Exploring the Relationship between Summer Tutoring and Reading and Math Scores:
An Analysis using the ECLS-K Dataset

MPP Professional Paper

In Partial Fulfillment of the Master of Public Policy Degree Requirements
The Hubert H. Humphrey School of Public Affairs
The University of Minnesota

Victoria Perkins

May 10th, 2016

Signature below of Paper Supervisor certifies successful completion of oral presentation and completion of final written version:

______________________________	____________________	___________________
Judy Temple, Professor, Paper Supervisor	Date, oral presentation	Date, paper completion

____
Mallory Warner-Richter, Senior Research Analyst at Child Trends, Committee Member

____
Date

Introduction:

The summer months between school years are eagerly awaited and thoroughly enjoyed by almost all students. They mean freedom from teachers' expectations, early mornings, and late evenings of homework. While students may enjoy the break afforded by the summer, these crucial months play an important role in academic achievement. Research has shown that many students experience a "summer slide" over the break – essentially losing some of the learning they acquired over the school year.[1] This means that when students return to school in the fall, there is a certain amount of "re-teaching" that needs to happen to get them back up to speed. This time spent on re-teaching skills takes away from time that could be spent on new learning experiences and building new skills. So, loss of learning over the summer months represents a real instructional inefficiency.

This problem is complicated, however, by the fact that not all students lose learning at the same rate. In fact, research has shown that less-advantaged students lose learning over the summer at a higher rate than their more advantaged peers, which contributes to the already existent academic achievement gap.[2] Summer school and other summer activities are an important part of combatting this summer learning loss and narrowing this gap. One particularly popular out-of-school time intervention is tutoring, which many students receive during the school year and some continue to receive over the summer months.

The effectiveness of tutoring interventions has been well documented, but it is still unclear what effect there is of attendance in tutoring.[3] [4] Practitioners in the out-of-school time field acknowledge that a combination of program quality, access to resources, and youth participation all are important factors for success. Much work and research has been devoted to program quality and access to resources, but less is known about the effect of program attendance on student outcomes. Of course, it stands to reason that any attendance is better than none, and that attending more tutoring sessions is better than attending some. However, previous research on the "dose effect" of out-of-school time program attendance has produced mixed results.[5]

Additionally, the relationship between attendance in tutoring over the summer months and academic achievement has yet to be explored. As such, this paper will attempt to explore this relationship for tutoring during the summer between kindergarten and first grade. The relevant research questions considered in this paper are:

1. Is there a difference in reading and math scores between tutored and non tutored students? Hypothesis: Tutored students will, on average, have higher math and reading scores than non-tutored students because summer tutoring mitigates the effects of summer slide.
2. Is there a difference in the impact of dosage levels on reading and math scores? Hypothesis: More tutoring will result, on average, in higher reading and math scores because the positive effects of tutoring grow with exposure.
3. To what extent does summer tutoring effect mastery of specific reading and math proficiency levels? Hypothesis: Summer tutoring will result in increases for all three foundational reading and math proficiency levels, on average, with the highest impact on the first two proficiency levels.
4. To what extent does dosage of summer tutoring effect mastery of specific reading and math proficiency levels? Hypothesis: On average, there will be little to no effect on proficiency level mastery of the lowest level of dosage. As dosage increases, there will be an increasingly positive impact on proficiency levels. The highest level of dosage will produce large positive impacts for all three foundational reading and math proficiency levels.

5. Are there differences in proficiency level mastery between subgroups of students receiving the same level of dosage? Hypothesis: On average, differences in proficiency level mastery between black and white students and high-income and low-income students will be larger at the lower levels of dosage, with differences declining as dosage increases.

Literature Review:

Academic achievement gap – The academic achievement gap between white students and minority students (particularly black students) is one of the most important educational equity problems of our time.[6] Previous research about how children and families spend their time over the summer has primarily focused on the differences between socio-economic classes. This research has found large differences in time spent over the summer months between low-SES and high-SES children and families.[7] Those who are concerned about the effects of summer learning loss also tend to tie it to the broader racial academic achievement gap as a means of garnering more attention for the problem. While it's true that the academic achievement gap is most wide between white and black students (even among income and SES groups), we cannot ignore the effect of summer learning loss on widening the gap. As the American Psychological Association notes, "SES and race and ethnicity are intimately intertwined".[8] Further, the APA states that African American children are three times as likely to live in poverty than white children.[9] This clear link between SES and race paired with the link between access to summer learning and SES points to the importance of addressing summer learning loss if we want to close the achievement gap.

Summer learning loss – It has been well documented that students have a tendency to lose some of their learning gains over the summer break. In their article *Summer Learning Loss: The Evidence and a Possible Solution*, Trevor Kerry and Brent Davies explore the empirical evidence supporting summer learning loss. Most notably, they cite a study by Cooper et al, which unearthed some important findings about which types of learning are most likely to be lost over the summer and the typical size of the loss. According to Kerry and Davies, one of the key findings of this study was that, even in the best circumstances, students will tend to exhibit "little or no academic growth over the summer".[10] This in and of itself is not particularly worrisome, however, Cooper et al also found that in the worst of cases, students lost one to three months of learning.[11] So, not only were students not progressing academically over the summer, they were regressing in their learning by up to three months. Additionally, Cooper et al found that the loss was greatest in mathematics. This is supported by the fact that "factual and procedural knowledge is more prone to decay than conceptual understanding," meaning that learning that requires a high level of procedural knowledge and fact recollection is most likely to be effected by the long summer vacation.[12] Most damning, though, is the evidence suggesting that summer learning loss disproportionally effects low income children, thus widening the already wide academic gap that exists between less advantaged students and their more advantaged peers.

