2875F

PART IIA PAPER 3 PROJECT

2. Factors influencing post-Kindergarten measured educational attainments in the United States, 2011 – 2014.

Using the Early Childhood Longitudinal Study (ECLS) database of the U.S. Center for Educational Statistics for the 2011 Kindergarten intake, together with any additional data and information you select, identify and evaluate the principal factors influencing individual differences in measured gains in any two of:

a) Reading

b) Mathematics

c) Science

Your analysis should take account of individual, family and school characteristics.

Word Count: 1862 + (7x13 half-line equations) + (1x25 full-line equations) = 1978

I. Abstract

Controlling for all other factors and using a fixed effects (FE) estimation, BMI, attention span, time spent on maths homework, music lessons and attending a public school all have a strong positive impact on the growth of maths scores for the 2010-2011 kindergarten cohort. Similarly, BMI, attention span, reading outside of school, attending a public school and being in a rural school compared to a town/city all had a significant positive impact on the growth of reading scores in the same period. For both maths and reading, teaching experience is only significant at the 10% level. However, due to the limitations highlighted in the paper, my estimators are likely to be biased due to omitted variables and therefore should be inferred with caution.

II. Background and Literature Review

There has been extensive research undertaken especially with the 1998-9 ECLS-K database to identify factors influencing kindergarten maths and reading scores. Denton and West (2002) find that poorest student score consistently below the national average in maths and reading as well as finding a gender divergence in both maths and reading. Penner and Paret (2008) also analyse the gender disparities with boys scoring higher in maths among kindergarteners. Byrnes and Wasik's (2009) opportunity-propensity framework is also used in my analysis.

III. Data Description

Data was collected from the ECLK4 database from the National Center for Education Statistics (NCES). A nationally representative sample of 18,174 kindergarteners starting from Fall 2010 was used.

	Table 1	
	Variable	Description
y_{it}	*MSCALE*	Maths IRT Scale Score.
	RSCALE	Reading IRT Scale Score.
c_{it}	*BMI*	$BMI = \frac{Weight\ (lbs)}{Height^2(in^2)} \times 703$
	BMISQUARED	The squared term of BMI. This is used to capture a potential quadratic (concave) relationship between BMI and maths/reading scores.
	X_CHSEX_R	A dummy variable = 1 if the child is male.
	X_BLACK_R	A dummy variable = 1 if the child is black.
	X_ASIAN_R	A dummy variable = 1 if the child is Asian.
	X_HAWPI_R	A dummy variable = 1 if the child is Hawaiian/Pacific Islander.
	X_AMINAN_R	A dummy variable = 1 if the child is American Indian/Alaskan native.
	X_MULTR_R	A dummy variable = 1 if the child is multi-race.
	RAPID	The number of days per week that the child exercises for a minimum of 20 minutes.
	ATTENTION	This is a categorical explanatory variable describing attention span: 0: UNABLE TO ATTEND, 1: DIFFICULTY ATTENDING, 2: ATTENTIVE, 3: VERY ATTENTIVE, 4: COMPLETE AND FULL ATTENTION
	TIMEMTLESSHALF	A dummy variable = 1 if the child spends between 0<x≤30 minutes on maths homework.
	TIMEMTMOREHALF	A dummy variable = 1 if the child spends between x>30 minutes on maths homework.
	OUTREADONCEORTWICE	A dummy variable = 1 if the child reads once or twice a week outside of school.
	OUTREADTHREETOSIX	A dummy variable = 1 if the child reads between 3 to 6 times a week outside of school.
	OUTREADEVERYDAY	A dummy variable = 1 if the child reads every day.
f_{it}	*NUMSIB*	The number of siblings that the child has.
	LOWERCLASSINCOME	A dummy variable = 1 if the family income is between $25,000 and $35,000.
	MIDDLECLASSINCOME	A dummy variable = 1 if the family income is between $35,000 and $100,000.
	UPPERCLASSINCOME	A dummy variable = 1 if the family income is greater than $100,000.
	MUSIC	A dummy variable = 1 if the child has music lessons.
	ABSENCES	This is a categorical explanatory variable describing the number of absences: 0: NO ABSENCES, 1: 1 TO 4 ABSENCES, 2: 5 TO 7 ABSENCES, 3: 8 TO 10 ABSENCES, 4: 11 TO 19 ABSENCES, 5: 20 OR MORE ABSENCES
	FEDUCHIGHSCHOOLPLUS	A dummy variable = 1 if the father's highest education is completed high school, vocational training or some college.
	FEDUCBAGRAD	A dummy variable = 1 if the father's highest education is a BA or graduate qualification.

