No. 85

UNIVERSITY OF COLOGNE

WORKING PAPER SERIES IN ECONOMICS

LEISURE AND LEARNING - ACTIVITIES AND THEIR EFFECTS ON CHILD SKILL DEVELOPMENT

PETER FUNK
THORSTEN KEMPER

Department of Economics
University of Cologne
Albertus-Magnus-Platz
D-50923 Köln
Germany

http://www.wiso.uni-koeln.de

Leisure and Learning—
Activities and Their Effects on Child Skill Development

Peter Funk Thorsten Kemper
University of Cologne*

February 8, 2016

Abstract

This paper studies how variations in leisure time allocation help explain the variations in school children's cognitive skills. We use representative data on the time use of American children from the Child Development Supplement (CDS) to the the Panel Study of Income Dynamics (PSID). Our findings suggest that 1) including time use data significantly contributes to explaining the variation in math and reading test scores; 2) in a relative ranking of the effect of raising the time spent on a given activity on the math test score music is placed at the top, followed by learning, reading, sports, watching television, attending school and sleep (in descending order). For the reading test score music ranks first again and reading second, before learning, school, television, sports and sleep; 3) when comparing the effect of child activities with that of parental investments on test scores in the PSID data, it turns out that activities have no less explanatory power than investments, proxied by an established investment measure, with higher explanatory power for the production of math skills.

JEL Classification: D13, I21, J13, J24
Keywords: Child development, leisure time activities

*Center for Macroeconomic Research, University of Cologne, Albertus-Magnus-Platz, D-50923 Köln, Germany, Tel.: +49/221/4704496, Fax: +49/221/4705143, e-mail: funk@wiso.uni-koeln.de

1 Introduction

Cognitive and non-cognitive skills play an important role in explaining labor market outcomes (Heckman et al., 2006). These skills emerge and are influenced during different phases of childhood (see Almond and Currie (2011); Cunha et al. (2006) for overviews). In empirical studies, cognitive skills are usually proxied by scores obtained from test batteries. A large number of studies have assessed the importance of various factors in explaining differences in test score achievement of children. One literature strand focuses on parental investments into child skills, which are conceptualized in different ways. A number of studies focus on aggregate investment measures, notably Todd and Wolpin (2007), Cunha and Heckman (2008), and Cunha et al. (2010). These include some information about the joint time use of parents and children. Other studies analyze more specifically how parental time inputs affect child skill development. Recently, an emerging literature strand extends the latter by studying how the allocation of children's time affects cognitive skill development, see for instance Fiorini and Keane (2014).

The present paper extends the empirical literature by analyzing how the allocation of spare time towards extracurricular educational activities and a variety of leisure activities (in a broad sense) helps explain the variation in school children's cognitive skills. We use representative data on the time use of American children from the Child Development Supplement (CDS) to the Panel Study of Income Dynamics (PSID). In short we want to answer the following questions:

1. Do children's time inputs matter for skill acquisition?
2. If so, which activities prove relatively beneficial or detrimental for skill acquisition? In particular, what are the relative benefits of time devoted to studying and time spent on leisurely activities?
3. How does the consideration of time allocation patterns affect the assessment of the importance of parental investments?

The third question links the paper to the aforementioned literature on the role of parental investments in the production of child skills. In this literature, parental investment is often proxied by a scalar index composed of, or a latent variable manifest in, a large number of variables that cover different aspects of parental inputs, such as investment into a home environment conducive for learning, parental time inputs, and parental pedagogical practices. Some of these investments may be strongly related to the time allocation of children. As a result, estimates of the effects of parental investments could mask effects that ultimately reflect child activities. When incorporating activities explicitly, this may change the assessment of the quantitative effects of established investment measures.

The first question has been partly addressed before, albeit with a strong focus on parental time inputs. Guryan et al. (2008) study how parental time inputs differ along dimensions such as education and income. The effect of parental time inputs on child outcomes has been studied in different settings, and with different time input measures. For example, mothers' trade-off between raising income and parenting their children along the extensive margin in the labor market decision is discussed in Blau and Grossberg (1992). The literature on early childhood development has approximated motherly time input by work force participation as well (Carneiro et al., 2010). Bernal and Keane (2010) study the influence of labor force participation and childcare on skill development. Other studies have exploited more detailed input measures stemming from time use studies. Del Boca et al. (2014) estimate a structural model of child development that includes measures of "active" and "passive" parental child care time inputs that are obtained from the CDS. Another strand singles out specific parental time inputs, notably reading to children (Price, 2008; Kalb and Van Ours, 2014). Hsin and Felfe (2014) study the effects of parents' time inputs to educational, "structured", and "unstructured" activities.

An emerging literature explicitly introduces child activities in one way or another and provides more detailed answers to questions one and two. Fiorini and Keane (2014) study the effects of eight different activities on skill development simultaneously in a sample of young Australian children, with a strong focus on time spent jointly with parents or other adults.[1] They find that educational activities, in particular with parents, are most effective in raising cognitive skills. Del Boca et al. (2012) focus on relative contributions of one child activity measure that lumps together different activities a priori thought to positively influence skill acquisition, and one parental time input measure, in a sample of adolescents obtained from the CDS. They find that the child time inputs explain a higher share of the variation in test scores than than mothers' time inputs for adolescents, while the role is reversed for younger children. Other studies concentrate on the effects of a single extracurricular activity, measured ordinally or dichotomously, on cognitive and non-cognitive skill formation. Felfe et al. (2012) study the effect of exercising sports on school grades, mainly in one large cross section of German children aged 3 to 10 years. Hille and Schupp (2015) study the effects of learning a musical instrument, operationalized by regular attendance of musical lessons outside of the school, and Cabane et al. (2015) build on the latter work by simultaneously studying the effects of music and sports. These contributions define and utilize binary "treatment" indicators of specific activities from a representative sample of German adolescents from the German Socio-Economic Panel.

We study the influence of a large number of (groups of) child activities on skill development. We partition the set of all possible time use activities into the groups Learning, Music, Reading, School, Sleep, Sports, TV/Video Games, and Other activities and include these simultaneously in a regression framework. This way we can measure the effect of shifting time among a large number of pairs of activities. Each activity (group of activities) can assume a continuum of values directly measured by its share in the total time budget of 24 hours. Studying the effects of different, continuously measured, activities simultaneously allows properly taking into account the effects of allocating time away from other activities. While the effect of engaging in, say, an extra hour of sports per day holding everything else equal is not identifiable, we can identify the relative effect of substituting between all pairs of activities. In contrast, studies that focus on a single activity cannot account for which activities are typically reduced in favor of a given activity. The measured difference in outcomes between those engaging in one particular activity and those who do not could be entirely driven by an activity left out of the analysis. E.g., a positive relationship between test scores and making music could be driven by children "buying" the necessary time for music by watching less television, even if there was no causal effect of

[1] These comprise of educational activities with parents, educational activities with other adults, general care with parents, general care with other adults. The remaining categories are sleeping, time at school/day care, social activities, and media consumption.

making music at all.

In this respect our approach is closest to Fiorini and Keane (2014) and Del Boca et al. (2012). Our methodology is similar to the former: Fiorini and Keane study simultaneously the effect of several mutually exclusive activities on cognitive and noncognitive skills in a sample of young children (about 4 to 7 years old). The authors choose the Longitudinal Study of Australian Children (LSAC), in particular because its sample size is larger than that of the PSID-CDS, its rhythm is biannual, and it comprises two cohorts instead of many. The LSAC only contains data on comparatively young children. In contrast, we are interested in the effect of activities throughout childhood, including adolescence, an age range which is covered by the PSID-CDS.[2] Furthermore, the direct availability of the HOME score variables in the CDS allows us to compare the relative importance of activities and this widely used proxy for parental investments. In turn, Del Boca et al. raise similar questions as we do, and employ the same data set. However, their focus lies on the relative contributions of maternal time investments and children's time investments. The definition of the investment measure, however, differs strongly from our approach. The authors lump activities that are thought to positively influence skill acquisition a priori into one additive activity category. We defer from building an index and study individual child activities' contributions instead. Conversely, we include one global parental time input measure of total time spent together with the respective child by at least one parent.

A first and preliminary look at the data supports the hypothesis that different extracurricular activities may influence skills differently. Figure 1 displays mean math skills for children who belong to the top and bottom 50 percent of each activity distribution at a different age. Children who pursue musical activities have substantially higher (about 0.5 standard deviations) math scores than children who don't. Slight positive differences (about 0.2 standard deviations) can be found for sports and reading. In turn, children who watch television a lot perform somewhat worse than those who don't. Sleeping seems to be slightly associated with lower test performance as well. Learning, time spent at school, and pursuing other activities does not seem to lead to different outcomes at first sight. Figure 2 repeats the exercise for reading scores, with similar results. Music is associated with strong differences in scores; reading, learning, and television exhibit slight differences. Here, sports shows no correlation with test scores.

The association between activities and cognitive skills displayed in figures 1 and 2 may reflect mere correlations. In fact, the challenge in answering the above questions lies in the potential endogeneity of several key explanatory variables. It is plausible that factors unobservable to the econometrician both affect test score achievement and time allocation. The Child Development Supplement (CDS) to the PSID is particularly well suited to address endogeneity concerns in two ways. First, it offers the advantage of giving quite detailed insight not only into the time allocation of children but also to other standard parental investment variables as well as family background variables, which should reduce omitted variable problems. Second, the three wave panel structure allows to control for unobserved heterogeneity by conditioning on past outcomes or time-invariant, unobserved confounders. In particular, we exploit standard panel data methods such as fixed effects and value-added models as do for instance Todd and Wolpin

[2] The groups of activities predefined in the LSAC reflect the age structure of the study's population in not including certain leisure activities typically pursued by older children which are of particular interest for the present study (e.g. musical activity or sport).

