (h/day)	2.51	(1.99)	2.56	(2.00)	2.53	(2.02)
White	0.558		0.600		0.595	
Black	0.276		0.293		0.305	
Hispanic	0.097		0.050		0.039	
Other race	0.037		0.029		0.031	
Multiracial	0.031		0.028		0.029	
Female	0.493		0.494		0.495	
Subsidized lunch	0.448		0.437		0.512	
Mother education			12.6		(2.5)	
Mother age			26.0		(5.9)	
Mother married			0.656		0.661	
Mother immigrant			0.072		0.057	
No father information			0.137		0.146	
Father education			12.7		(2.5)	
Father age			29.0		(6.6)	
Father immigrant			0.085		0.074	
First born			0.437		0.323	
Birth weight			7.29		(1.31)	
Alcohol when pregnant			0.012		0.012	
Tobacco when pregnant			0.172		0.172	

Sources: North Carolina Education Research Data Center and North Carolina State Center for Health Statistics. Sample: Students in grades 4–8, years 2003–2013. Standard deviations in parentheses

use outside of school. This questionnaire is administered at the time of testing. The education data cover all students in the third through eighth grade attending public school in North Carolina from 1997 to 2013. Since my econometric model requires a lagged outcome, I omit grade 3. The NCERDC matched students to the birth records of children born in North Carolina from 1987 onward. While most third graders are matched to a birth record beginning in 1997, high match rates for eighth graders do not occur until 2003. Since I rely on sets of siblings to identify my parameters of interest, I use 2003 for the first year of my sample.<sup>13</sup> While test scores are available for all years, data on homework time, free reading, and computer use was only collected through 2011, and data on TV watching was only collected through 2006. These restrictions leave a sample of almost

six million test score observations from 1.8 million students, though the sample size is less for the time use outcomes. The education data include other student demographics: sex, race/ethnicity, and subsidized lunch status. The North Carolina Department of Public Instruction (NCDPI) stopped requiring that schools report subsidized lunch status for individual students after 2006. Finally, each student-year record indicates the school attended.

I make some adjustments to the outcome variables for ease of interpretation. The test scores are reported on a developmental scale that changed twice during the sample period. I normalize test scores to be mean zero and standard deviation one by grade and year. In the raw data, the response options for the time use variables are given in ranges. In Table 6 in the Appendix, I report the original categorical responses, their frequencies, and the conversion scale.<sup>14</sup> I report results for time use outcomes in time units and standardized units. Like the test score outcomes, I standardize the time use variables to be mean zero and standard deviation one by grade and year, as students' activities likely change as they age and as technology changes over the years.<sup>15</sup> Means and standard deviations for the time use variables before standardization are in Table 1. In this pooled sample of the fourth through eighth graders, the average student spends 2.4 h per week on homework, reads in his free time 48 min per day, uses a computer at home for school work 3 days per month, and watches television 2.5 h per day.

The match rate of student-year records to birth records is 64%.<sup>16</sup> The birth records contain mother and father education, mother marital status, mother and father age, and other characteristics of the pregnancy and birth, such as alcohol and tobacco consumption. The NCERDC provided a unique mother identifier, which links the birth records and education records of siblings. The sample contains 696,632 unique mothers. I identify family innovations from the subsample of student-year observations matched to a sibling in that year.

The results presented below should be interpreted in the context of the estimation sample. To estimate a sibling-year covariance, I must observe two sibling outcomes in the same year. Since I estimate regressions for students in the fourth through eighth grade, observed sibling pairs are generally spaced no more than 5 years apart. In Section 5, I explore how the results vary by sibling spacing. Furthermore, to observe a pair of siblings, both siblings must be born in North Carolina and attending public school in the state. Finally, a child must be observed with a sibling, so family innovation results do not include only children.

In Table 1, I explore the impact of the latter two restrictions by comparing summary statistics for the full sample with the sample matched to birth data and the sibling-pair sample. While the matched sample mostly looks similar to the full sample, the sibling-pair sample appears to be negatively selected. In contrast to the mean-zero test scores for the full sample, the sibling-pair sample has a mean math score of  $-0.026$  standard deviations and a mean reading score of  $-0.061$  standard deviations. The differences between the mean time use variables for the full sample and sibling-pair sample are equally or less substantial, relative to their standard deviations. Students in the sibling-pair sample spend less time working on homework, free reading, and using a computer for school and slightly more time watching television. The ethnic composition is overall similar across samples, though the matched sample has a lower proportion of Hispanic students. The discrepancy is likely because North Carolina experienced a boom in Hispanic



immigration during this time period.<sup>17</sup> The rate of subsidized lunch eligibility is higher in the sibling-pair sample, indicating that these students are poorer. Since children from larger families tend to be poorer and have lower human capital, we would expect that students in the sibling-pair sample are negatively selected.