	FEDUCMAORABOVE	A dummy variable = 1 if the father's highest education is MA or above.
	MEDUCHIGHSCHOOLPLUS	A dummy variable = 1 if the mother's highest education is completed high school, vocational training or some college.
	MEDUCBAGRAD	A dummy variable = 1 if the mother's highest education is a BA or graduate qualification.
	MEDUCMAORABOVE	A dummy variable = 1 if the mother's highest education is MA or above.
s_{it}	*ENROL*	Number of children currently in the class.
	TOWN	A dummy variable = 1 if the school is located in a town.
	SUBURB	A dummy variable = 1 if the school is located in a suburb.
	CITY	A dummy variable = 1 if the school is located in a city.
	PRIVATE	A dummy variable = 1 if the school is private.
	TEACHEXP	Number of years of teaching experience that the teacher has.

IV. Methodology

i. Theoretical model

Initially, I interpreted the measured gains in maths and reading scores by looking at the factors which influenced scores in the Spring 2013 period. The naïve OLS model (equation 1) indicates that maths/reading scores are a combination of individual child factors c_{it}, family characteristics f_{it} and school characteristics s_{it}. a_i denotes the individual unobserved heterogeneity which is fixed over time. u_{it} denotes the remaining idiosyncratic error.

$$(1)\ y_{i6} = \beta_0 + \beta_1 c_{i6} + \beta_2 f_{i6} + \beta_3 s_{i6} + a_i + u_{it}$$

The main issue with the naïve OLS model is the unobservable heterogeneity present which violates the assumption of zero conditional expectation of the total error term $E[a_i + u_{it}|c_{it,}f_{it,}s_{it}] \neq 0$ in the Gauss Markov theorem. In order to deal with the omitted variable bias caused by unobserved heterogeneity, I used a 3-period panel (Spring 2011, Spring 2012 and Spring 2013) and a FE method to measure gains in maths and reading scores.

Theoretically, a basic FE model works by subtracting 3 from 2 in order to remove the a_i.

$$(2)\ y_{it} = \beta_0 + \beta_1 c_{it} + \beta_2 f_{it} + \beta_3 s_{it} + a_i + u_{it}$$

$$(3)\ \bar{y}_i = \beta_0 + \beta_1 \bar{c}_i + \beta_2 \bar{f}_i + \beta_3 \bar{s}_i + a_i + \bar{u}_i$$

$$(4)\ (y_{it} - \bar{y}_i) = \beta_1 (c_{it} - \bar{c}_i) + \beta_2 (f_{it} - \bar{f}_i) + \beta_3 (s_{it} - \bar{s}_i) + (u_{it} - \bar{u}_i)$$

ii. Diagnostic checks

Table 2		
Testing For	Heteroskedasticity	Correct Model Specification
Diagnostic Test	Breusche-Pagan	Ramsay RESET
Hypothesis	$H_0: Homoskedasticity$ $H_1: Heteroskedasticity$	$H_0: Model\ correctly\ specified$ $H_1: Model\ incorrectly\ specified$
Naïve OLS MSCALE6	chi2(1) = 1280.15 Prob > chi2 = 0.0000	F(3, 24788) = 42.24 Prob > F = 0.0000
Naïve OLS RSCALE6	chi2(1) = 875.53 Prob > chi2 = 0.0000	F(3, 25720) = 40.52 Prob > F = 0.0000