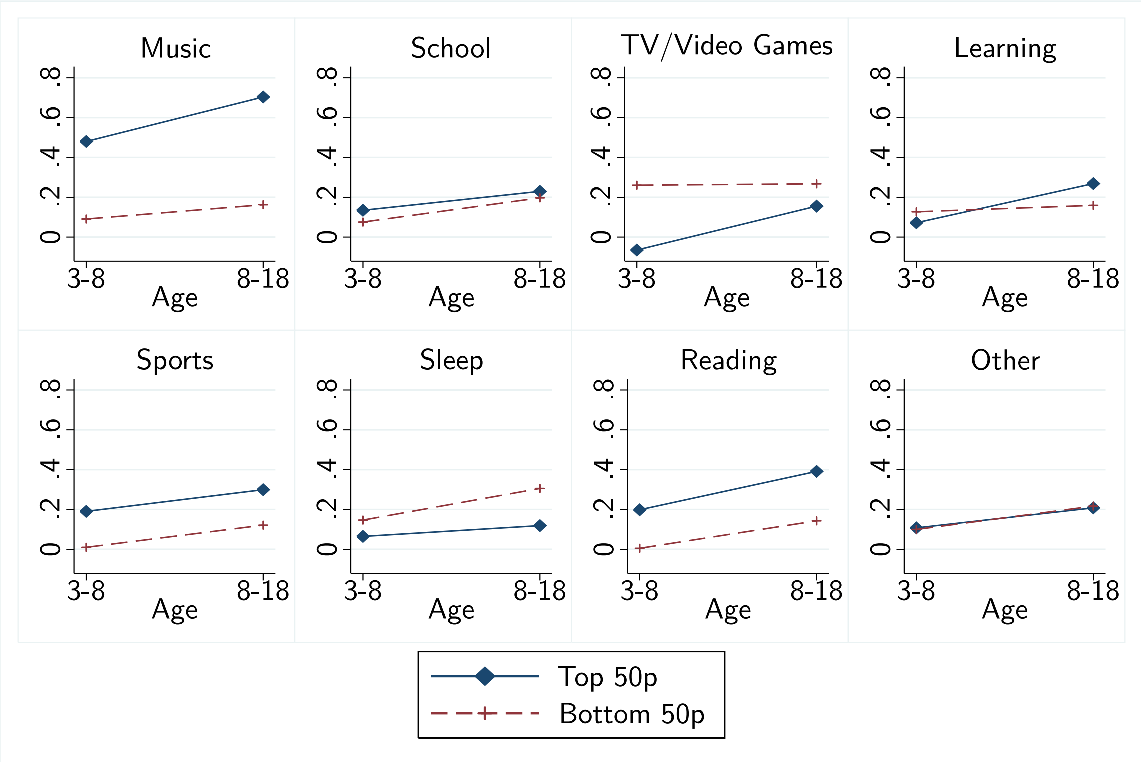

Figure 1: Math Test Scores by Activity and Age

(2007) and Fiorini and Keane (2014) when studying the effect of parental investments.

Our main findings suggest that

1. time inputs significantly contribute to explaining the variation in child math and reading test scores. The adjusted R^2 is higher with than without activities in all models explaining the variation in either math or reading test scores.

2. the relative ranking of the effect from raising the time spent on a given activity on the math test score is headed by music at the top, followed by learning and reading, exercising sports, watching television, attending school, and sleeping (in descending order). For the reading test score music ranks first again, followed by reading and learning. Time spent at school ranks fourth, followed by television or video games, sports and sleep. Not all differences in the effects are statistically significant. Musical activity, however, is more effective than any other remaining activity in raising reading tests and more effective than any other activity except learning in raising math tests, in a statistical sense.

3. when comparing the effect of child activities with that of parental investments on test scores in the PSID data, it turns out that (while both have statistically significant effects) including activities as a “proxy” for investments has explanatory power no less than parental investments. In case of math test scores, in all specifications the adjusted R^2 is lower if all activities are simultaneously excluded from the model than if parental investments are excluded.

The paper is organized as follows. Section 2 introduces the statistical model, section 3 describes the data, while section 4 presents the results summarized above. These results are

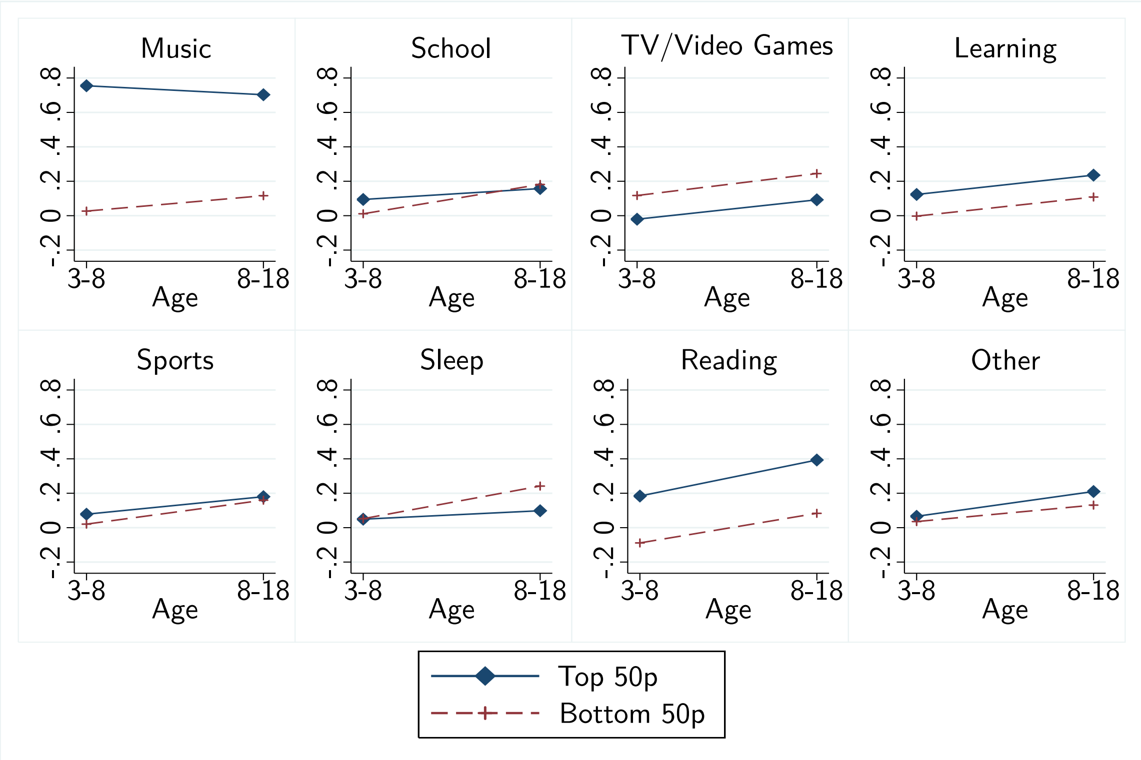

FIGURE 2: READING TEST SCORES BY ACTIVITY AND AGE

obtained under a standard conditional independence assumption. However, one could plausibly argue that our basic empirical model might be misspecified, leading us to conflate spurious regression results with causal effects of activities on test scores. Section 5.1 discusses the implications of the assumptions that underlie the results of section 4. In particular, we concede that the effect (and hence the ranking) of learning on test score achievement might be underestimated due to simultaneity bias. In section 5.2 we relax the assumption that unobserved factors that are relegated to the error term are manifest from very early age onwards. Section 6 concludes.

2 Statistical Model

We focus on a purely linear approximation to the production technology,

$$T_{it} = \alpha_0 + \boldsymbol{A}_{it}\boldsymbol{\alpha} + \boldsymbol{I}_{it}\boldsymbol{\beta} + \gamma B_{it} + \rho T_{i,t-1} + \boldsymbol{X}_{it}\boldsymbol{\delta} + c_{it} + \varepsilon_{it}, \qquad t = 1, 2, 3 \tag{1}$$

where i denotes a child and t stands for a survey wave. T and $\boldsymbol{A}$ represent a test score result and a $1 \times K$ vector of all but one mutually exclusive child activities, respectively; $\boldsymbol{I}$ denotes a $1 \times L$ vector of investments other than child i's time inputs; B is a proxy for noncognitive skills, and $\boldsymbol{X}$ represents a $1 \times M$ vector of observed control variables that do not have the interpretation of explicit investments into child skills. Unobserved heterogeneity is captured by the term c_{it}, and the idiosyncratic shocks are denoted ε.

Data limitations do not allow us to directly estimate the parameters of the population regression function, because only a small subset of observations can be observed in all three survey waves. We therefore estimate and compare the special cases with $\rho = 0$ and $c_{it} = 0$ for all i, t, respectively,

$$T_{it} = \alpha_0 + \boldsymbol{A}_{it}\boldsymbol{\alpha} + \boldsymbol{I}_{it}\boldsymbol{\beta} + \gamma B_{it} + \boldsymbol{X}_{it}\boldsymbol{\delta} + c_{it} + \varepsilon_{it} \qquad t = 1, 2, 3 \tag{1a}$$

$$T_{it} = \alpha_0 + \boldsymbol{A}_{it}\boldsymbol{\alpha} + \boldsymbol{I}_{it}\boldsymbol{\beta} + \gamma B_{it} + \rho T_{i,t-1} + \boldsymbol{X}_{it}\boldsymbol{\delta} + \varepsilon_{it} \qquad t = 2, 3 \tag{1b}$$

Equation (1a) assumes that "self-productivity" (captured by ρ) is absent, while equation (1b) abandons unobserved heterogeneity (captured by c_{it}).[3] Note that in all equations,

$$\boldsymbol{\alpha} = \begin{pmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_K\end{pmatrix}' = \begin{pmatrix}\tilde{\alpha}_1 - \tilde{\alpha}_{K+1} & \tilde{\alpha}_2 - \tilde{\alpha}_{K+1} & \dots & \tilde{\alpha}_K - \tilde{\alpha}_{K+1}\end{pmatrix}',$$

with $\tilde{\boldsymbol{\alpha}} = \begin{pmatrix}\tilde{\alpha}_1 & \tilde{\alpha}_2 & \dots & \tilde{\alpha}_K & \tilde{\alpha}_{K+1}\end{pmatrix}'$ the set of coefficients in a regression on all $K+1$ available, mutually exclusive activities. Since all $K+1$ activities sum up to 24 hours for every child in the population, $\tilde{\boldsymbol{\alpha}}$ is not identified. What is identified is $\boldsymbol{\alpha}$, where $\alpha_k = \tilde{\alpha_k} - \tilde{\alpha}_{K+1}$ measures the effect of substituting one hour away from the "residual" activity $K+1$ to activity k, when holding all remaining activities A_l, $l \in 1, 2, \dots, K, l \neq k$, fixed.