Disparities – Cooper et al discovered that summer learning loss disproportionally affects students of lower socioeconomic status, but how and why does this happen? One common explanation for the summer learning gap is the "faucet theory", which states that high-SES households are more able to provide activities during the summer that supplement school-day learning, thus "turning on the tap".[13] On the other hand, low-SES households are less able to supplement the turned-off tap of learning and thus cannot support the academic growth of their children over the summer.[14] Thus, gaps which already existed are wider and continue to grow as students experience summer break after summer break during their school careers. Some may not agree with the classic view of students as "empty vessels" waiting to receive knowledge from their teacher, and argue that there are other enriching activities that children can partake in over the summer that may have no academic content yet can support the positive growth and development of children. While this is certainly true, research suggests that even non-academic, enrichment summer experiences vary across socioeconomic status. Most notably, low-SES children spent, on average, two more hours a day watching TV than their high-SES peers.[15] So, not only are low-SES students disproportionately affected by summer learning loss because they have less access to learning supports over the summer, but also they seem to be spending less of their time on enrichment activities.

Objection: summer learning loss is not a problem – It is important to note that while many educators and parents believe in the existence of summer learning loss, some have argued that significant learning loss over the summer does not actually occur to the extent that many people believe it does. For example, in a 1972 study of three third-grade classes at the Los Arboles Elementary School, Woodrow Mousley found no significant differences in mean reading test scores from before and after summer vacation.[16] However, there are a number of limitations to this study which make its findings spurious. For one, there was no control group in this scenario, so there was no way to determine whether any gain or non-loss in reading ability was due to natural maturation. Additionally, the researchers did not account for activities over the summer, so there is no way to know whether or not these children were receiving supplemental reading support over the summer or maybe were simply engaged in reading for fun over the summer. Research which takes into account the different ways that children spend time over the summer and the academic supports they receive can provide a clearer picture of how learning loss may be mitigated by such activities and services.

Dosage – So, we know that having access to services, programs, and activities over the summer can potentially lessen the learning loss that takes place over the summer months. However, there are several key components of programs and activities that determine how successful they will be in making a difference for academic outcomes. Quality is one of these factors: quality of adult interactions with children as well as the quality of content and materials can have a large impact on the success of a program. Attendance, participation, and engagement are also important for success. This can be especially true over the summer months given that many students may rather be outside playing than inside reading a book or being drilled on the multiplication tables. As such, measures of how often and for how long students attend programming (typically known as dosage) can provide some information about a student's engagement with the program, though they do not provide the full picture.

Dosage in the academic context is a measure of students' engagement with programming, and there are a number of ways of measuring it. Chaput et al define dosage as consisting of three different domains: duration, intensity, and breadth.[17] These different measures indicate different types of engagement that students can experience with a program. Duration is the length of time (e.g., one year, two years, etc.) that a student has received programming. Intensity is the frequency (i.e., days/week, hours/day) at which a student receives programming. Breadth is the amount of different activities that a student is involved in within programming.[18]

Importance of adequate dosage – A 2010 study of the 21st Century Community Learning Centers reported that regular program attendance is the key to producing positive outcomes for students. Further, how students spend their time in programming is also an important factor in academic success. The study reports, "a review of the literature on time and learning found that there was no relationship between allocated time and student academic achievement, some relationship between engaged time and achievement, and a greater relationship between time spent on academic learning and achievement," which suggests that the higher the dosage of academic supports, the better the academic outcomes for students.[19]

The study also asserts that optimal dosage is highly dependent on the individual student's needs, "Regular attendance by students who are most in need of after-school programming is necessary for them to benefit from instruction".[20] This vague assertion suggests that students who tend to need after school programming the most (i.e., those falling behind in school) tend to need a higher dosage of programming – here denoted as "regular attendance" – in order to experience the full benefits of the program and make progress in catching up to their peers. What, then, is "regular attendance"? The Department of Education's annual performance guidelines for 21st Century Community Learning Centers define regular attendance as 30 nonconsecutive days or more, but the rationale behind this guideline is unknown at this time.[21]

A 2012 evaluation of a local tutoring program in Saint Paul, MN, conducted by Wilder Research does provide some support for a threshold model of attendance. While the evaluation was meant to be exploratory rather than confirmatory, the results do point to the importance of dosage in making academic gains with students. Specifically, the evaluation found that students who attended more than 40 days of tutoring made higher gains than those who attended less than 40 days.[22] While this is a higher threshold than the one put forth by 21stCCLC, it demonstrates the idea that when students hit a certain amount of tutoring it can have a positive impact on their academic scores. This higher threshold can also be explained by looking at the population served by the tutoring program evaluated: they are mostly minority, low-income students who are falling behind academically. So, these may be students who simply need more instructional time than other students in order to make reading gains.

Support for methodology – Studies that wish to measure the relationship between tutoring attendance and academic outcomes must grapple with serious methodological limitations. First is the issue of selection bias – it's likely that students who attend tutoring more regularly or receive more tutoring are systematically different than those students who receive less tutoring or attend less often. This issue is closely linked to the endogeneity of the outcome variables and explanatory variables of interest (tutoring dosage), as students are often selected into tutoring services due to lagging academic performance. Thus, the actual causal mechanism at work is unclear; does low achievement lead to higher tutoring doses or do higher tutoring doses lead to better academic outcomes? Alternately, are any positive outcomes of higher tutoring dosages attenuated by low academic performance? A related issue is that of selection on unobservable characteristics like parent involvement and student motivation.[1]

E. Michael Foster, in a 2003 study of therapy dosage and mental health outcomes, noted that patients who tend to receive the highest treatment dosages are those that tend to have the greatest needs, and thus may end up having worse mental health outcomes.[23] The same logic can be applied to tutoring dosage, as described above. Foster explored the application of propensity score methods in determining the treatment effect of dosage, finding that such methods are valuable in analyzing dosage effects. Further, his analysis using propensity score methods suggested that "added services improve treatment outcomes, especially child functioning. However, at least for the services and outcomes considered, the marginal benefits to high levels of treatment are limited".[24]

Propensity score methods have been used to identify dosage effects in the academic realm as well. In a 2014 study, Arteaga et al explored the long-term effects of zero, one, or

---

[1] For example, if students with highly engaged parents have better attendance in tutoring, then the observed relationship between tutoring and academic outcomes may overstate the causal impact of dosage.

two years of preschool. Recognizing that children who receive one or two years of preschool may be systematically different from those who receive none, the authors applied inverse propensity score weighting to correct for selection bias. They note that propensity score weighting is useful when dealing with multiple treatment groups (rather than simply comparing treated versus control) and that it handily deals with selection on observables like prior academic achievement without requiring the use of instrumental variables.[25] As such, propensity score weighting seems like a good method for dealing with both the non-random selection into tutoring as well as the multiple treatment categories inherent in this analysis.