As both Naïve OLS models were heteroskedastic, all models were adapted to use robust standard errors to ensure the standard errors are no longer biased. Heteroskedasticity could be present due to $Var(Maths\ Score \mid Income) = f(Income) \neq \sigma^2$. In this example, there may be a much larger spread of maths scores for individuals with high family incomes compared to the lower incomes because even those most high-income children have the opportunity to do well, not all have the propensity. Byrnes and Wasik (2009) use this opportunity propensity framework to analyse mathematical achievement.

In terms of the Ramsay RESET test, both naïve OLS models were incorrectly specified at 1% significance. However, this test doesn't state which variables are missing to correctly specify my model (Wooldridge 2015).

iii. Limitations and Biases

Data Limitations

In the panel data analysis, many variables of interest were not collected in the 3rd period (Fall 2011) and the 5th period (Fall 2012). This attrition of data was a result of subsampling 30% of those who participated in the base year in order to reduce costs. As a result, my panel data was restricted to the 2nd, 4th and 6th periods. The ECLS-K4 database was also subject to many missing data points. This may be an issue if this was a result of selection bias (e.g. those with the lowest incomes may not have answered questions on socioeconomic status).

Categorical variables

The explanatory variables ATTENTION and ABSENCES are categorical with integer scales between 0-4 and 0-5 respectively. Although this isn't a significant issue, the beta coefficients on the regression have a limited interpretation in the fact that:

1. They do not have an interpretation above the highest category e.g. an increase in ATTENTION from 6 to 7 has no true meaning.
2. They assume a fixed impact on scores when moving from one category to another e.g. an increase in ATTENTION from 0 to 1 has the same impact on results as an increase in ATTENTION from 2 to 3.

Omitted Variable bias

A major limitation with my method is that data on learning disabilities is suppressed in the ECLS-K4 database. This is a significant individual factor that would influence maths and reading scores as Morgan, Farkas and Wu (2009) find that children (in the ECLS-K 1998-9) with difficulties learning maths show the lowest growth trajectories in the subsequent years. I attempted to tackle this omitted variable bias through the use of premature birth as an instrumental variable (IV) for the ATTENTION variable. Although the relevance condition held $Cov(premature\ birth, attention\ span) \neq 0$ as premature birth could lead to an impact on brain development and therefore attention span as a child, I believed that the exogeneity condition $Cov(premature\ birth, \varepsilon_{it}) = 0$ didn't hold (ε_{it} is the entire error term in the regression). This is due to premature births highly likely to be correlated with learning disabilities.

To find completely unbiased estimators of the factors influencing maths and reading scores either data on the learning disabilities would have to be collected by contacting the NCES or an IV that satisfies both the relevance and exogeneity condition would have to be found.

Measurement error

Parental questionnaire proxies (attention span, time spent on maths, time spent on reading outside of school and absences) have been used. This could be an issue if parents tend to overestimate the time their children spend on maths for example. If the classical errors-in-variables (CEV) assumption holds, then regression estimators of these proxies will be biased and inconsistent (Wooldridge 2015).

V. Results

In the naïve OLS estimates, X_CHSEX_R and X_BLACK_R are both highly significant factors in determining both maths and reading scores in period 6 so I have added columns that separate the effects by gender and black subsamples in the FE regressions.

Individual factors

In table 4 column 1 and table 5 column 1, the coefficients on BMI and attention span are both positive and highly significant for the effect on maths and reading scores.

The increase in maths scores as a result of BMI could indicate the negative impact of being underweight on health and therefore academic performance. Although I expected a concave relationship between BMI and maths scores, the squared term of BMI was slightly positive and not significant which highlights that being overweight at a young age is unlikely to impact academic performance. Future panel data may be useful to find the long-term trend in middle school and onwards.