Activities and parental investments clearly represent choices in one way or another. It seems plausible that these choices are also endogenous in an econometric sense. In abstract terms, two well-known sources of endogeneity spring to mind. First, activities and other investments may be a function of unobserved variables that are correlated with test scores (omitted variables). In our model, $\boldsymbol{A}_{it}$ could be a function of c_{it}, which might capture, for instance, the general ability to learn and accumulate knowledge. Second, and closely related, the direction of causality might not only run from activities to test score achievements, but also from test score achievements to activities, again inducing correlation between, say, $\boldsymbol{A}_{it}$ and ε_{it}.

In either case, an instrumental variable strategy would be the desirable way to proceed. However, it would require at least $K + L$ instruments per age-group or wave, yet quasi-natural experiments on leisure-time allocation are hard to conceive. Policies affecting the total amount of time spent at school would be one possibility—but these would solely affect the time budget constraint in reducing or increasing the time left for leisure activities. Policies specifically directed at other child activities are hard to find. Of course, subsidies for certain leisure activities such as sports programs or musical lessons come to mind. However, in our data we cannot identify whether a child takes part in such programs.

Our options to address endogeneity concerns are limited. Regarding omitted variables, we study the classical fixed effects model and assume that $c_{it} = c_i$ in equation (1a). This assumption is valid if the unobserved, correlated variables affect test scores equally strongly in successive waves. However, abilities frequently modeled as a fixed effect might manifest themselves for the first time for a given child at different age than for another child. In a sensitivity analysis section 5.2 therefore reduces the sample by increasing the sample entry age and studies whether results change compared with the all-encompassing fixed-effect model.

Even when accounting for unobserved, time-invariant heterogeneity, the estimates will be

[3] Todd and Wolpin and Fiorini and Keane also study a variant with a lagged test score and lagged investment variables. While child development is a cumulative process, the value added specification is one convincing way to capture accumulated inputs. In principle past inputs could exhibit an additional influence on current skills besides the indirect one channeled through past skills. The empirical importance of this effect is negligible in our data, however.

consistent only if the usual conditional independence assumptions holds. This implies that we have to assume zero autocorrelation in the idiosyncratic error term. Reversely, correlation between the vector of choice variables $(\boldsymbol{A}, \boldsymbol{I})$ and time-variant components of the error term may induce bias.

3 Data

We construct a panel of children from the Panel Study of Income Dynamics' Child Development Supplement (CDS). In 1997, all PSID households with children younger than 13 were eligible for inclusion for the CDS modules. No more than two children per household were being interviewed. These children and their parents were re-interviewed in 2002 and 2007 unless the respective child had reached age 19.

The PSID-CDS provides a rich data source of various factors that may affect child development, such as the households' socioeconomic status, and parental assessment of child behavior. In addition to that, researchers conducted tests measuring math skills and reading capability. What makes the study particularly useful for our purpose is the inclusion of time-use diaries measuring the time allocation of respondent children.

Table 1 lists the variables included in the analysis. We measure math and reading test score achievement by the results in the Applied Problems test and the Letter-Word Identification test, taken from the Woodcock-Johnson Revised Tests of Achievement (Woodcock and Johnson, 1989), a widely used test of cognitive skills.

A detailed overview over the time-use categories employed is given in table 2, which lists the variables and their respective codes in the CDS data set. We have partitioned the set of all activities available in the CDS into nine activity categories. These categories consist of music and theater, learning/doing homework, pursuing individual or team sports, watching television or playing video games, attending school, and sleeping. The categories differ by the number of time use variables they comprise of. Different considerations have led us to end up with the partitioning presented. Sleeping and attending school, for instance, are single-category activities because of the large empirical role they play in the data. The definitions of learning and reading emerged naturally because either activity is directed at influencing at least one of the cognitive tests to be explained. Musical activity and sports are of first-order interest. For instance, while positive associations between music and noncognitive skills are well documented, Hille and Schupp (2015) cite Schellenberg (2011) who argues that playing music has a causally affects cognitive skills. Similarly, sports is widely seen as a positive force in child development in general, although its effect on cognitive skills is little studied.

Figure 3 shows how time allocation differs by age. The right panel contains activities that exhibit quite high mean values. The quantitatively most important activity is sleeping, shrinking from roughly 11 hours per day at the age of 3 to 9 hours at age 18. Attending school or preschool absorbs more than six hours per day between 6 and 14, after which mean school time declines slowly to an average of about five hours. Watching TV and/or playing video games is an important leisure activity from early childhood on; children aged 3 onwards watch about 2.5 hours per day. Last, in a sense, we have two "residual" activities at hand. The first, "Other", contains activities that are not suspect to affecting cognitive skills—time devoted to personal

Table 1: Descriptive Statistics

	Mean	Standard Deviation	Min	Max
Math score[1]	0.16	0.99	-2.83	2.43
Reading score[1]	0.13	0.97	-2.58	1.33
Noncognitive skills[1]	-0.03	1.00	-4.10	1.54
Music[2]	0.05	0.26	0	5.76
Learning[2]	0.57	0.82	0	6.62
Reading[2]	0.17	0.37	0	6.05
Sports[2]	0.62	0.93	0	11.74
TV/Video Games[2]	2.70	1.88	0	15.42
NA[2]	0.14	0.71	0	24.00
School[2]	4.01	1.91	0	8.14
Sleep[2]	9.56	1.30	0	18.08
Other[2]	6.18	2.27	0	20.71
Time w/ parents[2]	2.89	2.04	0	20.50
HOME factor score (3-6)[1]	0.01	0.28	-3.41	1.16
HOME factor score (6+)[1]	0.07	0.94	-3.68	2.01
Log ⊘ teacher sal. by state[3]	10.82	0.19	10.39	11.26
Pupils/teacher by state	16.56	2.68	11.78	23.68
Currently in public school	0.81	0.39	0	1
Log family income[3]	10.77	0.91	0.09	14.49
Mother's years of education	12.89	2.87	0	17
Household head in workforce	0.93	0.26	0	1
Household head employed	0.88	0.33	0	1
Household head self-employed	0.13	0.34	0	1
Mother's tot hrs worked prev yr	1,258.89	952.81	0	5,510
Child age (assmnt.)	11.61	3.98	3.00	19.00
Moved at least once since prior wave	0.28	0.45	0	1
# biol. sibl. in hh.	1.43	1.13	0	9
Father present	0.69	0.46	0	1
Birth weight (lbs)	6.98	1.36	1	15
Mother's age at birth $<$ 20	0.08	0.27	0	1
Mother's age at birth $<$ 30	0.62	0.49	0	1
Black	0.16	0.37	0	1
Hispanic	0.13	0.34	0	1
Female	0.50	0.50	0	1
Born first	0.40	0.49	0	1
Born second	0.35	0.48	0	1

Notes: [1] Normalized to zero mean and unity std. dev. for all observations of the variable; [2] All activities measured in hours; [3] In year-2000 $.

care, household chores, etc. The other residual activity is time not allocated in the time diary, "NA", and hence indicating partial nonresponse. One might be tempted to dismiss its relevance due to low mean values, but one of the most interesting activities, playing music, has an even lower age-contingent mean. Interestingly, on average time devoted to doing homework never exceeds the hour mark. The same holds for recorded sports activities.

Turning to parental investments, the variable "Time w/ parents" is one, continuously mea-

Table 2: Definitions of activity aggregates

Activity	Variables	Variable codes
Music	Music lessons, Playing an instrument	8870, 8610
Learn	Using the computer for homework, studying etc.; being tutored; homework; other education	5040, 5190, 5490, 5690
Read	Reading books; magazines; newspapers; unspecified; being read to	9390, 9410, 9590, 9420, 9430
Sport	All sorts of team sports, individual sports, outdoor activities etc.	8850, 8860, 8880, 8830, 8840, 8010, 8020, 8030, 8040, 8050, 8060, 8070, 8080, 8090, 8100, 8650, 8110, 8120, 8130, 8140, 8160, 8170, 8180, 8210, 8220, 8230, 8240, 8250, 8260, 8810
Tele	Television; playing computer or video games; other computer activities (e.g. surfing the net, chatting)	5020, 5030, 5050, 8790, 9190
NA	Activities of others reported; filling out time diary and the like; gap in diary	9840, 4810, 0000
School	Attending classes, school if full-time student	5090
Sleep	Night sleep, including in bed while awake	4590
Other	All other activities	*All remaining codes*

sured, proxy for parental time inputs and, similar to Del Boca et al. (2014), measures the hours during which at least one of the parent join their children in a primary activity. Several studies have emphasized the importance of parental time inputs.

Next, the Short Form of Home Observation Measurement of the Environment (henceforth HOME) score is a frequently used indicator of parental investments (Caldwell and Bradley, 1984). For instance, it is the central investment measure in Todd and Wolpin (2007), Cunha and Heckman (2008), and Cunha et al. (2010). It is often an additive index built from originally ordinally measured and then dichotomously recoded items which represent quite different dimensions of parental inputs. The latter two references extract a latent variable from these items. In similar fashion we distill loadings from factor analysis for each item and predict sum scores of the latent investment variable instead of equal weighting of the HOME measurement items. Note that the HOME score consists of different measurement items in the age groups three to six and six onwards, which is why we include both variables simultaneously, interacted with the respective age group dummy variable. In the basic regression we include predicted scores from a one-dimensional investment factor. In general, the share of variance attributed to measurement error (uniqueness) is high for many investment proxies. When extracting one latent investment variable only, those variables with the highest factor loads are those that reflect material investments into an environment conducive for child development. The different nature of the proxy variables suggests extracting a higher number of latent variables though. Table 3 assumes two underlying investment dimensions. In fact, a picture of two different dimensions emerges: The first factor manifests itself in rather material investments, whereas the second factor reflects parental time inputs.

We also include a measure of noncognitive skills. Cunha and Heckman (2008) and Cunha et al.