Data:

This analysis uses the Early Childhood Longitudinal Study—Kindergarten Class (ECLS-K), a nationally representative survey of the 1998-1999 kindergarten cohort. According the National Center for Education Statistics "the ECLS program provides national data on children's status at birth and at various points thereafter … The ECLS program also provides data to analyze the relationships among a wide range of family, school, community, and individual variables with children's development, early learning, and performance in school".[26] The study followed students from Kindergarten through grade 8 and includes information about the students, their parents, their teachers, and their schools. The ECLS-K dataset includes students from both public and private schools. Information was collected in the fall and the spring of kindergarten (1998-99) and first grade (1999-2000), and the spring of 3rd grade (2002), 5th grade (2004), and 8th grade (2007).

The ECLS-K dataset includes numerous different weights, strata variables, and primary sampling unit variables for use in analysis. Different weights are recommended for different types of analysis. However, according to the ECLS-K website, "Researchers are encouraged to use the same weight throughout all analyses in a publication or paper, even when there is a different ideal weight for each analysis."[27] As such, this analysis uses the C3CW0 weight and its associated strata and PSU variables, which is appropriate for the analysis of direct child assessment scores collected during the third round of data collection as well as demographic data connected to students. Since this analysis focuses on the relationship between summer tutoring attendance (collected during the third round of data collection) and student assessment scores during the third round of data collection (fall of first grade) this weight is the most useful.

This analysis focuses on math and reading test scores as the outcome variables of interest. The ECLS-K reports a number of different reading and math scores, though not all are appropriate for the purposes of this analysis. As such, this analysis uses the item response theory scale scores for math and reading. As the name suggests, these scores are based on item response theory, meaning that they take into account correct answers, incorrect answers, and characteristics of each test item to generate an overall ability score for each student. These scores are criterion referenced, meaning that interpretation of the score tells us something about the level of mastery or proficiency a student has attained on certain criteria. Even though the student assessment used to generate these scores is scaled for student's growth, the IRT scale scores are comparable across time due to how they are calculated.

In addition to overall math and reading IRT scale scores, this analysis explores the impact of tutoring on specific math and reading proficiency levels. The ECLS-K reports the highest level of proficiency reached by a particular student at the time of assessment based on IRT scores. Both math and reading are comprised on 9 proficiency levels, though for the purposes of this analysis, only the first three levels are considered. This is because the age group of students considered in this paper were overall mainly proficient in only the first three levels of math and reading proficiency. Each proficiency level is associated with particular math and reading skills which build upon each other and provide more descriptive information about a student's ability than would an overall test score. As such, proficiency level scores allow researchers to examine the specific skills which are impacted by a particular educational intervention. The proficiency levels for math and reading are:

- Math:
  - Level 1: count, number, and shape, which measures a student's mastery of the basic mathematical concepts related to recognizing numbers and shapes.
  - Level 2: relative size, which measures mastery of concepts related to size.
  - Level 3: ordinality and sequence, which measures mastery of concepts related to the order of numbers.
- Reading:
  - Level 1: letter recognition, which measures mastery of basic alphabetical shapes and their names.
  - Level 2: beginning sounds, which measures mastery of recognizing the pronouncing the beginning sounds of words.
  - Level 3: ending sounds, which measures mastery of recognizing the pronouncing the ending sounds of words.

Finally, this paper briefly explores the effect of summer tutoring dosage directly on summer slide. The summer slide variable is a binary indicator and was generated by assigning a value of 1 to any student whose fall first grade score in math and/or reading was lower than their spring of kindergarten score. A value of 0 was assigned to any student whose fall of first grade scores in reading and/or math were greater than or equal to their kindergarten spring scores. Thus, a 1 indicates that a student did experience summer learning loss, whereas a 0 indicates that they did not.

Table 1 shows average kindergarten spring test scores in overall reading and math as well as the first three reading and math proficiency levels. This gives a summary of where students were academically, on average, before entering into the summer break between kindergarten and first grade. For example, non-tutored students averaged about 45 in reading and 36 in math. Overall, students who received the lowest and highest levels of dosage scored lower in reading and math than their noon-tutored peers. The proficiency probability rates can be interpreted as the probability that students reached mastery of that level of skill. So, for example, the probability that non-tutored students were proficient in the first level of reading in the spring of kindergarten was about 83%. These baseline numbers are helpful in understanding the effects estimated for tutoring and dosage later in this paper.

Table 1:Mean spring kindergarten test scores - by tutored status

	Non-tutored	1-10 hours	12-30 hours	32-100 hours	110-480 hours
Reading	45.172	41.190	47.491	46.340	40.172
Math	36.196	32.734	37.675	37.594	34.180
Proficiency probability					
Read 1	0.834	0.737	0.950	0.835	0.832
Read 2	0.631	0.503	0.733	0.645	0.459
Read 3	0.454	0.337	0.537	0.471	0.295
Math 1	0.946	0.953	0.992	0.937	0.942
Math 2	0.802	0.777	0.854	0.796	0.826
Math 3	0.513	0.410	0.567	0.541	0.482

N: 5278 Weighted count: 3,729,994

The treatment variables used in this analysis measure different treatment effects of tutoring. A binary tutoring variable, “Tutored”, which equals one when a student received tutoring and 0 when a student did not, measures the impact of receiving any tutoring over the summer. A series of indicator variables, “Dose1-4”, is also used to measure the impact of tutoring dosage. A dose indicator is assigned to a student based on the amount of total tutoring they received over the summer with 1 indicating 1-10 hours, 2 indicating 12-30 hours, 3 indicating 32-100 hours, and 4 indicating 110-480 hours. These cut points were determined based on natural cut-offs of very low, moderate, high, and very high levels of dosage. The dose indicator variables were generated by multiplying the three indicators of dosage (hours per day, days per week, and weeks) into a continuous dosage variable which was then cut into the above categories. Table 2 shows the distribution of tutoring dosage.