Finally, Table 1 gives summary statistics for variables only in the birth records. At the student-year observation level, the average mother was 26 years old at the time of the child’s birth and had completed high school. Sixty-six percent of mothers were married. The probability a student’s mother was born outside of the USA is 7.2% in the matched sample but 5.7% in the sibling-pair sample. Since the NCERDC matched siblings through the mother, two siblings could in fact be half-siblings. Then the father information, if given, would not necessarily match. For 15% of observations in the sibling-pair sample, no information on the father is present. For the rest, the average father was 29 years old at the time of birth and had completed high school. The sibling-pair sample averages are not notably different for these variables with one mechanical exception: There are fewer first borns in the sibling-pair sample since some first borns are only children.

#### 4 Results

The goal of the analysis is to quantify the importance of time-varying family circumstances for achievement and time use. Table 2 reports the standard deviation of family innovations for the full set of outcomes I study, which come from the pooled sample of the fourth through eighth graders.<sup>18</sup> Starting with the baseline model in the first column, we can see that these time-varying circumstances matter.

For math scores, I estimate the standard deviation of family innovations to be 0.17; for reading scores, it is 0.19. In other words, a one standard deviation change in the distribution of family innovations is associated with a 0.17 (0.19) standard deviation change in a student’s math (reading) achievement. In the literature on student test scores, estimates of this size are considered economically meaningful. For example, in a recent paper on the impact of teachers on student test scores, Chetty et al. (2014) find that a one standard deviation increase in teacher quality increases math scores by 0.14 standard deviations and English scores by 0.1 standard deviations; Rothstein (2017) replicates those results with the NCERDC data.<sup>19</sup> An important caveat is that the family innovation parameter

**Table 2** Standard deviation of family innovations with model comparison

Outcome	Model				Sibling-pair observations
	Baseline <sup>a</sup> (SD units)	Baseline (time units)	Polynomial of lags <sup>b</sup> (SD units)	Lag-mom education interactions <sup>c</sup> (SD units)	
Math score	0.170		0.169	0.184	721,443
Reading score	0.188		0.187	0.208	715,210
Homework (h/week)	0.207	0.488	0.206	0.209	505,669
Free reading (h/day)	0.248	0.154	0.248	0.257	504,257
Computer use (days/month)	0.239	1.405	0.235	0.242	504,157
TV watching (h/day)	0.284	0.565	0.282	0.296	266,263

Sources: North Carolina Education Research Data Center and North Carolina State Center for Health Statistics

<sup>a</sup>Regression estimates for the baseline model are in Table 7 in the Appendix

<sup>b</sup>Model includes a third-order polynomial of lagged outcome

<sup>c</sup>Model includes interactions between lagged outcome and five categories of mother’s education

may pick up some common shocks to siblings that do not originate in the family. Still, the results in this paper suggest that the time-varying role of the family is more important than teacher assignment for student learning. They also reveal a bigger role for a bundle of time-varying factors than any single time-varying trait or event. For example, Lyle (2006) finds that a parental absence decreases a student's test score by one tenth of a standard deviation in the year of the absence. Contrasting the results for the two subjects, we see that time-varying family circumstances are relatively more important for reading skills compared to math skills. This evidence suggests that families have a comparative advantage in reading.

The results for student time use also reveal a substantial role for the time-varying family environment. I pair homework time and free reading together as educational uses of time, and I put computer use and TV watching together as media exposure. Family innovations to educational time use could include more or less oversight after school, a new system to ensure homework is completed, or moving closer to a library. Analysis by Carneiro and Ginja (2016) shows that income shocks influence parental time investments. Large changes to media exposure at home could be driven by the purchase of a new television or computer or an expansion of services (e.g., more TV channels or an upgrade in Internet speed). Changes in parental oversight could also be included in the family innovation to media exposure. There are many possibilities, though my measure only captures their net effect.

I present the time use results in standard deviation units and time units to show the relative and absolute importance of family innovations. In relative terms, year-to-year variation in the family environment is more important for TV watching ( $SD = 0.28$ ), somewhat important for free reading (0.25) and computer use (0.24), and less important for homework time (0.21). These results reflect the fact that homework load is driven by teachers, schools, and perhaps community preferences, as opposed to any single family. Television viewing, however, is determined outside of school hours, so there is more scope for time-varying family factors to influence TV time.

Turning to time units, the standard deviation of family innovations for homework time is about one-half hour per week. For time spent free reading, the standard deviation of family innovations is 0.15 h, or 9 min, per day. Although these may seem like small increases in time use if taken in isolation, the cumulative effects are more substantial. Assuming a 36-week school year where students meet 5 days per week, a one standard deviation move up in the family innovation distribution for homework is associated with 18 more hours of homework time. For free reading, that move is associated with 28 more hours of reading during the school year. Fiorini and Keane (2014) find that educational activities like these are the most productive uses of time for cognitive development, so increases to homework and reading time should translate into test score gains.