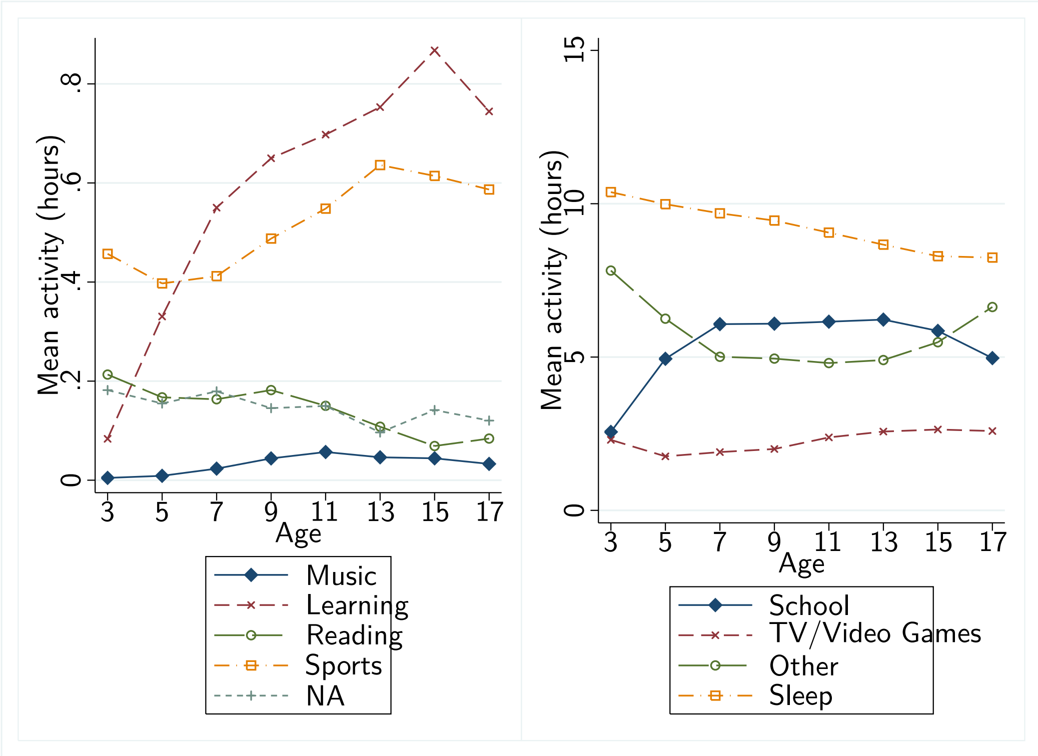

FIGURE 3: AVERAGE ACTIVITY BY AGE

TABLE 3: ROTATED FACTOR LOADINGS OF HOME MEASUREMENTS

	HOME Factor 1	HOME Factor 2	Uniqueness
Number books	0.16	0.36	0.85
Musical instrument	0.13	0.32	0.88
Child encouraged to pursue hobbies	0.03	0.21	0.95
Child enlisted for extrac act's	0.06	0.13	0.98
How often child taken to museum	0.05	0.61	0.62
How often child taken to theater	0.09	0.63	0.59
Parents discuss TV w child	0.02	0.17	0.97
Home (not) monotonous	0.63	0.15	0.58
Home (not) cluttered	0.76	0.02	0.41
Home clean	0.91	0.03	0.17
How often provision of toys/activities	0.10	0.17	0.96
Eigenvalues	1.95	1.07	

Note: Based on sample of children aged 6 or older.

(2010) show that noncognitive skills are an important factor in the cognitive skill development process. It seems plausible that noncognitive skills at the same time have a large influence on time allocation decisions by shaping preferences and providing (and limiting) resources for different activities. Like Cunha and Heckman (2008) and Cunha et al. (2010),[4] we use the Behavioral Problem Index (BPI), which originates from Peterson and Zill (1986). The BPI is a measure of behavioral problems of children at least 4 years old and is obtained via parental assessment.

[4] Fiorini and Keane (2014) use a battery of questions that are similar to the BPI.

In the CDS, behaviors are divided into two dimensions, externalizing or aggressive behavior on the one hand and internalizing, withdrawn or sad behavior on the other. The Externalizing question battery contains items such as difficulty concentrating/not paying attention for long, impulsiveness/acting without thinking, but also disobedience and aggressive behavior towards other children. Internalizing consists of items such as feeling worthless or not liked by others, among others. We use the total score in the CDS that is a sum score from a confirmatory factor analysis.

Regarding school quality, we follow Todd and Wolpin (2007) and add, first, information measured at state level, namely log average teacher salary and the average pupil-teacher ratio. Second, it would be desirable to measure school quality at the district or school level. Attending better schools will likely affect the time allocated to certain activities during school hours which we do not observe and that are lumped together in the catch-all category "time at school" in our data. Differences in within-school time allocation may in turn affect leisure time allocation which we can measure with high precision. Not accounting for activities at school could lead to bias in our estimates. For instance, a school in a district inhabitated by wealthy parents may provide extensive musical education. This may lead to more musical exercise during leisure hours. Since better schools may, via better education, enable children to achieve better test scores, increased musical activity and correlated increased test scores could be driven by the omitted variable school quality. In our data set, due to data limitations, we proxy school quality by a dummy variable indicating private and public schools.

In addition to that we employ a standard set of control variables that are thought to correlate with or directly capture other sources of differential child development. First, we control for maternal education measured in years of schooling. It is not a perfect, yet the bext proxy for parental cognitive skills available in the data. Parental cognitive skills are an important and well-documented factor in shaping children's cognitive skills. Adding fathers' education would have resulted in a substantial reduction of the sample size; however, correlation between both variables is very high. We further control for labor market outcomes of the household head (typically the father) by a dummy for participation in the workforce, a dummy for self-employment, and a dummy for the employment status of the household head (typically the father). A further performance indicator is the log of annual household income. Strictly speaking, income cannot be a direct input into the production technology. However, in general it is widely accepted proxy for the socioeconomic status of economic agents, and its inclusion may proxy unobserved confounding factors.[5] The mothers' participation in the labor market is captured by her total hours worked during the previous year.[6] Mothers' labor force participation on one hand affects a household's resources, but may also affect the time available for child care and joint activities, a mechanism studied in various contributions cited in the introduction.

We also control for certain initial conditions each child faced. Like Todd and Wolpin, we include dummies for the position in the birth order and birth weight, two factors shown to affect cognitive skills by Rosenzweig (1986). We include dummy variables for the mother's age

[5]Todd and Wolpin, in a more structural interpretation, call skill production function that include household income "hybrid" functions.

[6] In the 2002 CDS wave these variables do not refer the previous but to the current year due to the biannual rhythm of the PSID.

at birth. The list also contains dummy variables for the respective child's gender and dummy variables for afroamerican or hispanic racial background. Differences in test scores between white children and non-white children from very young age onwards are well-documented, see for instance Carneiro et al. (2003).

The family composition is described by a dummy for whether the father lives in the household and the number of biological siblings living in the household. Lastly we control for the child age by a second-degree polynomial in age measured (precise at the day level) plus a full set of age dummies measured in years. We also add wave or year dummies.

In general, response rates for the different questionnaire modules differ somewhat. Still, our estimates are consistent when possible selection is a function of the (strictly) exogenous variables and the unobserved heterogeneity. In turn, correlation between the idiosyncratic error term and the selection mechanism may lead to selection bias and require correction procedures. We test for this necessity along the lines of Nijman and Verbeek (1992) by adding a selection indicator for the respective following wave in waves 1 and 2. It turns out that the estimated coefficient of next period's missingness indicator is not significantly different from zero whether or not missingness is defined as having dropped out of the sample as such or as resulting from nonresponse in at least relevant questionnaire module despite eligibility, with $\hat{\omega}_{\text{Missing}_{t+1}} = -0.01$ [0.04] in the former case and -0.01 [0.04] in the latter in the case of math test scores, and 0.03 [0.04] and 0.03 [0.04], respectively, in the case of reading scores.

4 Basic results

We now analyze whether time allocation patterns make a difference in the production of test score achievements, which activities matter most, and their relative importance vis-à-vis other investments.

4.1 Do activities matter?

Tables 4 and 5 display the main results. The tables are structured as follows. In both tables, column one presents results for the pooled specification, and column four shows the results for the fixed effects model when a scalar HOME score is included, respectively. Column two displays coefficients for the value added model in which we condition on past test scores, again with one HOME score. Columns three and five present results for the value-added and fixed effects regressions, respectively, when two investment scores are included. They will be discussed in section 4.3. We present coefficients for our main regressors of interest only. We estimate asymptotic variances robust to both serial correlation and heteroskedasticity by means of sandwich estimators. Below that we report the R^2 and the adjusted R^2, $\tilde{R}^2$. In addition, we report F tests of joint insignificance of the activity and the HOME score variables, respectively. Below that we report the change in $\tilde{R}^2$ once all activities are excluded from the respective equation ($\boldsymbol{\alpha} = \mathbf{0}$) and once the HOME score is excluded ($\boldsymbol{\beta} = \mathbf{0}$) in order to assess changes in model fit.