Table 2: Distribution of Dosage

Hours per day	Percentage
1	64.20%
2	17.80%
3	9.38%
4	2.26%
5	3.11%
6	1.31%
8	1.89%
Days per week	
1	21.10%
2	30.30%
3	18.80%
4	7.41%
5	22.40%
Weeks	
1	3.44%
2	10.70%
3	5.72%
4	17.10%
5	7.91%
6	7.87%
7	1.53%
8	19.10%
9	3.98%
10	11.30%
11	0.72%
12	10.00%
14	0.72%
Total dosage	
1-10 hours	40.80%
12-30 hours	28.00%
32-100 hours	15.60%
110-480 hours	15.70%

N: 183 Weighted count: 120,251.8

Some important covariates are needed as controls in the estimation equation. The covariates used in this analysis are race, gender, income quintile, whether or not the student attended a public or private school, and reading and math test scores for the previous period of data collection. These variables are used as controls in this analysis because they are hypothesized to be important confounding variables between academic interventions and achievement scores. By controlling for these variables in the estimation model, we can produce less-biased estimates of the effect of summer tutoring on reading and math scores. Tables 3 and 4 show how these characteristics are distributed across tutored and non-tutored students as well as the different levels of dosage.

Table 3: Characteristics of first grade students – by summer tutoring status

Characteristic	Received tutoring		Did not receive tutoring	
	Rate/Mean	SE	Rate/Mean	SE
Gender				
Male	48.2%	4.99	51.3%	0.846
Female	58.1%	4.99	48.40%	0.863
Race				
White	45.8%	10.4	57.80%	2.52
Non-white	54.2%	10.4	42.20%	2.52
Income				
First Quintile	92.1%	3.98	17%	2.70
Second Quintile	5.36%	3.45	23.7%	3.08
Third Quintile	1.36%	0.973	20.1%	3.61
Fourth Quintile	0.628%	0.65	20.3%	3.26
Fifth Quintile	0.52%	0.536	19.4%	2.46
Mean spring Kindergarten test score				
Reading	48.95	1.93	47.46	0.71
Math	47.80	1.61	49.62	0.39

N: 5278 Weighted count: 3,729,994

Table 4: Characteristics by dosage level

Characteristic	1-10 hours	12-30 hours	32-100 hours	110-480 hours
Female	57.8%	41.8%	54.9%	50.5%
Race				
White	71.5%	66.7%	48.6%	68.3%
Non-white	28.5%	33.3%	51.4%	31.7%
Income				
Bottom Quintile	25.7%	34.1%	24.8%	11.1%
Top Quintile	27.8%	8.78%	30.8%	24.3%
Mean spring K test scores				
Reading	54.92	50.61	53.58	55.12
Std. dev.	1.24	2.53	2.48	3.92
Math	53.64	52.99	54.20	52.22
Std. dev.	1.04	1.50	2.19	3.08

N: 183 Weighted count: 120,251.8

Methods:

The effect of summer tutoring on continuous math and reading test scores is analyzed using a multivariate linear regression model. This linear model is performed for each tutoring treatment variable. The model which examines the effect of any summer tutoring versus no summer tutoring is: $Y = \beta_0 + \beta_1 TUTORING + \beta_2 X + \epsilon$

Where Y = child's test score in math or reading

X = individual demographic covariates

TUTORING = the summer tutoring variable which equals one when the student did receive tutoring over the summer and 0 when they did not.

β_1 = coefficient on the tutoring variable which estimates the size and direction of tutoring's impact on test scores. Any coefficient with a p-value less than 0.1 is considered statistically significant for the purposes of this paper.

The model which estimates the impact of each separate dosage level on test scores is:

$Y = \beta_0 + \beta_1 DOSE_{1-4} + \beta_2 X + \epsilon$

Where $DOSE_1$ =1 if the student received 1-10 hours of tutoring and 0 otherwise

$DOSE_2 = 1$ if the student received 12-30 hours of tutoring and 0 otherwise

$DOSE_3 = 1$ if the student received 32-100 hours of tutoring and 0 otherwise

$DOSE_4 = 1$ if the student received 110-480 hours of tutoring and 0 otherwise

β_1 = coefficient on the dummy dosage variable which estimates the size and direction of each dosage level's impact on test scores.

The effect of summer tutoring on reading and math proficiency level mastery is analyzed using an ordered logit model. The model which examines the effect of any tutoring on reading and math proficiency is:

$$\Pr(Y = 1,2,3) = \frac{1}{1 + \left(\frac{1}{e^{\beta_0 + \beta_1 TUTORING + \beta_2 X + \epsilon}}\right)}$$

Where Pr(Y=1,2,3) = the probability that a student will have mastered either the first, second, or third level of proficiency.

Since the coefficients produced from this model cannot be directly interpreted, the marginal effect of tutoring was estimated for each level of mastery. These marginal effects are interpreted as tutoring's impact on a student's mastery of proficiency levels in reading and math.

The ordered logit model estimating the impact of each dosage level on proficiency level mastery is much the same as the above model for the impact of any tutoring. However, the dosage model replaces the binary tutoring variable with dummy variables for each level of tutoring dosage. In this case, when marginal effects are estimated, they are interpreted as each dosage level's impact on a student's mastery of a proficiency level.

Students were not randomly selected into tutoring nor were they randomly selected into different dosage levels. As such, students who received tutoring over the summer are likely to be systematically different than those who did not receive tutoring. This introduces bias into the model. This analysis employs inverse propensity score weighting (IPSW) as a way to counteract this bias effect. IPSW assigns a weight to each individual that is equal to the inverse of their probability of being selected into a tutoring dosage level. The propensity score is generated using a multinomial logit model that includes all variables thought to influence selection into the treatment group. The equation used to produce propensity scores is: $pdoselevel = \beta_0 + \beta_1 X + \epsilon$

Where pdoselevel = the probability of receiving a particular level of dosage for the four dosage categories

X = demographic and family variables which may affect selection into different levels of tutoring dosage.