Time spent watching TV and using a computer reflect media exposure. The standard deviation of family innovations for TV watching is 0.57 h per day, or 34 min. For computer use, it is about 1.4 days per month. The variation in computer use seems less economically significant, and in any case, it only reflects how often a student uses a computer for school work, not total use. However, the time-varying family environment plays a substantial role in how much television a child watches year-to-year. A one standard deviation change in the distribution of family innovations is associated with an almost 102-h change in TV watched on school days over the course of the school year.

In the context of the decomposition of the family innovation, the variation in family innovations comes from both the underlying family innovation  $\theta_{ft}$  and the impact of the underlying innovation  $\alpha_{ift}^k$  for the children observed. On one extreme, the dispersion in family innovations could arise from families responding differently to the same underlying innovation  $\theta_{ft}$ . On the other extreme,  $\alpha_{ift}^k$  could be the same for each child in each family for a particular outcome, but every family could experience a different underlying innovation. There is likely some role for both the underlying innovation and its child-specific impact. In the example of parental oversight after school, children may respond differently to the same increase in oversight, and parents may increase or decrease their oversight in a given year. Here,  $\theta_{ft}$  is only observed through the impact it has, so I analyze the dispersion of  $\alpha_{ift}^k \theta_{ft}$ . In Section 5, I explore whether estimates for the standard deviation of family innovations are lower, for example, if age differences between siblings are greater. All together, these results reveal that there is a larger role for the time-varying family environment than perhaps previously thought.

Next, I examine how family innovations to different outcomes move together.<sup>20</sup> For instance, suppose there is a positive correlation between family innovations to free reading and family innovations to reading scores. This would tell us that family innovations that increase free reading time also tend to increase reading scores. As noted above, the correlation could be driven by the underlying family innovation as well as the child-specific impacts. I explore these possibilities through the correlations in Table 3.

The correlation between family innovations to math and family innovations to reading is 0.73, indicating that the time-varying family environment has a similar effect on a student’s cognitive development across subjects. I also find positive relationships between family innovations to achievement and family innovations to educational time use. In other words, time-varying family factors that lead to increases in homework time and free reading also lead to increases in test scores. By the same token, family innovations that lead to decreases in educational time use tend to lead to decreases in test scores. The relationship between family innovations to free reading and family innovations to reading scores ( $\rho = 0.27$ ) is stronger than the relationship between family innovations to free reading and family innovations to math scores ( $\rho = 0.19$ ). Conversely, family innovations to homework time are more strongly associated with family innovations to math scores ( $\rho = 0.24$ ) compared to reading scores ( $\rho = 0.21$ ). These correlations fit with the idea that a student can boost his reading skills from reading more on his own, but improving math skills comes from more structured practice, which schools facilitate.

Turning to the relationship between family innovations to achievement and family innovations to media time, we see that the magnitudes are all lower compared to the

**Table 3** Correlations between family innovations to different outcomes

	Math score	Reading score	Homework	Free reading	Computer use	TV watching
Math score	1.000					
Reading score	0.727	1.000				
Homework	0.235	0.206	1.000			
Free reading	0.186	0.274	0.214	1.000		
Computer use	0.023	−0.009	0.122	0.020	1.000	
TV watching	−0.114	−0.070	−0.072	−0.020	−0.072	1.000

Sources: North Carolina Education Research Data Center and North Carolina State Center for Health Statistics

correlations with educational uses of time. They are also all close to zero or negative, ranging from 0.02 to  $-0.11$ . The correlations between family innovations to different uses of time are generally weak, with the exception of free reading and homework time ( $\rho = 0.21$ ). In a given family, a year that sees an increase in free reading beyond any expected growth is also a year that sees an increase in time spent on homework. Certain changes in parent behavior, like less oversight after school, could decrease time spent on homework as well as free reading.

#### 4.1 Robustness to differential trends

Given that students from different families may be on different learning trajectories, I estimate two alternative versions of my model that allow the lagged outcome to enter more flexibly into the model. I present these results in the third and fourth columns of Table 2. In the first alternative model, I include a third-order polynomial of the lagged outcome. This allows outcomes to grow more flexibly from the baseline. In the second, I include a full set of interactions between the lagged outcome and mother's education, a key measure of family socioeconomic status. This specification lets students from higher SES families have faster or slower growth in their outcomes. Comparing the columns of Table 2, we see that these alternative specifications do not noticeably alter the results. Thus, I have no evidence that differential growth trends contaminate my estimates.

## 5 Heterogeneity

In this section, I discuss heterogeneity in (i) the dispersion of family innovations by sibling characteristics and (ii) the correlations between family innovations by family characteristics.