Before we have a detailed look at the sign and magnitude of the coefficients, we study whether leisure activities as a whole affect skill acquisition—in other words, whether they "matter" or not. For this, we test the restriction $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ by means of a Wald test. As noted above,

Table 4: Estimates of math skill production function

	Pooled	Value added		Fixed Effects	
	(1)	(2)	(3)	(4)	(5)
Music	0.100**	0.068*	0.069*	0.091***	0.089***
	[0.043]	[0.040]	[0.041]	[0.028]	[0.029]
Learning	0.039**	0.024	0.024	0.042***	0.042***
	[0.018]	[0.016]	[0.016]	[0.014]	[0.014]
Reading	0.010	0.013	0.014	0.015	0.021
	[0.030]	[0.029]	[0.029]	[0.028]	[0.028]
Sports	0.024*	0.002	0.002	0.008	0.008
	[0.013]	[0.015]	[0.014]	[0.012]	[0.012]
TV/Video Games	-0.003	-0.009	-0.009	0.003	0.002
	[0.007]	[0.008]	[0.008]	[0.007]	[0.007]
NA	-0.042*	-0.049**	-0.050**	0.002	0.002
	[0.024]	[0.020]	[0.020]	[0.010]	[0.010]
School	0.002	-0.009	-0.008	-0.001	-0.001
	[0.008]	[0.009]	[0.009]	[0.006]	[0.006]
Sleep	-0.035***	-0.026**	-0.026**	-0.030***	-0.029***
	[0.011]	[0.011]	[0.012]	[0.010]	[0.010]
Time w/ parents	-0.006	-0.006	-0.006	-0.000	0.001
	[0.006]	[0.007]	[0.007]	[0.005]	[0.005]
HOME factor score (3-6)	0.020			0.043	
	[0.033]			[0.035]	
HOME factor 1 score (3-6)					0.040
					[0.036]
HOME factor 2 score (3-6)					0.037
					[0.040]
HOME factor score (6+)	0.043**	0.019		0.024	
	[0.017]	[0.017]		[0.018]	
HOME factor 1 score (6+)			0.016		0.028*
			[0.016]		[0.016]
HOME factor 2 score (6+)			-0.002		-0.009
			[0.014]		[0.015]
Math score (Lag 1)		0.652***	0.653***		
		[0.046]	[0.046]		
Noncognitive skills	0.060***	0.042***	0.043***	0.050***	0.050***
	[0.015]	[0.014]	[0.014]	[0.017]	[0.017]
Log family income	0.072***	0.039	0.041	0.017	0.018
	[0.024]	[0.026]	[0.026]	[0.020]	[0.020]
# Observations	2,036	1,176	1,176	2,036	2,036
# Children	946	946	946	946	946
R^2	0.85	0.70	0.70	0.95	0.95
$\tilde{R}^2$ (Adj. R^2)	0.85	0.68	0.68	0.94	0.89
F Test: $\boldsymbol{\alpha} = 0$	3.9	2.2	2.2	4.0	30.3
Prob $> F$	0.000	0.028	0.028	0.000	0.000
F Test: $\boldsymbol{\beta} = 0$	3.1	1.2	0.5	1.3	4.9
Prob $> F$	0.048	0.265	0.580	0.275	0.293
$\Delta\tilde{R}^2$ given $\boldsymbol{\alpha} = 0$	-0.0033	-0.0048	-0.0048	-0.0018	-0.0029
$\Delta\tilde{R}^2$ given $\boldsymbol{\beta} = 0$	-0.0009	-0.0003	0.0000	-0.0002	-0.0003

Note: Column 1 presents estimates from the contemporaneous specification without child fixed effects. Column 2 includes a lag of the dependent variable as a regressor. Estimates in column 4 are obtained after demeaning all variables. Columns 3 and 5 present results for two instead of one parental investment variable. Additionally, all models include the mother's education, the number of biological siblings in the household, dummies for whether the household head is in the workforce and employed, whether the household head is selfemployed, the mother's working hours, whether the father lives in the household, child age and age squared at assessment, a full set of age dummies, and wave dummies. The contemporaneous and value-added specifications also include birth weight, a dummies for mother's age at birth, dummies for black and hispanic origin, child sex, and dummies for the birth order. Clustered standard errors in parentheses. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < .01$.

Table 5: Estimates of reading skill production function

	Pooled	Value added		Fixed Effects	
	(1)	(2)	(3)	(4)	(5)
Music	0.020	0.018	0.017	0.066***	0.065***
	[0.035]	[0.021]	[0.021]	[0.019]	[0.018]
Learning	0.008	0.002	0.002	0.011	0.013
	[0.013]	[0.012]	[0.012]	[0.011]	[0.011]
Reading	0.036*	-0.013	-0.014	0.015	0.015
	[0.021]	[0.018]	[0.018]	[0.021]	[0.020]
Sports	-0.005	-0.011	-0.011	-0.003	-0.003
	[0.011]	[0.010]	[0.010]	[0.011]	[0.011]
TV/Video Games	-0.006	-0.011*	-0.011*	0.006	0.005
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
NA	-0.035**	-0.045***	-0.045***	-0.017**	-0.018**
	[0.016]	[0.016]	[0.016]	[0.008]	[0.008]
School	-0.002	-0.012*	-0.012*	0.009	0.008
	[0.006]	[0.007]	[0.007]	[0.006]	[0.006]
Sleep	-0.017*	-0.015	-0.015	-0.012	-0.012
	[0.009]	[0.011]	[0.011]	[0.011]	[0.010]
Time w/ parents	-0.003	-0.001	-0.001	0.011**	0.011**
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
HOME factor score (3-6)	-0.014			-0.049	
	[0.028]			[0.041]	
HOME factor 1 score (3-6)					-0.051
					[0.040]
HOME factor 2 score (3-6)					0.021
					[0.027]
HOME factor score (6+)	0.078***	0.051***		0.017	
	[0.018]	[0.013]		[0.018]	
HOME factor 1 score (6+)			0.046***		0.021
			[0.013]		[0.015]
HOME factor 2 score (6+)			0.022**		-0.007
			[0.011]		[0.013]
Reading score (Lag 1)		0.436***	0.436***		
		[0.035]	[0.035]		
Noncognitive skills	0.042***	0.019*	0.019*	0.010	0.010
	[0.013]	[0.011]	[0.011]	[0.015]	[0.015]
Log family income	0.020	0.022	0.021	-0.031	-0.032
	[0.018]	[0.017]	[0.018]	[0.022]	[0.021]
# Observations	2,041	1,180	1,180	2,041	2,041
# Children	948	948	948	948	948
R^2	0.88	0.65	0.65	0.96	0.96
$\tilde{R}^2$ (Adj. R^2)	0.87	0.64	0.64	0.95	0.91
F Test: $\boldsymbol{\alpha} = 0$	1.8	1.6	1.6	2.9	24.1
Prob $> F$	0.081	0.123	0.135	0.004	0.002
F Test: $\boldsymbol{\beta} = 0$	9.9	14.8	7.6	1.4	5.1
Prob $> F$	0.000	0.000	0.001	0.244	0.276
$\Delta\tilde{R}^2$ given $\boldsymbol{\alpha} = 0$	-0.0006	-0.0029	-0.0029	-0.0006	-0.0009
$\Delta\tilde{R}^2$ given $\boldsymbol{\beta} = 0$	-0.0035	-0.0077	-0.0078	-0.0002	-0.0004

Note: Column 1 presents estimates from the contemporaneous specification without child fixed effects. Column 2 includes a lag of the dependent variable as a regressor. Estimates in column 4 are obtained after demeaning all variables. Columns 3 and 5 present results for two instead of one parental investment variable. Additionally, all models include the mother's education, the number of biological siblings in the household, dummies for whether the household head is in the workforce and employed, whether the household head is selfemployed, the mother's working hours, whether the father lives in the household, child age and age squared at assessment, a full set of age dummies, and wave dummies. The contemporaneous and value-added specifications also include birth weight, a dummies for mother's age at birth, dummies for black and hispanic origin, child sex, and dummies for the birth order. Clustered standard errors in parentheses. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < .01$.

we can only identify coefficients relative to some left-out activity. We have chosen the activity category "Other". Note that the test statistic does not depend on the choice of the left-out activity.

As tables 4 and 5 indicate, the set of activity variables clearly matters for both math and reading test score achievement. The F test leads to rejection of the null of joint insignificance in almost all columns. Only in the case of the value-added reading production function, which is based on almost half as many observations compared with the pooled or the fixed effects specifications, do we fail to reject the null at conventional significance levels. Likewise, placing the restriction $\boldsymbol{\alpha} = \mathbf{0}$ unambiguously leads to a decrease in the adjusted R^2 in all columns of tables 4 and 5.

We conclude that activities do matter for both math and reading skills production. We reject the notion that activities taken together do not affect test scores in five out of six specifications.

4.2 Which activities do matter?

Tables 4 and 5 are partly instructive for studying the relative contribution of specific activities to skill production. Irrespective of the chosen "residual" activity, the ordering of activities along their conduciveness for skill acquisition becomes immediately apparent. However, the standard errors, in parentheses, only provide information about the difference $\alpha_k = \tilde{\alpha}_k - \tilde{\alpha}_{\text{Other}}$, $k \in \{\text{Music, Learning, ...}\}$, respectively. In order to learn more about the trade-offs when choosing leisure activities we perform a Wald test of equal coefficients for every possible combination of activities. Table 6 gives the results based on the fixed effects model specification in column four of tables 4 and 5, with robust standard errors below. The associated p values represent the probability of wrongly failing to reject the null hypothesis $H_0 : \alpha_{k,l} = 0$, where $\alpha_{k,l} = \tilde{\alpha}_k - \tilde{\alpha}_l$ is the relative coefficient for activities k, l, $k \neq l$. In column 1 all activities are ranked by the size of their coefficients.

Surprisingly, for both math and reading skills acquisition, playing a musical instrument seems to be the most effective activity. Substituting time away from playing music towards any other activity is almost always associated with lower skills, see column 2 in either panel. For example, reducing "Other" activities by one hour and investing this time in playing an instrument leads to an average increase of math test scores by 9.1 percent of a standard deviation and to an increase in reading test scores by about 6.6 percent of a standard deviation. Almost all coefficients relative to music are different from zero at conventional significance levels. The effect of music is only indistinguishable from the one of learning in case of math skills. Similarly, studying at home or doing one's homework is also comparably effective in raising both cognitive skills, ranking second in the case of math and third in the case of reading. This activity is significantly more effective in raising math test scores than every other lower-ranking activity except reading. Its effect is measured with less accuracy in case of the reading production function; here its effect is indistinguishable from those of the activities reading, school, TV, and sports; while it is more effective than sleeping. While reading ranks third and second, respectively, its relative coefficient compared with any other activity (except music) is indistinguishable from zero for both math and reading production. Turning to the bottom of the rank order, the least effective true activity in raising either skill is sleeping. (In the lower panel, only time slots not filled rank