The variables used to estimate the propensity score are race, gender, income (in quintiles), prior achievement measured by spring Kindergarten reading and math scores, whether the student attended a public or private school, and family engagement as measured by how many parent-teacher meetings and other similar meetings were attended by a member of the student's family during the school year. After estimating a propensity score for each dosage level, students are assigned a weight that is the inverse of the probability of the dosage they actually received. For example, if a student was in the 12-30 hour category and their probability of being in that category was estimated at 0.4, then their assigned weight would be 1/0.4 = 2.5.

Finally, the dosage effect of summer tutoring on the summer slide in reading and math is analyzed using a logit model:

$$\Pr(Y=1) = \frac{1}{1 + \left(\frac{1}{e^{\beta_0 + \beta_1 DOSE_{1-4} + \beta_2 X + \epsilon}}\right)}$$

Where Pr(Y=1) is the probability that the student experienced learning loss between spring of kindergarten and fall of first grade

$DOSE_{1-4}$ =the dummy dosage variable indicating the amount of summer tutoring received by the student.

Results:

Table 5 shows the results of a linear regression model regressing overall fall first grade math and reading scores on the binary tutoring variable. Overall, summer tutoring was not found to have a statistically significant impact on fall math and reading test scores. However, with all covariates added to the model, summer tutoring appears to have a negative impact on fall reading scores. This suggests that, after accounting for potential confounding demographic factors, students who receive tutoring during the summer perform slightly worse on the fall first grade reading assessment, on average, than students who do not receive tutoring. Additionally, results from this analysis suggest that black, Hispanic, and Native American students all perform worse, on average, than white students on the reading and math assessments, but these effects may be overestimated when prior math and reading achievement is not controlled for.

There are a number of ways in which the accuracy of the results produced from this base linear model may be in doubt. For one, even though controlling for demographic confounders, this model does not take into account selection bias into tutoring. Further, the model does not allow us to explore differing levels of attendance in tutoring, which previous research suggests can have an impact on student outcomes.[28] Further, this model only reports tutoring's impact on overall test scores, ignoring the fact that it may have a different effect on specific math and reading skills, a quirk which is not picked up in overall test scores. So while, this model can provide a snapshot of summer tutoring's effect on academic achievement, taking these results as the ultimate test of summer tutoring is not advisable.

Table 5: Effect of tutoring on math and reading scores - estimates from OLS regression (standard errors in parentheses)

	1		2		3		4		5	
	Math IRT scale	Read IRT scale	Math IRT scale	Read IRT scale	Math IRT scale	Read IRT scale	Math IRT scale	Read IRT scale	Math IRT scale	Read IRT scale
Summer Tutoring	-2.054	-2.516	-2.041	-2.596	-0.682	-1.109	-0.532	-0.767	0.262	-1.822*
	(2.646)	(2.294)	(2.639)	(2.297)	(3.275)	(2.738)	(3.296)	(2.838)	(1.240)	(1.028)
Female	--	--	-0.402	2.798***	-0.562	2.690***	-0.577	2.670***	-0.683*	0.076
	--	--	(0.525)	(0.660)	(0.546)	(0.689)	(0.548)	(0.695)	(0.346)	(0.474)
Black	--	--	--	--	-8.690***	-6.821***	-8.702***	-6.835***	-2.076***	-1.662***
	--	--	--	--	(0.805)	(0.969)	(0.811)	(0.977)	(0.462)	(0.501)
Hispanic	--	--	--	--	-9.153***	-7.357***	-9.181***	-7.397***	-1.325***	2.184***
	--	--	--	--	(0.885)	(0.844)	(0.891)	(0.858)	(0.487)	(0.682)
Native	--	--	--	--	-12.256***	-13.080***	-12.217***	-12.979***	-3.906***	-3.523***
	--	--	--	--	(1.661)	(2.027)	(1.672)	(2.082)	(0.929)	(1.089)
Income level (quintiles)	--	--	--	--	--	--	0.082	0.184	-0.142	-0.013
	--	--	--	--	--	--	(0.242)	(0.309)	(0.128)	(0.206)
Spring K Math test score	--	--	--	--	--	--	--	--	1.034***	--
	--	--	--	--	--	--	--	--	(0.031)	--
Spring K Read test score	--	--	--	--	--	--	--	--	--	1.098***
	--	--	--	--	--	--	--	--	--	(0.056)

****=significant at the 0.01 level, **=significant at the 0.05 level, *=significant at the 0.10 level.*
Reading and math scores were measured in the fall of 1st grade.

In order to get a more granular understanding of the relationship between tutoring and academic achievement, it may be helpful to examine the impact of differing levels of tutoring attendance. Table 6 reports results from an analysis of the effects of tutoring dosage on overall math and reading scores. After accounting for all covariates, 1-10 hours of tutoring was found to have no effect on math scores, on average. However, this level of tutoring was found to have a large, negative impact on reading test scores, on average. This negative impact may suggest that such a small amount of summer tutoring is not effective in mitigating the effects of summer slide in reading skills. Further, 12-30 hours of tutoring is found to have a small, negative impact on math scores and a positive impact on reading scores, on average. The marked reversal in the impact on reading scores between these two levels of dosage suggests a stark contrast between them, potentially revealing an important threshold of dosage at 12 hours, particularly for reading ability. The 32-100 hour dosage level presents a conundrum, however, as it shows no impact on reading scores and actually increases the negative impact on math scores, on average. This effect is reversed, however, at the 110-480 level of dosage, which produces positive effects for both reading and math scores, on average. The overall negative impact on math scores suggests that, unless a student receives the highest level of dosage, summer tutoring may not be as effective at raising math scores as it is for reading scores.