Heterogeneity in the dispersion of family innovations reveals how closely sibling outcomes track based on the degree of similarity between the siblings. They also indicate how sensitive results are to the ages available in my sample. Table 4 reports the standard deviation of family innovations by the gender mix of sibling pairs and the grade difference of sibling pairs.<sup>21</sup> The standard deviation of family innovations to achievement is relatively stable across sibling gender composition, with male pairs of siblings having slightly higher

**Table 4** Standard deviation of family innovations by sibling characteristics

	Math score	Reading score	Homework	Free reading	Computer use	TV watching
Gender						
Male pair	0.183	0.202	0.206	0.253	0.224	0.313
Female pair	0.168	0.189	0.222	0.271	0.271	0.296
Mixed pair	0.165	0.180	0.201	0.232	0.230	0.260
Grade difference						
0 years	0.209	0.217	0.258	0.287	0.309	0.299
1 year	0.168	0.193	0.218	0.245	0.240	0.300
2 years	0.169	0.186	0.202	0.252	0.235	0.288
3 years	0.159	0.175	0.187	0.233	0.217	0.262
4 years	0.154	0.174	0.175	0.222	0.200	0.245
All	0.170	0.188	0.207	0.248	0.239	0.284

Sources: North Carolina Education Research Data Center and North Carolina State Center for Health Statistics. All variables in standard deviation units

standard deviations. Recall from Eq. (4) that the variance of family innovations is calculated from the covariance of error terms for sibling pairs. Using the covariance to inform interpretation, deviations from expected growth in test scores are more closely related for male siblings, as compared to female siblings or mixed-gender siblings. Male siblings also have greater co-movement in their error terms for TV watching, while female siblings' covariance of error terms are higher for the other three time use measures (homework, free reading, and computer use). Overall, the mixed-gender pairs have the lowest co-movement in their error terms across outcomes, which is consistent with same-gender children within a family having more similar experiences.

The grade difference for siblings is close to their age difference. Table 4 reveals that siblings who are closer in age have closer deviations from their expected growth in all outcomes. These results suggest that if my sample included a wider range of ages, the estimates of the standard deviation of family innovations would be lower. In addition, siblings with closer spacing appear more times in the sample<sup>22</sup>, which also suggests that the true dispersion in family innovations is lower than the estimates reported in Section 4. To get a sense of how much the sample composition affects the main results, consider that the average sibling age difference is about 3 years.<sup>23</sup> For each outcome, the standard deviation of family innovations for the sample with a 3-year grade difference is lower than the standard deviation for the full sibling-pair sample, but it is not substantially lower. Thus, the conclusions of this paper hold for a sample of siblings with the average age difference, which suggests that the limited range of age differences in my data is not a major concern.

Next, I turn to heterogeneity in the correlation results. The correlations between family innovations to different outcomes tell us whether time-varying family circumstances that affect one outcome also tend to affect others, and if so, in what direction. However, these correlations need not be the same for all types of families. It could be the case, for example, that family innovations to free reading and family innovations to reading scores move together more strongly for high-SES families than low-SES families or vice versa.

Table 5 presents correlations between family innovations conditioning on selected family characteristics. Mother's education offers the finest classification of families by

**Table 5** Correlations between family innovations by family characteristics

	Homework		Free reading		Computer use		TV watching	
	Math	Read	Math	Read	Math	Read	Math	Read
Mother's education								
Less than high school	0.292	0.236	0.239	0.296	0.007	0.016	-0.130	-0.045
High school	0.276	0.226	0.192	0.304	0.109	-0.028	-0.063	-0.029
Some college	0.192	0.210	0.129	0.215	0.027	-0.014	-0.102	-0.076
College	0.212	0.213	0.145	0.243	0.024	0.011	-0.217	-0.184
Graduate school	0.096	0.051	0.200	0.286	0.153	0.018	-0.185	-0.267
Income								
Subsidized lunch	0.302	0.257	0.216	0.284	-0.001	0.000	-0.073	-0.024
No subsidized lunch	0.260	0.221	0.189	0.233	0.033	-0.009	-0.162	-0.126
Mother's ethnicity								
White	0.227	0.186	0.159	0.251	0.008	-0.003	-0.176	-0.155
Black	0.237	0.244	0.215	0.297	0.019	-0.016	0.001	0.050
Hispanic	0.308	0.381	0.177	0.351	0.007	0.038	0.010	0.083
All	0.235	0.206	0.186	0.274	0.023	-0.009	-0.114	-0.070

Sources: North Carolina Education Research Data Center and North Carolina State Center for Health Statistics

socioeconomic status, but the patterns are supported by the available measure of income, subsidized lunch status. The correlations between family innovations to homework time and family innovations to math scores vary somewhat by mother's education, from 0.29 for high school dropout mothers to 0.10 for mothers who went to graduate school. A similar pattern is present for reading. For low-SES families, family innovations that increase homework time also tend to increase test scores. While the positive association is still present for high-SES families, it is less strong. The correlations by mother's ethnicity also support this pattern since black and Hispanic families are on average more disadvantaged. Although there is some heterogeneity in the association between family innovations to free reading and family innovations to achievement, a clear pattern by SES does not emerge.

Why might there be stronger correlations between family innovations to homework time and family innovations to test scores in low-SES families? The association could come from differences in the types of family innovations that low-SES and high-SES families experience (i.e., differences in the underlying innovation  $\theta_{ft}$ ). For example, parental job loss could be more likely in low-SES households, or family income might vary more widely. Another possibility is that low-SES and high-SES families are impacted differently by the same events or the same changes in inputs (i.e., differences in  $\alpha_{ift}^k$ ). For example, a negative health shock may be more devastating in a low-SES family that has little savings to rely on. These correlations help tell a story about the role of the time-varying family environment for different outcomes, but work on specific events, input changes, and shocks is needed to better understand these associations.