Table 6: Relative activity coefficients in skill production functions

Substitute... ...for	Other	Music	Learning	Reading	Sports	TV/Video Games	NA	School
Music	0.091***							
	[0.028]							
Learning	0.042***	-0.049						
	[0.014]	[0.030]						
Reading	0.015	-0.076*	-0.027					
	[0.028]	[0.040]	[0.032]					
Sports	0.008	-0.083***	-0.034**	-0.007				
	[0.012]	[0.030]	[0.017]	[0.030]				
TV/Video Games	0.003	-0.088***	-0.039***	-0.012	-0.005			
	[0.007]	[0.029]	[0.014]	[0.029]	[0.013]			
NA	0.002	-0.090***	-0.040**	-0.014	-0.007	-0.002		
	[0.010]	[0.030]	[0.017]	[0.031]	[0.015]	[0.011]		
School	-0.001	-0.093***	-0.043***	-0.017	-0.010	-0.005	-0.003	
	[0.006]	[0.028]	[0.015]	[0.028]	[0.013]	[0.007]	[0.010]	
Sleep	-0.030***	-0.121***	-0.072***	-0.045	-0.038***	-0.033***	-0.031***	-0.028***
	[0.010]	[0.029]	[0.016]	[0.030]	[0.015]	[0.011]	[0.012]	[0.010]

(a) Math

Substitute... ...for	Other	Music	Reading	Learning	School	TV/Video Games	Sports	Sleep
Music	0.066***							
	[0.019]							
Reading	0.015	-0.051*						
	[0.021]	[0.028]						
Learning	0.011	-0.055***	-0.004					
	[0.011]	[0.021]	[0.023]					
School	0.009	-0.057***	-0.006	-0.002				
	[0.006]	[0.018]	[0.021]	[0.012]				
TV/Video Games	0.006	-0.060***	-0.009	-0.005	-0.003			
	[0.006]	[0.019]	[0.021]	[0.011]	[0.006]			
Sports	-0.003	-0.070***	-0.018	-0.014	-0.012	-0.009		
	[0.011]	[0.021]	[0.023]	[0.013]	[0.011]	[0.011]		
Sleep	-0.012	-0.078***	-0.027	-0.023*	-0.021**	-0.018*	-0.008	
	[0.011]	[0.021]	[0.021]	[0.013]	[0.010]	[0.011]	[0.013]	
NA	-0.017**	-0.084***	-0.033	-0.029**	-0.026***	-0.023***	-0.014	-0.006
	[0.008]	[0.020]	[0.022]	[0.013]	[0.009]	[0.008]	[0.012]	[0.010]

(b) Reading

Note: Based on the fixed effects estimates from column 4 of tables 4 and 5. Robust standard errors in parentheses. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < .01$.

even lower.) Substituting sleep by any other activity will, on average, lead to an increase in both math and reading skills, most of the time the effect is statistically significant. While this is surprising, sleep also ranks very low in most specifications in Fiorini and Keane (2014). Other leisure activities such as watching television or exercising sports assume a medium position in either ranking. Sports, for instance, is significantly less productive than music and learning, and more productive than sleeping for math test achievement, while its effect is not significantly different from that of reading, watching TV, and attending school. The picture is very similar for reading test scores. Note that hardly any activities in Fiorini and Keane (2014) have statistical significant relative effects on child development once child fixed effects are taken into account. In contrast, in the present environment with older children, the effects of some activities, in particular music, are measured more precisely and remain statistically significant also in the child fixed effect model specification.

The ranking of coefficients in the value-added specifications is very similar to the one in the fixed effects specifications, see table 7. Again, music ranks highest for math and reading tests in the value-added specification. Here, learning ranks second in both cases. In case of math tests, the effects of music and learning are significantly larger then those of time at school, watching television, and sleeping hours, while music is also more productive than the residual category of "other" activities. The effect of reading cannot be distinguished from that of any other real activity. Sports is more productive than sleeping, which, in turn, is only indistinguishable from reading, sports, school and TV. The predictive capacity of activities is generally much lower for reading test scores. Apart from time slots not assigned ("NA"), the only relative effects measured with some precision are those of TV and school relative to other activities, respectively.

The relatively large coefficient of musical activity in the fixed effects and value added specifications confirms the first impression based on bivariate plots in figures 1 and 2. Furthermore, our results corroborate the findings of Cabane et al. (2015) and Hille and Schupp (2015) in that, first of all, exercising music is a quite productive means of raising child skills, and that, second, it is relatively more effective than sports in raising cognitive skills.

4.3 The relative importance of child activities and parental investments

We examine the relative importance of activities and parental investments in two ways. First, in order to secure high comparability to Todd and Wolpin (2007), we treat the battery of HOME measurement items as proxies of one latent parental investment variable. We then assess the models' fit after excluding either set of variables of interest successively. However, the home score aggregates investments that differ substantially by their nature. Certain measurement items predominantly proxy material investments. Other items clearly represent parental time investments. As described above, extracting two factors produces predicted latent variables one of which loads higher on "material" investments, while the other variable loads higher on variables that measure parental practices. We will include both latent variables and again assess model fit.

In the lower part of tables 4 and 5 we do not only report the F statistic for the hypothesis $\boldsymbol{\alpha} = \mathbf{0}$, but also the corresponding statistic for the test that parental investments proxied by the HOME score do not make any difference, i.e. $\boldsymbol{\beta} = \mathbf{0}$. In the lines below we report the difference

Table 7: Relative activity coefficients in skill production functions

Substitute... ...for	Other	Music	Learning	Reading	Sports	School	TV/Video Games	Sleep
Music	0.068*							
	[0.040]							
Learning	0.024	-0.044						
	[0.016]	[0.042]						
Reading	0.013	-0.055	-0.011					
	[0.029]	[0.049]	[0.033]					
Sports	0.002	-0.066	-0.022	-0.011				
	[0.015]	[0.041]	[0.021]	[0.031]				
School	-0.009	-0.077*	-0.033*	-0.021	-0.011			
	[0.009]	[0.040]	[0.019]	[0.029]	[0.016]			
TV/Video Games	-0.009	-0.077*	-0.033**	-0.022	-0.011	-0.001		
	[0.008]	[0.040]	[0.016]	[0.029]	[0.015]	[0.010]		
Sleep	-0.026**	-0.094**	-0.050***	-0.038	-0.028*	-0.017	-0.016	
	[0.011]	[0.041]	[0.019]	[0.030]	[0.016]	[0.012]	[0.012]	
NA	-0.049**	-0.117***	-0.073***	-0.062*	-0.051**	-0.040**	-0.040**	-0.023
	[0.020]	[0.044]	[0.025]	[0.036]	[0.024]	[0.020]	[0.019]	[0.021]

(a) Math

Substitute... ...for	Other	Music	Learning	Sports	TV/Video Games	Reading	School	Sleep
Music	0.018							
	[0.021]							
Learning	0.002	-0.016						
	[0.012]	[0.023]						
Sports	-0.011	-0.029	-0.013					
	[0.010]	[0.022]	[0.014]					
TV/Video Games	-0.011*	-0.029	-0.013	0.000				
	[0.006]	[0.020]	[0.012]	[0.010]				
Reading	-0.013	-0.031	-0.015	-0.002	-0.002			
	[0.018]	[0.027]	[0.022]	[0.018]	[0.018]			
School	-0.012*	-0.030	-0.014	-0.001	-0.001	0.001		
	[0.007]	[0.020]	[0.013]	[0.010]	[0.007]	[0.018]		
Sleep	-0.015	-0.034	-0.017	-0.004	-0.005	-0.003	-0.003	
	[0.011]	[0.021]	[0.014]	[0.012]	[0.010]	[0.017]	[0.008]	
NA	-0.045***	-0.063***	-0.047**	-0.034**	-0.034**	-0.032	-0.033**	-0.030*
	[0.016]	[0.024]	[0.019]	[0.017]	[0.015]	[0.023]	[0.015]	[0.016]

(b) Reading

Note: Based on the value-added estimates in column 2 of tables 4 and 5. Robust standard errors in parentheses. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < .01$.

in the adjusted R^2, denoted by $\Delta\tilde{R}^2$, again after consecutively setting $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\beta} = \mathbf{0}$.

In the case of the math score production function, the picture is quite clear: parental investments into the home environment seem to play a smaller role than the activities children are pursuing. In all specifications the adjusted R^2 shrinks more after excluding activities than after excluding the HOME score. Likewise, we cannot reject the null hypothesis that the respective HOME score variables are in fact zero. In case of reading, results are mixed. In the pooled specification, the HOME score of children aged 6 or older is significantly larger than zero. All in all, the adjusted R^2 drops more when the scores for both age groups are excluded from the regression than when activities are. The HOME (age 6 and older) score's coefficient drops from 0.078 to 0.051 in the value-added specification, but it has larger predictive capacity than activities. When unobserved heterogeneity is accounted for, however, we cannot reject the null that both HOME score variables are zero. Here, activities explain a larger share of the variance in reading test scores than the HOME score does.

These considerations indicate that child activities are at least as important as the often-used HOME score which serves as a proxy for parental investments. Up to now, we have left the composition of this score unchanged. We have pointed out above, however, that it is possible to disentangle the score via factor analysis and divide it into two scores that could be labeled material inputs and parental time inputs. We now explicitly turn to columns 3 and 5 of both tables, which distinguish between the predicted sum scores "HOME factor 1" and "HOME factor 2".

In general, the more "material" latent parental investment variable explains the variation in test scores better than the time or parenting variable. In case of math test scores, the first latent investment variable is significantly different from zero in the fixed-effects specification for children aged 6 or older (with coefficient 0.028 [0.016]). The coefficients on the second investment factor are all close to zero, both for 3 to 6 year olds and for children older than 6. For reading test scores, both factors exhibit coefficients significantly different from zero in the value-added specifications, but the coefficient on the "material" factor is roughly twice the size of the one on rather time-related investments. In the fixed effects regression, all investment variables are indistinguishable from zero.

We conclude that while there is evidence that a child's pure time allocation is a better predictor of test scores than parental investments captured by the HOME score, among the variables that contribute to the latter those that reflect rather material investments have higher explanatory value than those that reflect parental time inputs or parenting practices.

5 Robustness

In the previous section, we have argued that activities do matter for skill acquisition, given the assumptions placed on the specifications studied. A more detailed look at the ranking of activities along their conduciveness for skill acquisition in the baseline child fixed effects regressions suggests a surprisingly strong positive effect of playing music on skill development, compared with the other activities considered. Sleep, unlike conventional wisdom would have it, ranks very low among all mutually exclusive child activities.