Table 6: Effect of dosage on math and reading scores - estimates from OLS regression (standard errors in parentheses)

	1		2		3		4		5	
	Math IRT scale	Read IRT scale	Math IRT scale	Read IRT scale	Math IRT scale	Read IRT scale	Math IRT scale	Read IRT scale	Math IRT scale	Read IRT scale
1-10 hours	1.620	-2.129	0.914	-5.661***	--	--	--	--	--	--
	(1.161)	(1.380)	(0.902)	(0.190)	--	--	--	--	--	--
12-30 hours	-4.200**	-5.441*	--	--	-1.491*	2.918**	--	--	--	--
	(0.481)	(1.450)	--	--	(0.381)	(0.420)	--	--	--	--
32-100 hours	1.796	5.507**	--	--	--	--	-2.825**	1.420	--	--
	(0.684)	(0.811)	--	--	--	--	(0.294)	(0.813)	--	--
110-480 hours	1.694	6.809***	--	--	--	--	--	--	3.252*	4.159***
	(1.105)	(0.490)	--	--	--	--	--	--	1.065	0.237
Female	--	--	0.748	1.167	0.708	1.118	0.883	0.984	0.842	1.114
	--	--	(0.359)	(1.030)	(0.342)	(0.958)	(0.338)	(0.958)	(0.386)	(0.936)
Black	--	--	2.170	1.167	2.112	1.425	2.409	1.267	2.273	1.581
	--	--	(1.765)	(0.970)	(1.743)	(1.258)	(1.841)	(1.228)	(1.775)	(1.146)
Hispanic	--	--	-0.052	7.233	-0.005	7.010	0.359	6.680	0.543	7.420
	--	--	(1.245)	(3.548)	(1.188)	(3.814)	(1.421)	(3.620)	(1.134)	(3.515)
Native	--	--	1.731	-1.207	1.464	0.425	2.114*	0.033	1.201	-0.250
	--	--	(0.814)	(1.016)	(0.537)	(0.872)	(0.622)	(0.915)	(0.607)	(0.839)
Income level (quintiles)	--	--	0.066	1.006	0.033	1.095	0.115*	0.915	0.046	1.066
	--	--	(0.034)	(0.499)	(0.027)	(0.542)	(0.035)	(0.555)	(0.053)	(0.571)
Spring K Math test score	--	--	1.308***	--	1.304***	--	1.325***	--	1.313***	--
	--	--	(0.047)	--	(0.043)	--	(0.045)	--	(0.040)	--
Spring K Read test score	--	--	--	1.582***	--	1.591***	--	1.570***	--	1.567***
	--	--	--	(0.075)	--	(0.073)	--	(0.071)	--	(0.071)

***=significant at the 0.01 level, **=significant at the 0.05 level, *=significant at the 0.10 level
Reading and math scores were measured in the fall of 1st grade.

In an effort to understand summer tutoring's differing effects on reading and math scores, an ordered logit model regressing math and reading proficiency levels on the tutoring treatment variables was run. Table 7 shows the results of that model, presenting the marginal effects of any tutoring as well as each level of dosage on mastery of each proficiency level. Any tutoring at all and the lowest level of tutoring dosage are shown to have no effect on any proficiency level, across the board. Additionally, both the 12-30 and 21-100 hour level of dosage have no effect on math proficiency levels. They do have an effect on reading proficiency levels, though. These two levels of dosage show similar patterns, with negative effects on mastery of the first two levels of reading proficiency and positive effects on the third level. This suggests that students receiving low to moderate levels of tutoring over the summer may have spent the majority of their time working on some higher level reading skills.

Statistically significant effects are found for all reading and math proficiency levels at the highest level of dosage, 110-480 hours. This level of dosage's effects on reading proficiency levels are surprising though, as the effect is positive for the first two levels and negative for the third level. This could suggest something about the types of students who receive each level of dosage. Perhaps those receiving the highest amount of dosage are indeed students in most need of tutoring. This would explain why tutoring would have a positive impact on the lower levels of proficiency, which struggling students would spend the most time working on, and a negative effect of the higher level of proficiency. Later analysis using inverse propensity score weighting will provide more clarity around this dose response for particular proficiency levels. Finally, the effects on math proficiency levels follow a pattern of negative effects for lower levels and positive effect for the higher level of proficiency. Taken in conjunction with the results of the above model, this provides more evidence for extremely high levels of tutoring dosage being most effective for raising math scores.

Table 7: Marginal effects of tutoring and dosage on Proficiency levels (standard errors in parentheses)

	Any tutoring	1-10 hours	12-30 hours	32-100 hours	110-480 hours
Read 1	0.014	-0.007	-0.017***	-0.006***	0.034***
	(0.0257)	(0.005)	(0.001)	(0.001)	(0.005)
Read 2	0.012	-0.018	-0.041***	-0.015***	0.085***
	(0.022)	(0.012)	(0.003)	(0.004)	(0.013)
Read 3	-0.026	0.025	0.058***	0.022***	-0.118***
	(0.048)	(0.017)	(0.003)	(0.005)	(0.018)
Math 1	-0.010	0.006	0.005	-0.001	-0.017***
	(0.018)	(0.005)	(0.006)	(0.004)	(0.005)
Math 2	-0.013	0.016	0.015	-0.004	-0.051***
	(0.022)	(0.014)	(0.020)	(0.011)	(0.015)
Math 3	0.023	-0.022	-0.020	0.005	0.069***
	(0.040)	(0.019)	(0.027)	(0.014)	(0.020)

***=significant at the 0.01 level, **=significant at the 0.05 level, *=significant at the 0.10 level
Proficiency level mastery was measured in fall of 1st grade.
Analysis included all covariates from Tables 5 and 6, not shown here.

Tables 8 and 9 explore predicted probabilities of mastering the three foundational proficiency levels for each level of dosage by race and income. Based on previous research indicating gaps between white and black students as well as low income and high-income students, this analysis focuses specifically those groups of students. For reading proficiency, the largest gaps that exist between white and black students are at the second and third proficiency levels. Across dosage levels, these differences in predicted probabilities remain fairly stable. At the second level of reading proficiency (beginning sounds) black students are found to have a much higher predicted probability of mastery. Moving on to the third level of proficiency, though, indicates that white students have a much higher probability of mastery. This pattern suggests that, even within the same dosage category, black students are not predicted to reach as high levels of reading proficiency as white students are. Moving on to the analysis of math proficiency, predicted probabilities of mastery are much more even across white and black students. While white students, across dosage levels, are about 5 percentage points more likely to mastery the third level of proficiency than black students, this difference is not nearly as marked as the difference observed in reading proficiency.