The last four columns of Table 5 examine the correlations between family innovations to media exposure and family innovations to achievement. Here, we see some trends by socioeconomic status and even the occasional reversal of sign. While it is generally true that the relationship between family innovations to computer use and family innovations to test scores is close zero, there is a more substantial positive relationship in one instance: family innovations to math scores for children with a mother with a graduate degree. For family innovations to TV watching, the correlations with family innovations to achievement are more negative for high-SES families. However, when comparing across ethnic groups, we see that they are positive for family innovations to reading scores for black and Hispanic families. These results suggest that changes to media exposure have different impacts in different families, perhaps due to differences in the content itself but perhaps due to what the media time replaces (Morgan and Gross 1980).

## 6 Conclusions

While a growing literature examines the impact of specific events and input changes on student outcomes, this paper uses the co-movement of sibling error terms from the same year to quantify the total importance of the time-varying family environment. I find that this bundle of annual innovations plays a substantial role in student time use and achievement. The results indicate that my measure of family innovation is more important for student achievement than teacher assignment. In addition, a one standard deviation move in the distribution of family innovation parameters is associated with changes in time use of 18 h of homework time, 28 h of free reading, and 102 h of TV watching over the course of the school year. Correlations between family innovations suggest that when a family places more emphasis on homework and free reading, children's test scores are likely to

increase. However, the scope for these increases would be greater for low-SES families than high-SES families. The estimates presented here likely pick up some factors that are not common to a family in a given year, in which case they would overstate the true variation in family innovations. Still, this research serves as a complement to studies on specific family-level changes by providing a big-picture view on the relative importance the time-varying family environment.

In this paper, I examine the role of family innovations for contemporaneous outcomes. Another important question is whether the effects of any of these year-to-year changes in family life persist into adulthood. Studies of the effects of specific inputs and events suggest that the answer is yes. For example, Oreopoulos et al. (2008) find that experiencing parental job loss as a child lowers earnings by 9%. While the data used in this study are not well-suited to study long-run outcomes, the results presented here suggest that the time-varying role of the family for adult outcomes is an important area for future work.

### Endnotes

<sup>1</sup> Plausibly causal estimates indicate a positive relationship between parent's schooling and various student outcomes (Björklund and Salvanes 2011). With the increased prevalence of mixed-race families, mother's race is a better measure of family race than child's race. Arcidiacono et al. (2015) find that children of black and Hispanic mothers score about 0.4 standard deviations lower on math tests than children of white mothers after controlling for various family and school characteristics.

<sup>2</sup> Black et al. (2017) examine the effect of having a third child with a disability on the first two children. The unique data sets and novel identification strategy employed in the paper illustrate how difficult it can be to isolate these channels.

<sup>3</sup> The only other similarly linked data set comes from Florida. Figlio et al. (2014), among others, use this data to study the effect of neonatal health on educational outcomes. A handful of recent papers also use the matched North Carolina data, mostly to study the impact of early childhood education on later outcomes.

<sup>4</sup> Fiorini and Keane (2014) provide evidence on the importance of time use, especially the time spent in educational activities, for cognitive development.

<sup>5</sup> These estimates are typical in this literature. Rockoff (2004), Rivkin et al. (2005), Aaronson et al. (2007), and Kane and Staiger (2008) all put the impact of a one standard deviation better teacher on test scores between 0.08 and 0.15 standard deviations. Rothstein (2017) replicates the Chetty et al. (2014) results with the school data used in this paper and reproduces all key results on teacher value-added.

<sup>6</sup> There are some similarities in the approach taken by Carneiro and Ginja (2016) and this paper. Carneiro and Ginja (2016) assume exogeneity of income shocks (defined as fluctuations in income conditional on child fixed effects and time-varying controls) but recognize that this is a strong assumption. If I assumed that the family innovation was exogenous, the interpretation of that parameter would be different—any shared changes in sibling outcomes would come from exogenous factors, and I would rule out the possibility that parents plan changes in inputs. With that channel shut down, the conclusions reached in this paper would be stronger; for example, I might recommend policies that help families insure against negative shocks. Acknowledging that the family innovation is not completely exogenous, I interpret the family innovation parameter more carefully.

<sup>7</sup> Consider, for example, the interpretation of sibling correlations described in Björklund et al. (2010): “A sibling correlation can thus be thought of as an omnibus measure of the importance of family background and community effects. It includes anything shared by siblings: parental income and parental influences such as aspirations and cultural inheritance, as well as things not directly experienced in the home, such as school, church and neighborhood effects.”

<sup>8</sup> Many papers in the teacher quality literature also use lagged test score for this purpose (e.g., Aaronson et al. 2006; Clotfelter et al. 2007; Todd and Wolpin 2007).