As a first step in assessing the robustness of these findings we have reestimated the models studied so far using a number of alternative controls or alternative variable definitions, based on the fixed effects model specification. In particular, we have analyzed age-standardized math and reading test scores instead of the respective raw scores. The overall picture remains stable: musical activity is by far the most effective activity in raising either test score analyzed, and learning ranks high, although it ranks third in the math production function instead of second, while reading ranks second in both functions. The relative importance of activities compared with HOME investments is similar, except that activities also explain a higher share of the reading score's variance in the value-added specification. In turn, the relative contribution of single activities is measured a bit less accurately overall. Next, we replaced the BPI by a sum score obtained from confirmatory factory analysis over a set of variables including the BPI and a positive behavior scale that comprises of variables that measure such traits as the ability to rely on oneself, curiousness, whether a child usually carries out work carefully, etc. Similarly, we have estimated the model with an ad-hoc additive index of the HOME measurements instead of a predicted score obtained from factor analysis. Furthermore, we estimated our model without noncognitive skills, and we included general health status of the child. Again, our results change little. We have also varied functional form assumptions and specified more flexible specifications within the linear-in-parameters paradigm, e.g. by introducing interaction terms and second-order polynomials of the key explanatory variables. However, these regressions have not led to additional insights, mostly due to a lack of precision in the estimates.

In addition to these variations we have analyzed in greater detail why sleep ranks very low. To begin with, the perception that more sleep is always better may be premature. One objection to the conventional wisdom may be that the connection between sleep and the ability to carry out cognitive processes is not necessarily monotone. Voderholzer et al. (2011), for example, show that sleep deprivation need not result in lower cognitive capacity. Less time devoted to sleep may be balanced by a higher sleep intensity. This calls into question ex ante hypotheses of a strictly positive relationship between sleep and test score achievement. Still, the clearly negative "substitution" effect apparent in the basic specifications between sleep and almost any other activity calls for further investigation. We have therefore modified the basic specification and replaced the absolute sleep duration per day by the time when the observed child got up and when he or she went to bed in order to learn whether the results are driven by the way the day is structured. In addition, we have included dummies for the weekday at which activities were recorded. We have also controlled for information about general tiredness of the child and time passed since the last consultation of a doctor. None of these modifications change our results.

Moreover, visual inspection of test scores plotted against reported sleep after purging either quantity of any correlation with all other covariates does not reveal any anomalies—the bulk of observations is centered around a negative regression line. As a consequence, we tentatively accept that increasing sleeping hours is not a particularly effective means in raising cognitive skills.

Still, in general, our estimated coefficients might be confounded not because of data deficiencies but on a conceptual level. As pointed out in section 2, our estimates could be inconsistent due to various sources of endogeneity, none of which would disappear as a result of the above alternative data employed. As outlined, unfortunately we are not able to find exogenous variation in activities and investments in order to apply an IV approach.

In the next subsections we discuss two sources of endogeneity. Section 5.1 discusses conditions under which reverse causality affects our estimates and argue that the partial correlation between, say, music and test scores indicates that music must causally affect tests. Section 5.2 discusses a relaxation of the assumption that unobserved factors that are relegated to the error term are manifest from very early age onwards.

5.1 Reverse causality

When test results elicit behavioral responses in child-parent pairs that lead to readjustments of time allocation patterns, our production technology parameters may suffer from simultaneity bias. On the one hand, one might call into question whether children react at all to the kind of tests elicited in the study. First of all, the time diary forms have been completed prior to the child interviews in most (about 75 percent) of the cases, so that in a strict sense (future) scores cannot have caused past activities in these cases. Second, it seems more plausible that children and parents view school exam results as valuable indicators of skills, and not a test that has been administered by an outside research institute. On the other hand, insisting on the timing pattern could still be misleading. Exam results may be the expression of a slow-moving process of skill evolution, and the CDS test may be a "good" indicator of the true skills of the past weeks, to which households may have reacted already, possibly after receiving exam results in prior weeks, so that ruling out reverse causality on these grounds seems premature.

Thus, in general, we cannot rule out simultaneity bias. Still, our results suggest that activities other than learning, in particular musical activity, are likely to influence test scores. Consider an extreme counter-position denying any causal influence of musical activity on math skills. In this scenario, children and parents know that music has no effect on math skills. Children study at home to raise their math skills, but do not practice music with the *intention* of improving math skills. A positive idiosyncratic shock on test scores, by signaling higher than expected math skill and thus reducing the necessity and pressure to learn hard, may reduce the time allocated towards learning. Due to the fixed time budget, other activities will necessarily be adjusted as a reaction to reduced learning hours. As a result the time spent practicing music may rise. This mechanism alone could induce positive correlation between music and test scores. Would the models studied in section 4 lead us to wrongly detect a causal effect of music on test scores based on this correlation? Not necessarily—a possible bias would reflect partial correlation between a regressor and the error term rather than unconditional correlation: The expectation of the estimator for the effect of music is $E[\hat{\alpha}_{\text{Music}}] = \alpha_{\text{Music}} + \pi_{\text{Music}}$,[7] where π_{Music} is the coefficient in the projection of the error term onto the column space spanned by the regression covariates, including time spent on learning. In the counter-position story, only learning hours will *directly* respond to shocks on expected scores. Once learning is controlled for, correlation between ε_t and other activities would still have to result in zero or very low partial correlation, leading to zero projection coefficients for all activities other than learning, if all other activities in turn adapt to changed learning hours (in an approximately linear fashion). This would imply

[7] Recall that coefficients not marked by a tilde represent effects relative to those of "other" activities, $E[\hat{\alpha}_j] = \tilde{\alpha}_j - \tilde{\alpha}_{\text{Other}} + \tilde{\pi}_j - \tilde{\pi}_{\text{Other}}$.

a (likely downward) bias in the coefficient on learning only and zero bias in the one on, say, musical activity in the regression of test scores on activities. However, despite controlling for learning hours, we do find substantial partial correlation between certain other activities, music in particular, and test scores. We conclude that musical activity has to affect skills causally to a certain extent. Of course this conclusion depends on the assumption that *if* music is irrelevant for math skills ($\tilde{\alpha}_{\text{Music}} = 0$), *then* raising math skills is not an important rationale for parents and children to practice music, in a way that leads to $\tilde{\pi}_{\text{Music}}$ close to zero and consequently to a low position in the ranking of relative coefficients. Under this assumption, if $\alpha_{\text{Music}} = 0$, then $E[\hat{\alpha}_{\text{Music}}] = \alpha_{\text{Music}} + \pi_{\text{Music}}$ is small. Remember that in the regression of section 4, $\hat{\alpha}_{\text{Music}}$ actually is significantly larger than all other coefficients. We find the assumption (if $\alpha_{\text{Music}} = 0$, then $\pi_{\text{Music}} = 0$) sufficiently plausible to conclude that the extreme counter-position denying any causal influence of musical activity on math skills is not particularly compelling.

If one accepts the first of our hypotheses from the introduction (activities other than learning, e.g. music, have a direct influence), this leaves open the answer to the second (how strong is this influence?). Even if we do believe that music "matters", the strength of the effect may still be misjudged. If $\alpha_{\text{Music}} \neq 0$, then we have little reason to speculate that $\pi_{\text{Music}} = 0$. We cannot distinguish between combinations of α_{Music} and π_{Music} summing up to the same estimated coefficient $\hat{\alpha}_{\text{Music}}$. Like previous works on the subject, e.g. Todd and Wolpin (2007) and Fiorini and Keane (2014), we necessarily leave open the question of how large possible simultaneity bias ultimately is.

5.2 Unobserved ability

In section 3 we have described the rich set of conditioning variables we employ in the analysis, thus reducing the likelihood of omitted variable bias. In particular, we hope that the inclusion of the BPI as a measure of noncognitive skills is a powerful proxy for motivation, perseverance and the like. We may still miss to control for fundamental characteristics of a given child, characteristics that are frequently treated as endowed, time-invariant abilities—after all, this is why we use fixed effects estimation in the first place. This produces unbiased estimates if the assumption is correct that the c_i's are indeed time-invariant within the population under study. Now suppose instead an innate ability or disposition which remains unobservable to the researcher throughout the sample period either automatically reveals itself to the parents and the child not at birth but at some later age in childhood reached at $a = \tau_i$, or is activated by certain random or coincidences at $a = \tau_i$. The distinction is irrelevant to the extent that both scenarios will lead to a behavioral response of parent-child pairs after τ_i. In both cases it would only make sense to treat $c_{it} = c_i$ in equation (1a) for $a \geq \tau_i$, but not before.

A low minimum age for inclusion in the sample could induce omitted variable bias. In particular this would be the case if, for a fraction of children, the late disclosure of the ability in question directly would raise mathematical aptness and increase the likelihood of persistently engaging in musical activity. Suppose further that for another fraction of children, a negative trait becomes manifest/known in the meantime that negatively affects math aptness and decreases the likelihood of enduring in musical activities. Even if there was no causal influence of music on math test scores, holding studying fixed, computing changes in math scores and musical activity would lead to positive correlation between the two quantities.

We can test whether our fixed effects estimates are robust by increasing the lowest age for inclusion in the sample: If we assume that latent abilities unfold their impact only from some random point in time onwards, and that the likelihood that they manifest themselves/are discovered is a monotonically increasing function in age, then bias from erroneously treating abilities as fixed from age 3 should decrease when the age that qualifies for sample inclusion increases. In contrast, if the estimates remain stable, we interpret this as evidence that the fixed effects estimates unlikely to be inconsistent.

We restrict the sample to children at least 8 years old. By then, most children are in third grade. By reducing the sample this way, we "lose" about forty percent of the observations included in the unrestricted estimation. Table 8 presents results for the fixed effects estimation based on the reduced sample. In general, compared to the results in tables 6 and 7, the relative coefficients are estimated with less precision. Still, musical activity ranks highest in both production functions. The relative coefficient in the case of math is 0.053 [0.031], which is smaller than the coefficient in the unrestricted sample. Learning remains the second-most effective activity in raising math test scores ($\hat{\alpha}_{\text{Learn}}$ =0.048 [0.016]). In case of reading skills, music's coefficient is now 0.036 [0.016]which is again smaller than the "unrestricted" coefficient. As in the unrestricted regression, reading and learning rank second and third. It is notable that the explanatory power of activities relative to that of the parental investment measure is not diminished compared with section 4. The null that all relative coefficients are zero is rejected ($p = .027$). Still, for both the set of activities as well as the HOME score the adjusted R^2 increases when either (set of) variable(s) is dropped from the regression. The increase is equally large, however, which hints at similarly large roles in explaining math test score achievement.