Table 8: Predicted probabilities of proficiency level by dosage and race (standard errors in parentheses)

	1-10 hours		12-30 hours		32-100 hours		110-480 hours	
	White	Black	White	Black	White	Black	White	Black
Read 1	0.39%***	3.02%***	0.38%***	2.99%***	0.39%***	3.06%***	0.35%***	2.80%***
	(0.000)	(0.007)	(0.000)	(0.006)	(0.000)	(0.007)	(0.000)	(0.006)
Read 2	12.32%***	50.66%***	12.11%***	50.56%***	12.31%***	50.67%***	11.85%***	50.50%***
	(0.016)	(0.066)	(0.018)	(0.057)	(0.017)	(0.065)	(0.016)	(0.055)
Read 3	87.29%***	46.32%***	87.51%***	46.45%***	87.30%***	46.26%***	87.80%***	46.70%***
	(0.016)	(0.073)	(0.018)	(0.063)	(0.017)	(0.072)	(0.016)	(0.061)
Math 1	0.02%***	0.05%***	0.02%	0.05%	0.02%***	0.05%***	0.02%***	0.05%***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Math 2	2.88%***	7.93%***	2.89%***	7.45%***	2.93%***	7.79%***	2.69%***	7.00%***
	(0.006)	(0.015)	(0.005)	(0.015)	(0.006)	(0.015)	(0.005)	(0.011)
Math 3	97.10%***	92.02%***	97.10%***	92.50%***	97.05%***	92.16%***	97.29%***	92.95%***
	(0.006)	(0.015)	(0.005)	(0.015)	(0.006)	(0.015)	(0.005)	(0.011)

***=significant at the 0.01 level, **=significant at the 0.05 level, *=significant at the 0.10 level
Proficiency level mastery was measured in fall of 1st grade.
Analysis included all covariates from Tables 5 and 6, not shown here.

The predicted probabilities of mastering each proficiency level by high-income and low-income subgroups of students do not show as extreme differences as those observed between white and black students. Reading and math both show a pattern of steadily increasing predicted probabilities, with both groups of students exhibiting the highest probability of mastery at the third level of proficiency in reading and math. Again, we see that differences between the two groups of students remain stable across dosage levels. The particularly surprising part of this analysis is that high-income students, on average, have a lower predicted probability of mastery of the second and third levels of proficiency for reading math even within dosage levels.

Table 9: Predicted probabilities of proficiency level by dosage and income level (standard errors in parentheses)

	1-10 hours		12-30 hours		32-100 hours		110-480 hours	
	Low	High	Low	High	Low	High	Low	High
Read 1	0.91%***	1.27%***	0.93%***	1.28%***	0.94%***	1.28%***	0.79%***	1.20%***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Read 2	24.66%***	31.20%***	25.03%***	31.39%***	24.85%***	31.00%***	23.16%***	31.35%***
	(0.051)	(0.015)	(0.051)	(0.017)	(0.052)	(0.015)	(0.046)	(0.018)
Read 3	74.43%***	67.53%***	74.05%***	67.33%***	74.21%***	67.72%***	76.05%***	67.44%***
	(0.053)	(0.016)	(0.053)	(0.019)	(0.054)	(0.017)	(0.048)	(0.020)
Math 1	0.02%***	0.10%***	0.02%***	0.11%***	0.02%***	0.11%***	0.02%***	0.10%***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Math 2	3.78%***	14.53%***	3.46%***	15.37%***	3.70%***	15.40%***	3.49%***	13.89%***
	(0.005)	(0.043)	(0.005)	(0.047)	(0.005)	(0.043)	(0.006)	(0.047)
Math 3	96.19%***	85.37%***	96.52%***	84.52%***	96.27%***	84.49%***	96.49%***	86.01%***
	(0.005)	(0.044)	(0.004)	(0.048)	(0.005)	(0.043)	(0.006)	(0.048)

***=significant at the 0.01 level, **=significant at the 0.05 level, *=significant at the 0.10 level
Proficiency level mastery was measured in fall of 1st grade.
Analysis included all covariates from Tables 5 and 6, not shown here.

Since the coefficients reported by the OLS regression model may be biased due to selection, it is important to try to address this problem in order to get an even more accurate view of the relationship between tutoring attendance and test scores. To that end, propensity score weighting was used to explore this relationship while correcting for bias in selecting into different levels of tutoring dosage. Table 10 shows the variables that were used to generate the each student's propensity scores as well as each variable's marginal effect on the probability of being selected into each category of tutoring.

Students were more likely to receive 1-10 hours of tutoring if they were Hispanic, had higher reading scores, and their family was more engaged in school. Conversely, Asian and Native American students were less likely to receive this level of dosage. Additionally, students from higher income families, with higher math scores, and who attended public school were less likely to receive 1-10 hours of tutoring. At the 12-30 hour dosage level, the only variables with a statistically significant effect on selection were being black and having a higher math score in the spring of kindergarten. Females, students with higher reading scores, public school attendees, and students whose parents were more engaged in their schooling were less likely to receive 12-30 hours of tutoring. The 32-100 hours dosage level had the most factors that positively affected selection into this category. Hispanic, Asian, and Native American students were all more likely to receive 32-100 hours of tutoring. Additionally, females, students from higher income families, with higher math scores in the spring of kindergarten, and who attended public school were also more likely to select into this level of dosage. Higher reading scores and more family engagement had a negative effect on selection into this dosage level. Finally, at the highest level of dosage, 110-480 hours, only higher reading scores and attending public school were found to have a positive effect on selection. Conversely, black students, Hispanic students, females, and student with higher kindergarten spring math scores were less likely to receive 110-480 hours of tutoring.

This analysis reveals some noteworthy patterns among the factors that affect selection into different levels of tutoring. For instance, attending a public school has a negative effect on selection into the two lower levels of dosage and a positive effect on selection into the two higher levels of dosage. Additionally, family engagement, which is often thought of as a bell weather for academically-strong students, has a positive effect on selection into the lowest level of tutoring dosage, a negative effect on 12-30 hours and 32-100 hours, and no effect on the highest level of dosage. Kindergarten spring reading and math scores also appear to have the opposite effect on selection into dosage categories, with higher reading scores positively effecting selection into the top and bottom dosage levels whereas higher math scores have a negative effect on selection into these categories.