<sup>9</sup> Todd and Wolpin (2003) offer a thorough exploration of the assumptions behind various forms of the educational production function.

<sup>10</sup> The NCERDC identifies the individual who proctors the test. For elementary school students, the proctor is usually the student’s classroom teacher, but this is not the case for middle school students. Including middle school students in my sample is critical for my analysis since many of the sibling pairs include a student in middle school.

<sup>11</sup> A barking dog outside the classroom on the day of the test would be a common shock to a group of classroom peers. The students in the class would have similarly low test scores, but not through the mechanism of peer effects.

<sup>12</sup> Note that in my estimation of the educational production function, I estimate the family-year effect as a fixed effect. However, I use the composite error defined in Eq. (2), which is the sum of the family innovation and the rest of the error, in my estimation of the dispersion of family innovations.

<sup>13</sup> While starting my sample in 1997 would yield more sibling pairs, these pairs would be biased toward siblings who were closer in age. Starting the sample in 2003 makes my estimates more representative of sibling pairs who were born up to 5 years apart.

<sup>14</sup> I convert the ranges to a continuous scale using the midpoint of the range. For the top option, I use a value close to the lower bound. For computer use, the student had the option to indicate that he does not have a computer at home, and for homework time, he could report that his teacher does not assign homework. In both of these instances, I convert the response to zero.

<sup>15</sup> The regression model includes a school-grade-year effect, so time trends and grade-level trends are still removed from the family innovation results that I report in time units. The descriptive statistics are affected by units, however.

<sup>16</sup> The match rate is perhaps lower than one might expect due to a high rate of migration into North Carolina during this time period.

<sup>17</sup> Hull (2017) details the differences between North Carolina public school students that do and do not match to a North Carolina birth record, finding that the matched Hispanics are positively selected on test scores.

<sup>18</sup> Estimated coefficients from the outcome regressions appear in Table 7 in the Appendix.

<sup>19</sup> In the NCERDC data, students can only reliably be matched to teachers in elementary grades, and so, Rothstein (2017) limits his sample to grades 3–5. For the purpose of this analysis, including teacher fixed effects in Eq. (1) is undesirable since they could not be estimated for many of the students included in the analysis.

<sup>20</sup> The students answer the time use questions at the same time that they take their achievement tests. Because of the potential for endogeneity, I do not estimate the effect of time use on achievement.



<sup>21</sup> Due to the large sample size, the differences by gender and grade are almost always statistically significant, except in cases where the point estimates are very close.

<sup>22</sup> For example, a sibling pair that is one grade apart could potentially appear in the sibling-pair sample four times, from when the younger sibling is in the fourth grade and the older sibling is in the fifth grade to when the younger sibling is in the seventh grade and the older sibling is in the eighth grade. In contrast, a sibling pair that is four grades apart would appear only once, when the younger sibling is in the fourth grade and the older sibling is in the eighth grade.

<sup>23</sup> Using children of mothers in the NLSY79 and the 1988 Natality Detail files, Buckles and Munnich (2012) find that mean birth spacing is 40.8 months and the median is 34 months in both data sets.

## Appendix

**Table 6** Conversion of time use categories to continuous values

Response <sup>a</sup>	Relative frequency	Assigned value
Time on homework		h/week
Has homework, but does not do it	0.012	0
Less than 1 h each week	0.318	0.5
Between 1 and 3 h	0.405	2
More than 3 but less than 5 h	0.145	4
Between 5 and 10 h	0.088	7.5
More than 10 h	0.017	12
No homework is ever assigned	0.015	0
Amount of time spent free reading		h/day
None	0.122	0
About 30 min	0.484	0.5
About 1 h	0.206	1
Between 1 and 2 h	0.118	1.5
More than 2 h	0.070	2.5
Student uses computer at home		days/month
I use a computer at home for school work almost every day	0.065	24
Once or twice a week	0.172	6
Once or twice a month	0.178	1.5
Hardly ever	0.324	0.5
Never, even though there is a computer at home	0.155	0
There is no computer at home	0.106	0
TV watched at home each school day		h/day
None	0.048	0
1 h or less each school day	0.270	0.5
2 h	0.256	2
3 h	0.194	3
4 to 5 h	0.139	4.5
6 h or more	0.094	7