We interpret this as evidence that increases in musical activity indeed lead to higher test scores, even at an age at which a lot of personality traits should be quite stable.

6 Conclusion

Time is a central resource for acquiring skills, yet the role of child leisure activities throughout childhood has not been studied extensively in the literature on child development. We contribute to the emerging literature that analyzes the effects of certain child activities on cognitive and noncognitive skills by employing a large sample of American school children and incorporating after-school activities in a child skill production function framework. Our contribution complements previous studies that quantify the effect of parental investments on child skills. When neglecting child activities, estimates in these contributions will likely measure the joint effects of increases in home investments and the effects of adjustments in the allocation of leisure time. This need not pose a problem, depending on the question at hand. We believe, however, that time allocation is so central to skill accumulation that an investigation isolating both channels is justified.

By employing time use diaries of a large panel of American children, we can thoroughly incorporate the time budget constraint and hence estimate effects of different activities, holding all other activities but one "residual" activity fixed. On the one hand, this sets our contribution apart from studies that quantify the effects of a very limited number of activities on child

Table 8: Relative activity coefficients in skill production functions for reduced sample

Substitute... ...for	Other	Music	Learning	Reading	Sports	TV/Video Games	NA	School
Music	0.053*							
	[0.031]							
Learning	0.048***	-0.005						
	[0.016]	[0.034]						
Reading	0.010	-0.043	-0.038					
	[0.031]	[0.044]	[0.036]					
Sports	0.025*	-0.028	-0.023	0.015				
	[0.015]	[0.032]	[0.020]	[0.034]				
TV/Video Games	0.003	-0.050	-0.045***	-0.007	-0.022			
	[0.008]	[0.032]	[0.017]	[0.032]	[0.016]			
NA	-0.007	-0.060*	-0.055***	-0.017	-0.032*	-0.010		
	[0.012]	[0.032]	[0.019]	[0.034]	[0.018]	[0.012]		
School	0.005	-0.048	-0.043**	-0.005	-0.020	0.002	0.012	
	[0.009]	[0.031]	[0.019]	[0.031]	[0.016]	[0.009]	[0.011]	
Sleep	-0.009	-0.062*	-0.057***	-0.019	-0.034**	-0.013	-0.002	-0.014
	[0.012]	[0.032]	[0.019]	[0.032]	[0.017]	[0.013]	[0.014]	[0.012]

(a) Math

Substitute... ...for	Other	Music	Reading	Learning	School	TV/Video Games	Sports	Sleep
Music	0.036**							
	[0.016]							
Reading	0.004	-0.033						
	[0.025]	[0.031]						
Learning	0.003	-0.033*	-0.001					
	[0.011]	[0.019]	[0.028]					
School	0.001	-0.035**	-0.003	-0.002				
	[0.006]	[0.017]	[0.025]	[0.012]				
TV/Video Games	-0.003	-0.039**	-0.007	-0.006	-0.004			
	[0.006]	[0.018]	[0.026]	[0.012]	[0.007]			
Sports	-0.014	-0.050**	-0.017	-0.016	-0.015	-0.010		
	[0.011]	[0.020]	[0.027]	[0.013]	[0.011]	[0.011]		
Sleep	-0.006	-0.042**	-0.009	-0.008	-0.007	-0.002	0.008	
	[0.010]	[0.019]	[0.026]	[0.014]	[0.009]	[0.010]	[0.012]	
NA	-0.027***	-0.064***	-0.031	-0.030**	-0.028***	-0.024***	-0.014	-0.022**
	[0.008]	[0.018]	[0.027]	[0.013]	[0.009]	[0.008]	[0.013]	[0.010]

(b) Reading

Note: Based on the fixed effects estimates with minimum age for inclusion in the sample of 8 years. Robust standard errors in parentheses. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < .01$.

development from early childhood to adolescence. In these studies, time use is usually measured ordinally. On the other hand, we can differentiate between relative effects of different activities instead of analyzing one activity measure that aggregates different activities according to a rule of thumb.

We have accounted for potential endogeneity of activities by including a large set of covariates and by standard panel data methods such as value added and fixed effects specifications. Under standard assumptions on the error term, we have found that activities account for a substantial part of the variation in children's cognitive test scores relative to parental investments as measured by the HOME score.

Based on our assumptions, we find that, not surprisingly, learning has a positive impact when almost any other activity is "substituted" against. More surprisingly, musical activities seem to be quantitatively even more important than learning. Our findings suggest that one additional hour of music per day (at the cost of "other" activities) leads to a short-term rise in math test scores by about 9.1 percent of a standard deviation and reading test scores by about 6.6 percent. In another comparison of time allocation and home investments, one tenth of the reported average math test score gap between white and black children could be closed by allocating 28 minutes per day away from "other" activities towards musical activities, compared with a necessary increase of the HOME score by 1.76 standard deviations. As a caveat, learning might be the activity more likely to be prone to simultaneity bias than any other activity. As a consequence, we cannot rule out that its position in the ranking of activities is in fact first instead of second or third. In general, while time use data offer the advantage of convincingly capturing time allocation decisions, one has to interpret findings based on these data with caution. Even when employing standard panel data methods to alleviate endogeneity bias, certain sources of such bias cannot be ruled out entirely in lack of a large number of instrumental variables.

References

Almond, Douglas and Janet Currie, "Human Capital Development before Age Five," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4, Part B, Elsevier, 2011, pp. 1315–1486.

Bernal, Raquel and Michael P. Keane, "Quasi-structural estimation of a model of childcare choices and child cognitive ability production," *Journal of Econometrics*, 2010, *156*, 164–189.

Blau, Francine D. and Adam J. Grossberg, "Maternal Labor Supply and Children's Cognitive Development," *The Review of Economics and Statistics*, 1992, *74* (3), 474–481.

Cabane, Charlotte, Adrian Hille, and Michael Lechner, "Mozart or Pelé? The effects of teenagers' participation in music and sports," Technical Report, SOEPpapers on Multidisciplinary Panel Data Research 2015.

Caldwell, R. H. and B. M. Bradley, *Home Observation for Measurement of the Environment*, Univ. Arkansas, Center Child Development and Educ., 1984.

Carneiro, Pedro, James J Heckman, and Dimitriy V Masterov, "Labor market discrimination and racial differences in premarket factors," Technical Report, National Bureau of Economic Research 2003.

Carneiro, Pedro Manuel, Katrine Vellesen Løken, and Kjell G. Salvanes, "A Flying Start? Long Term Consequences of Maternal Time Investments in Children During Their First Year of Life," IZA Discussion Paper 5362, University College London 2010.

Cunha, Flavio and James J. Heckman, "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 2008, *43* (4), 738–782.

_ , _ , and Susanne M. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

_ , _ , Lance Lochner, and Dimitriy V. Masterov, "Interpreting The Evidence on Life Cycle Skill Formation," in Eric Hanushek and Finis Welch, eds., *Handbook of the Economics of Education*, Vol. 1, Amsterdam: Elsevier, 2006, chapter 12, pp. 697–812.

Del Boca, Daniela, Chiara Monfardini, and Cheti Nicoletti, "Self investments of adolescents and their cognitive development," Technical Report, Quaderni DSE Working Paper 2012.

_ , Christopher Flinn, and Matthew Wiswall, "Household choices and child development," *The Review of Economic Studies*, 2014, *81* (1), 137–185.

Felfe, Christina, Michael Lechner, and Andreas Steinmayr, "Sports and Child Development," Technical Report, Swiss Institute for Empirical Economic Research, University of St. Gallen December 2012.

Fiorini, Mario and Michael P. Keane, "How the Allocation of Children's Time Affects Cognitive and Noncognitive Development," *Journal of Labor Economics*, 2014, *32* (4), pp. 787–836.

Guryan, Jonathan, Erik Hurst, and Melissa Kearney, "Parental Education and Parental Time with Children," *The Journal of Economic Perspectives*, 2008, *22* (3), 23–46.

Heckman, James J., Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 2006, *24* (3), pp. 411–482.

Hille, Adrian and Jürgen Schupp, "How learning a musical instrument affects the development of skills," *Economics of Education Review*, 2015, *44*, 56 – 82.

Hsin, Amy and Christina Felfe, "When does time matter? Maternal employment, children's time with parents, and child development," *Demography*, 2014, *51* (5), 1867–1894.

Kalb, Guyonne and Jan C Van Ours, "Reading to young children: A head-start in life?," *Economics of Education Review*, 2014, *40*, 1–24.

Nijman, Theo and Marno Verbeek, "Nonresponse in Panel Data: The Impact on Estimates of a Life Cycle Consumption Function," *Journal of Applied Econometrics*, 1992, *7*, 243–257.

Peterson, James L. and Nicholas Zill, "Marital Disruption, Parent-Child Relationships, and Behavior Problems in Children," *Journal of Marriage and the Family*, 1986, *48* (2), 295–307.

Price, Joseph, "Parent-Child Quality Time Does Birth Order Matter?," *Journal of Human Resources*, 2008, *43* (1), 240–265.

Rosenzweig, Mark R, "Birth spacing and sibling inequality: Asymmetric information within the family," *International Economic Review*, 1986, pp. 55–76.

Schellenberg, E Glenn, "Examining the association between music lessons and intelligence," *British Journal of Psychology*, 2011, *102* (3), 283–302.

Todd, Petra E. and Kenneth I. Wolpin, "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps.," *Journal of Human Capital*, 2007, *1* (1), 91 – 136.

Voderholzer, Ulrich, Hannah Piosczyk, Johannes Holz, Nina Landmann, Bernd Feige, Barbara Loessl, Marta Kopasz, John Peter Doerr, Dieter Riemann, and Christoph Nissen, "Sleep restriction over several days does not affect long-term recall of declarative and procedural memories in adolescents," *Sleep Medicine*, 2011, *12*, 170–178.

Woodcock, R. W. and M. B. Johnson, *Tests of Achievement, Standard Battery (Form B)*, Riverside Publishing (Houghton Mifflin), 1989.