As expected, the time spent on maths homework (both more and less than 30 minutes) is positive and highly significant on maths scores. This highlights the need for 'practise makes perfect' with a subject like maths.

In table 5 column 1, reading outside of school for more than 3 days per week is positive and highly significant. Reading outside of school should enables children to pick up new vocabulary and increase the speed at which they can read enabling them to score higher in reading tests. The coefficient on the dummies for reading outside of school is significant for males but not for females which indicates the gender disparity where girls already have an aptitude for reading and therefore further reading outside school doesn't have a significant impact.

For both maths and reading, attention span is a positive and highly significant factor. This matches the results of Byrnes and Wasik (2009) that indicate that regardless of whether children are given the opportunity to score highly, they need the skills to be able to take advantage of the opportunity.

Family factors

The greatest gains in maths and reading scores (Table 4 Column 1, Table 5 Column 1) are seen by individuals with more siblings, those who take music lessons and those who have an income greater than $100,000.

The impact of music matches the findings of Southgate and Roscigno (2009) who use ECLS-K 1998-9 data to show that involvement in music related activities does statistically and positively impact academic performance potentially through influencing habits or creating a stronger mindset (propensity impact). However, Southgate and Roscigno (2009) also find that music involvement is a strong mediator of family background so there may not be a significant causal link between music activities and maths and reading scores.

The children who have a family income greater than $100,000 are more likely than those in lower income families to have opportunities present to score higher in maths and reading tests (Byrnes and Wasik 2009 Opportunity-Propensity framework). For example, the parents of these children have the finance to hire

private tutors and are more likely to finance extra-curricular activities such as music lessons which could contribute to a greater gain in scores.

The positive impact of siblings on maths and reading scores could be a result of the ability of older siblings to help the younger siblings with both maths and reading although this contradicted my initial belief that have more siblings would lead to less time for parental involvement.

School factors

Using table 4 column 1 and table 5 column 1, private schools have a significant negative impact on maths and reading scores. This may seem counter-intuitive as private school are likely to be funded better but having controlled all other factors, this result is in line with Carbonaro's (2006) finding that "public school gains in achievement are either the same as private schools' or slightly greater".

Class size has a significant positive (yet small) impact on both scores. This is a difficult result to analyse considering that the impact should be negative; a smaller class size should lead to greater one-to-one time per student. However, an explanation for this could be that class size becomes a much more important factor at a later stage in school when ideas become more conceptually difficult.

Conclusion and Policy Implications

Results from this paper show that there are a combination of individual, family and school characteristics which impact the gain in maths and reading scores over the Spring 2011, Spring 2012 and Spring 2013 period. BMI, time spent on maths homework (maths only), attention span and reading outside of school (reading only) are the significant individual factors. Families with income at the highest end of the distribution as well as those with more siblings and those who take music lessons are more likely to score higher and show greater gains in maths and reading tests. While at a school level, public schools and class size both positively impact maths and reading scores. From a policy perspective, greater parental involvement (e.g. in music lessons) could lead to greater growth trajectories in kindergartener's maths and reading scores. This parental involvement could improve early learning/behavioural habits to improve factors such as attention span which have been shown in the paper to significantly improve maths and reading scores. As there are potential family finance restrictions, subsidies could be provided to those with the lowest incomes to provide these opportunities to all children.