Source: North Carolina Education Research Data Center

<sup>a</sup>Variable titles and labels as they appear in the codebook

**Table 7** Outcome regression results

	Math score	Reading score	Homework	Free reading	Computer use	TV watching
Lagged outcome	0.769 (0.000)	0.735 (0.000)	0.193 (0.001)	0.409 (0.001)	0.167 (0.001)	0.403 (0.001)
Mother's education						
High school	0.025 (0.003)	0.027 (0.004)	0.020 (0.007)	0.004 (0.007)	0.018 (0.007)	-0.013 (0.008)
Some college	0.043 (0.005)	0.045 (0.005)	0.034 (0.009)	0.006 (0.009)	0.030 (0.010)	-0.019 (0.011)
College	0.068 (0.007)	0.070 (0.007)	0.059 (0.013)	0.025 (0.013)	0.017 (0.014)	-0.072 (0.017)
Graduate school	0.075 (0.009)	0.082 (0.010)	0.061 (0.017)	0.037 (0.017)	0.029 (0.018)	-0.108 (0.022)
Father's education						
High school	0.026 (0.004)	0.029 (0.004)	0.022 (0.007)	0.003 (0.007)	0.005 (0.007)	-0.009 (0.008)
Some college	0.038 (0.005)	0.039 (0.005)	0.027 (0.009)	0.012 (0.009)	0.008 (0.010)	-0.035 (0.012)
College	0.064 (0.007)	0.068 (0.007)	0.057 (0.013)	0.030 (0.012)	0.011 (0.013)	-0.071 (0.016)
Graduate school	0.064 (0.008)	0.078 (0.009)	0.069 (0.016)	0.042 (0.016)	0.028 (0.017)	-0.082 (0.021)
Black	-0.205 (0.001)	-0.233 (0.001)	-0.124 (0.002)	-0.092 (0.002)	0.045 (0.002)	0.371 (0.003)
Hispanic	-0.151 (0.001)	-0.230 (0.001)	-0.098 (0.002)	-0.117 (0.002)	-0.010 (0.002)	0.081 (0.003)
Other race	0.096 (0.002)	-0.053 (0.002)	0.063 (0.004)	0.075 (0.004)	0.175 (0.004)	-0.021 (0.005)
Multiracial	-0.064 (0.002)	-0.068 (0.002)	-0.055 (0.004)	-0.004 (0.004)	0.008 (0.004)	0.135 (0.006)
Female	0.020 (0.001)	0.054 (0.001)	0.072 (0.001)	0.194 (0.001)	0.070 (0.001)	-0.061 (0.002)
Birth data	-0.081 (0.147)	-0.007 (0.161)	0.268 (0.298)	0.005 (0.291)	0.117 (0.305)	0.062 (0.351)
Mother age	0.002 (0.001)	0.001 (0.001)	0.006 (0.001)	-0.001 (0.001)	0.006 (0.001)	0.003 (0.001)
Mother married	0.002 (0.003)	-0.004 (0.003)	0.022 (0.006)	0.000 (0.006)	0.018 (0.006)	0.002 (0.008)
Mother immigrant	0.019 (0.020)	0.047 (0.021)	0.067 (0.039)	0.088 (0.038)	-0.002 (0.040)	0.030 (0.050)
No father information	0.011 (0.009)	0.004 (0.010)	-0.003 (0.016)	-0.014 (0.016)	-0.024 (0.017)	0.011 (0.017)
Father age	-0.001 (0.000)	-0.001 (0.000)	-0.002 (0.001)	0.000 (0.001)	-0.001 (0.001)	0.002 (0.001)
Father immigrant	0.026 (0.009)	0.026 (0.010)	0.037 (0.020)	0.027 (0.019)	-0.012 (0.020)	-0.007 (0.027)
First born	0.020 (0.002)	0.035 (0.002)	0.031 (0.003)	0.039 (0.003)	0.025 (0.003)	-0.001 (0.004)
Birth weight	0.007 (0.001)	0.006 (0.001)	0.001 (0.002)	-0.001 (0.001)	0.002 (0.002)	0.005 (0.002)
Alcohol when pregnant	-0.006 (0.008)	-0.009 (0.008)	0.013 (0.014)	-0.000 (0.014)	-0.005 (0.014)	0.004 (0.015)
Tobacco when pregnant	-0.012 (0.003)	-0.013 (0.004)	-0.006 (0.006)	0.003 (0.006)	0.000 (0.007)	-0.004 (0.008)
Observations	5,886,058	5,855,758	4,394,084	4,387,494	4,387,242	2,077,296
R <sup>2</sup>	0.848	0.822	0.590	0.613	0.573	0.659

Sources: North Carolina Education Research Data Center and North Carolina State Center for Health Statistics. All outcomes measured in standard deviation units. Standard errors in parentheses. Regressions also include missing indicators, school-grade-year fixed effects, and family-year fixed effects

### Acknowledgements

The author would like to thank Peter Arcidiacono, Jared Ashworth, Esteban Aucejo, Jeremy Bray, Brian Clark, Julie Cullen, Mary Daly, Hugh Macartney, Arnaud Maurel, Tyler Ransom, Jesse Rothstein, Seth Sanders, and the seminar participants at Duke, UC Berkeley, Utah State, SUNY Buffalo, UNC Greensboro, the 2015 Southern Economic Association Meetings, and the 2016 European Association of Labour Economists Conference for the helpful discussions and comments on the earlier drafts. The author would also like to thank the anonymous referees and the editor for the helpful comments. Responsible editor: Joni Hersch

### Funding

The author has no sources of funding to declare.