Table 3: Spring 2013 Naïve OLS Estimators (Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 – For all tables ahead)

VARIABLES	MSCALE6	RSCALE6
X6BMI	1.180***	0.335
	(0.406)	(0.375)
X6BMISQUARED	-0.0314***	-0.00962
	(0.0102)	(0.00942)
X_CHSEX_R	3.399***	-1.461***
	(0.289)	(0.269)
X_BLACK_R	-6.001***	-1.283***
	(0.515)	(0.459)
X_ASIAN_R	2.711***	0.384
	(0.521)	(0.462)
X_MULTR_R	2.107***	1.006
	(0.768)	(0.660)
P6RAPID	0.164**	-0.0445
	(0.0732)	(0.0663)
C6ATTENTION	6.173***	5.552***
	(0.175)	(0.154)
A6TIMEMTLESSHALF	2.270**	
	(1.057)	
A6TIMEMTMOREHALF	2.106	
	(1.517)	
P6OUTREADONCEORTWICE	6.203***	5.805***
	(1.413)	(1.217)
P6OUTREADTHREETOSIX	8.573***	8.852***
	(1.384)	(1.191)
P6OUTREADEVERYDAY	9.383***	11.71***
	(1.386)	(1.191)
X6NUMSIB	-0.434***	-0.974***
	(0.140)	(0.129)
X6LOWERCLASSINCOME	3.265***	3.259***
	(0.600)	(0.552)
X6MIDDLECLASSINCOME	4.923***	4.545***
	(0.478)	(0.437)

VARIABLES	MSCALE6	RSCALE6
X6UPPERCLASSINCOME	5.438***	4.832***
	(0.561)	(0.517)
P6MUSIC	2.207***	2.106***
	(0.371)	(0.342)
T6ABSENCES	-0.857***	-0.561***
	(0.139)	(0.132)
FEDUCHIGHSCHOOLPLUS	3.548***	3.778***
	(0.557)	(0.508)
FEDUCBAGRAD	6.079***	6.566***
	(0.647)	(0.591)
FEDUCMAORABOVE	7.055***	7.008***
	(0.717)	(0.667)
MEDUCHIGHSCHOOLPLUS	1.620***	1.431***
	(0.423)	(0.383)
MEDUCBAGRAD	3.124***	3.345***
	(0.546)	(0.489)
MEDUCMAORABOVE	5.041***	4.985***
	(0.612)	(0.576)
A6ENROL	0.136***	0.0540
	(0.0373)	(0.0335)
	(0.606)	(0.528)
X6SUBURB	-0.660*	-0.0341
	(0.376)	(0.347)
X6PRIVATE	0.0245	1.316***
	(0.427)	(0.404)
A6TEACHEXP	0.0254*	0.0362***
	(0.0145)	(0.0136)
Constant	39.53***	72.02***
	(4.506)	(3.935)
Observations	8,275	8,585
R-squared	0.371	0.368

Table 4: Maths FE Estimators – Spring 2011, Spring 2012, Spring 2013 Panels

VARIABLES	MSCALE (ALL)	MSCALE (MALE)	MSCALE (FEMALE)	MSCALE (BLACK)	MSCALE (NOT BLACK)
BMI	4.623***	5.284***	3.819**	0.264	5.180***
	(1.202)	(1.644)	(1.733)	(3.067)	(1.238)
BMISQUARED	0.0180	0.000264	0.0396	0.107	0.00833
	(0.0293)	(0.0406)	(0.0417)	(0.0698)	(0.0308)
RAPID	0.0759	0.0270	0.133	-0.0950	0.0865
	(0.0926)	(0.132)	(0.129)	(0.229)	(0.101)
ATTENTION	1.877***	2.088***	1.667***	1.778***	1.886***
	(0.214)	(0.305)	(0.300)	(0.599)	(0.230)
TIMEMTLESSHALF	8.412***	8.869***	7.733***	3.770	8.832***
	(0.911)	(1.286)	(1.285)	(3.260)	(0.952)
TIMEMTMOREHALF	8.933***	8.283***	9.260***	8.522*	8.579***
	(1.714)	(2.290)	(2.584)	(4.689)	(1.858)
OUTREADONCEORTWICE	0.212	1.402	-2.073	1.914	-0.422
	(1.137)	(1.437)	(1.809)	(2.771)	(1.233)
OUTREADTHREETOSIX	1.692	2.938**	-0.676	1.577	1.312
	(1.145)	(1.445)	(1.815)	(2.898)	(1.232)
OUTREADEVERYDAY	1.509	2.795*	-0.762	2.320	0.932
	(1.160)	(1.474)	(1.836)	(2.962)	(1.249)
NUMSIB	4.626***	3.449***	5.699***	2.471	5.164***
	(0.640)	(0.970)	(0.839)	(1.713)	(0.694)
LOWERCLASS	0.743	0.373	1.200	-1.384	1.177
	(0.832)	(1.218)	(1.133)	(2.064)	(0.903)
MIDDLECLASS	0.601	0.241	1.255	-1.029	0.891
	(0.946)	(1.378)	(1.310)	(2.293)	(1.038)
UPPERCLASS	2.314*	1.069	3.682**	-2.025	2.708**
	(1.196)	(1.706)	(1.690)	(3.980)	(1.276)
MUSIC	4.143***	3.178***	5.005***	4.529***	4.084***
	(0.569)	(0.838)	(0.776)	(1.537)	(0.613)
ABSENCES	-0.0330	-0.155	0.0894	0.404	-0.105
	(0.185)	(0.256)	(0.268)	(0.555)	(0.194)
ENROL	0.618***	0.616***	0.631***	0.616***	0.623***
	(0.0595)	(0.0830)	(0.0859)	(0.155)	(0.0646)
TOWN	-6.997***	-6.455**	-7.522***	-5.953	-6.711***
	(1.891)	(2.926)	(2.355)	(5.943)	(1.986)
SUBURB	-1.550	-4.377**	1.471	-2.036	-1.439
	(1.543)	(2.170)	(2.137)	(3.301)	(1.745)
CITY	-3.106	-6.816***	1.030	-7.692*	-1.458
	(2.103)	(2.554)	(3.351)	(4.261)	(2.295)
PRIVATE	-8.425***	-10.32***	-6.032	-20.49***	-6.718**
	(2.751)	(3.796)	(3.932)	(4.522)	(2.966)
TEACHEXP	0.0276*	0.0601**	-0.00445	0.0784	0.0257
	(0.0166)	(0.0238)	(0.0231)	(0.0507)	(0.0176)
Constant	-38.84***	-41.29**	-34.44*	9.296	-45.38***
	(12.03)	(16.16)	(17.66)	(32.45)	(12.17)
Observations	19,190	9,720	9,470	2,575	16,512
R-squared	0.214	0.212	0.222	0.217	0.219
Number of CHILDID	11,671	5,907	5,764	1,681	9,916