### Availability of data and material

This paper uses data provided for the North Carolina Education Research Data Center. Details of how to obtain the data can be found at <https://childandfamilypolicy.duke.edu/research/nc-education-data-center/>.

### Competing interests

The *IZA Journal of Labor Economics* is committed to the IZA Guiding Principles of Research Integrity. The author declares that she has observed these principles.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 15 August 2017 Accepted: 25 September 2017

Published online: 26 October 2017

### References

- Aaronson D, Barrow L, Sander W (2007) Teachers and student achievement in the Chicago Public High Schools. *J Labor Econ* 25(1):95–135
- Ananat EO, Gassman-Pines A, Gibson-Davis CM (2011) The effect of local employment losses on children's educational achievement. In: Duncan GJ, Murnane RJ (eds). *Whither opportunity? rising inequality and the uncertain life chances of low-income children*. Russel Sage Publications, New York. pp 299–314
- Arcidiacono P, Beauchamp A, Hull M, Sanders S (2015) Exploring the racial divide in education and the labor market through evidence from interracial families. *J Human Capital* 9(2):198–238
- Björklund A, Jäntti M (2012) How important is family background for labor-economic outcomes? *Labour Econ* 19:465–74
- Björklund A, Lindahl L, Lindquist MJ (2010) What more than parental income, education and occupation? An exploration of what Swedish siblings get from their parents. *B E J Econ Anal Policy (Contributions)* 10(1). Article 102
- Björklund A, Salvanes KG (2011) Education and family background: mechanisms and policies. In: Hanushek EA, Machin S, Woessman L (eds). *Handbook of the Economics of Education*. Elsevier, Amsterdam Vol. 3. pp 201–47
- Black SE, Breining S, Figlio DN, Guryan J, Karbownik K, Nielsen HS, Roth J, Simonsen M (2017) Sibling spillovers. NBER Working Paper 23062
- Buckles KS, Munnich ES (2012) Birth spacing and sibling outcomes. *J Human Resources* 47(3):613–42
- Carneiro P, Ginja R (2016) Partial insurance and investments in children. *Econ J* 126(596):F66–F95
- Chetty R, Friedman JN, Rockoff JE (2014) Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates. *Am Econ Rev* 104(9):2593–2632
- Clotfelter CT, Ladd HF, Vigdor JL (2006) Teacher-student matching and the assessment of teacher effectiveness. *J Human Resources* 41(4):778–820
- Cooper H, Nye B, Charlton K, Lindsay J, Greathouse S (1996) The effects of summer vacation on achievement test scores: a narrative and meta-analytic review. *Rev Educ Res* 66(3):227–68
- Correia S (2014) REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. *Statistical Software Components*, Boston College Department of Economics, Revised 25 July 2015
- Dahl GB, Lochner L (2012) The impact of family income on child achievement: evidence from the earned income tax credit. *Am Econ Rev* 102(5):1927–56
- Downey DB, von Hippel PT, Broh BA (2004) Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *Am Sociol Rev* 69(5):613–35
- Figlio D, Guryan J, Karbownik K, Roth J (2014) The effects of poor neonatal health on children's cognitive development. *Am Econ Rev* 104(12):3921–55
- Fiorini M, Keane MP (2014) How the allocation of children's time affects cognitive and noncognitive development. *J Labor Econ* 32(4):787–836
- Hull MC (2017) The academic progress of Hispanic immigrants. *Econ Educ Rev* 57:91–110
- Kane TJ, Staiger DO (2008) Estimating teacher impacts on student achievement: an experimental evaluation. NBER Working Paper 14607
- Lyle DS (2006) Using military deployments and job assignments to estimate the effect of parental absences and household relocations on children's academic achievement. *J Labor Econ* 24(2):319–50
- Milligan K, Stabile M (2011) Do child tax benefits affect the well-being of children? Evidence from Canadian child benefit expansions. *Am Econ J Econ Policy* 3(3):175–205
- Morgan M, Gross L (1980) Television viewing, IQ and academic achievement. *J Broadcast* 24(2):117–33
- Oreopoulos P, Page M, Stevens AH (2008) The intergenerational effects of worker displacement. *J Labor Econ* 26(3):455–83
- Rivkin SG, Hanushek EA, Kain JF (2005) Teachers, schools, and academic achievement. *Econometrica* 73(2):417–58
- Rockoff JE (2004) The impact of individual teachers on student achievement: evidence from panel data. *Am Econ Rev* 94(2):247–52

- Rothstein J (2017) Measuring the impacts of teachers: comment. *Am Econ Rev* 107(6):1656–84
- Stevens AH, Schaller J (2011) Short-run effects of parental job loss on children's academic achievement. *Econ Educ Rev* 30:289–99
- Todd PE, Wolpin KI (2003) On the specification and estimation of the production function for cognitive achievement. *Econ J* 113(485):F3–F33
- Todd PE, Wolpin KI (2007) The production of cognitive achievement in children: home, school, and racial test score gaps. *J Human Capital* 1(1):91–136

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---