Table 5: Reading FE Estimators – Spring 2011, Spring 2012, Spring 2013 Panels					
VARIABLES	RSCALE (ALL)	RSCALE (MALE)	RSCALE (FEMALE)	RSCALE (BLACK)	RSCALE (NOT BLACK)
BMI	4.304***	4.737***	3.616**	1.029	4.719***
	(1.086)	(1.482)	(1.571)	(2.856)	(1.123)
BMISQUARED	0.0131	-0.000280	0.0334	0.0794	0.00598
	(0.0264)	(0.0364)	(0.0378)	(0.0656)	(0.0278)
RAPID	-0.0230	-0.0417	-0.00479	0.0422	-0.0494
	(0.0846)	(0.116)	(0.122)	(0.223)	(0.0914)
ATTENTION	1.958***	2.135***	1.788***	1.852***	1.984***
	(0.193)	(0.265)	(0.282)	(0.563)	(0.206)
OUTREADONCEORTWICE	1.585	2.577**	-0.216	2.957	1.226
	(1.088)	(1.274)	(2.055)	(2.481)	(1.196)
OUTREADTHREETOSIX	3.482***	4.163***	2.028	3.508	3.271***
	(1.114)	(1.311)	(2.090)	(2.588)	(1.221)
OUTREADEVERYDAY	3.875***	4.990***	2.097	5.316**	3.434***
	(1.128)	(1.340)	(2.093)	(2.636)	(1.236)
NUMSIB	3.924***	2.551***	5.398***	0.920	4.817***
	(0.675)	(0.912)	(0.828)	(1.236)	(0.652)
LOWERCLASS	0.569	0.412	0.743	0.103	0.666
	(0.730)	(1.034)	(1.031)	(1.937)	(0.790)
MIDDLECLASS	1.261	1.484	1.199	0.488	1.344
	(0.845)	(1.176)	(1.206)	(2.162)	(0.915)
UPPERCLASS	2.546**	1.933	3.165**	0.287	2.700**
	(1.080)	(1.470)	(1.582)	(3.816)	(1.147)
MUSIC	3.411***	2.800***	3.959***	3.881***	3.340***
	(0.526)	(0.790)	(0.705)	(1.453)	(0.566)
ABSENCES	-0.154	-0.315	-0.0144	0.0534	-0.193
	(0.169)	(0.234)	(0.245)	(0.538)	(0.176)
ENROL	0.559***	0.555***	0.572***	0.542***	0.568***
	(0.0553)	(0.0768)	(0.0798)	(0.156)	(0.0591)
TOWN	-5.949***	-5.875***	-5.939**	-5.494	-5.686***
	(1.635)	(2.165)	(2.457)	(4.573)	(1.731)
SUBURB	-2.787**	-3.611*	-1.816	-1.191	-3.437**
	(1.403)	(2.009)	(1.957)	(2.321)	(1.644)
CITY	-6.288***	-9.993***	-2.289	-11.91***	-4.254*
	(2.029)	(2.425)	(3.336)	(3.389)	(2.315)
PRIVATE	-11.04***	-12.45***	-9.117**	-22.34***	-9.068***
	(2.755)	(4.094)	(3.555)	(4.926)	(2.874)
TEACHEXP	0.0262*	0.0440**	0.00934	0.130***	0.0154
	(0.0150)	(0.0214)	(0.0211)	(0.0437)	(0.0160)
Constant	-3.206	-5.321	1.399	32.41	-8.642
	(10.91)	(14.64)	(16.12)	(29.76)	(11.13)
Observations	19,747	10,015	9,732	2,645	16,997
R-squared	0.200	0.199	0.205	0.215	0.203
Number of CHILDID	11,813	5,979	5,834	1,702	10,036

Bibliography

ECLS-K Sample Design, Weights, Variance, and Missing Data: https://nces.ed.gov/training/datauser/ECLS-K_04/assets/ECLK_04_slides.pdf

Morgan, P.L., Farkas, G. and Wu, Q., 2009. Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, *42*(4), pp.306-321.

User's Manual for the ECLS-K:2011 Kindergarten–Second Grade
Data File and Electronic Codebook, Public Version https://nces.ed.gov/pubs2017/2017285.pdf, pp(3-9)/54

Byrnes, J.P. and Wasik, B.A., 2009. Factors predictive of mathematics achievement in kindergarten, first and third grades: An opportunity–propensity analysis. *Contemporary Educational Psychology*, *34*(2), pp.167-183.

Southgate, D.E. and Roscigno, V.J., 2009. The impact of music on childhood and adolescent achievement. *Social science quarterly*, *90*(1), pp.4-21.

Wooldridge, J.M., 2015. *Introductory econometrics: A modern approach*. Nelson Education.

Carbonaro, W., 2006. Public-private differences in achievement among kindergarten students: Differences in learning opportunities and student outcomes. *American Journal of Education*, *113*(1), pp.31-65.

Denton, K. and West, J., 2002. Children's reading and mathematics achievement in kindergarten and first grade.

Penner, A.M. and Paret, M., 2008. Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, *37*(1), pp.239-253.