Table 10: Marginal Effects of Covariates on Selection into Tutoring Dosage Levels

Variable	1-10 hours Marginal Effect	1-10 hours SE	12-30 hours Marginal Effect	12-30 hours SE	32-100 hours Marginal Effect	32-100 hours SE	110-480 hours Marginal Effect	110-480 hours SE
Race								
Black	0.005	0.025	0.040**	0.019	0.040	0.030	-0.085***	0.017
Hispanic	0.117*	0.066	-0.028	0.100	0.081*	0.047	-0.170***	0.018
Asian	-0.357***	0.012	0.016	0.073	0.361***	0.018	-0.020	0.044
Native	-0.311***	0.039	0.067	0.042	0.209***	0.010	0.035	0.024
Female	0.033	0.037	-0.039**	0.016	0.032**	0.014	-0.026*	0.015
Income (in quintiles)	-0.019**	0.010	-0.009	0.009	0.030***	0.006	-0.003	0.009
Spring K reading score	0.007***	0.002	-0.012***	0.001	-0.003***	0.001	0.008***	0.002
Spring K math score	-0.010***	0.002	0.008***	0.001	0.009***	0.002	-0.007***	0.001
Public school	-0.133**	0.067	-0.217***	0.020	0.210***	0.039	0.139***	0.025
Family engagement	0.385***	0.054	-0.166**	0.072	-0.188***	0.017	-0.030	0.098

****=significant at the 0.01 level, **=significant at the 0.05 level, *=significant at the 0.10 level*

Table 11 reports the effects of each level of tutoring attendance on reading and math test scores. The results of the inverse propensity score weighted analysis show some marked differences from the OLS estimates. After running the model using IPSW, all but the 32-100 hour level of dosage are found to have no effect on fall of first grade math scores. At the 32-100 hour dosage level, there is a large negative effect on math scores. Examination of the data showed that this large negative effect is not likely due to outliers or to population weights. Further, when using IPSW, all but the 12-30 hour level of dosage are found to have a statistically significant effect on fall of first grade reading scores. Additionally, the negative effect on 1-10 hours of tutoring which is found in the OLS regression is reduced in the IPSW model. However, there is also a reduction in the magnitude of the effect of the highest level of dosage, 110-480 hours. Overall, when accounting for selection on observables into different levels of dosage, no amount of dosage is found to have a positive impact on math test scores, and 32-100 hours appears to have to largest positive impact on reading scores.

Table 11: Effects of dose on Math and Reading Scores - OLS and IPSW comparison

	OLS		IPSW	
	Math	Reading	Math	Reading
Dose				
1-10 hours	0.914	-5.661***	0.371	-2.958**
12-30 hours	-1.491*	2.918**	0.623	-1.846
32-100 hours	-2.825**	1.420	-3.206***	4.247*
110-480 hours	3.252*	4.159***	2.100	1.244*

****=significant at the 0.01 level, **=significant at the 0.05 level, *=significant at the 0.10 level*
Regressions included all covariates from Tables 5 and 6, not shown here.

Table 12 shows a similar comparison between OLS estimates and IPSW estimates, except that the estimates are not regression coefficients but rather marginal effects produced from running the an ordered logit model of dosage on math and reading proficiency levels. Taking into account the differential selection of students into different dosage categories using IPSW produces some marked differences from the original model. For instance, the 1-10 hour level of dosage is found to have significant negative effects on the first two levels of proficiency for reading and math, and a positive effect of mastery of the third proficiency level for both subjects. Additionally, the direction of the effect of 12-30 hours on reading proficiency reverses, showing a positive effect on the first two proficiency levels and a negative effect on the third. At the 32-100 hour level, the magnitude of the effect on reading proficiency levels increases, showing a larger negative effect for the first two levels of proficiency as well as a larger positive effect for the third level. There is also evidence of a statistically significant effect on math proficiency levels at this amount of dosage, a result that was not found in the original model. Finally, the 110-480 hours level of dosage is found to have no effect on math proficiency levels.

Table 12: Marginal effects of dosage on proficiency - comparison of OLS and IPSW estimates

	OLS			
	1-10 hours	12-30 hours	32-100 hours	110-480 hours
Read 1	-0.007	-0.017***	-0.006***	0.034***
Read 2	-0.018	-0.041***	-0.015***	0.085***
Read 3	0.025	0.058***	0.022***	-0.118***
Math 1	0.006	0.005	-0.001	-0.017***
Math 2	0.016	0.015	-0.004	-0.051***
Math 3	-0.022	-0.020	0.005	0.069***
	IPSW			
	1-10 hours	12-30 hours	32-100 hours	110-480 hours
Read 1	-0.038***	0.020*	-0.043***	0.063***
Read 2	-0.057***	0.029**	-0.061***	0.093***
Read 3	0.095***	-0.050*	0.104***	-0.156***
Math 1	-0.009***	0.007**	0.002***	-0.003
Math 2	-0.048***	0.041*	0.012***	-0.015
Math 3	0.057***	-0.047*	-0.015***	0.017

****=significant at the 0.01 level, **=significant at the 0.05 level, *=significant at the 0.10 level*
Regressions included all covariates from Tables 5 and 6, not shown here.

Table 13 presents the results of an analysis of tutoring dosage's effects not on overall math and reading scores but on a student's likelihood to experience summer slide. This analysis indicates that student who receive 1-10 hours of tutoring over the summer are 4.4 percentage points less likely to experience the summer slide in the their math scores as compared to students receiving other amounts of dosage. Further, students who received the highest level of dosage were 11.5 percentage points less likely to experience summer slide in math. At the 12-30 hour level though, students were 11.1 percentage points more likely to experience the summer slide. This suggests that in order for summer tutoring to have a significant and large positive impact on mitigating the summer slide in math, students must receive at least 110 hours of tutoring. For summer slide in reading, the first two levels of dosage show that students are more likely to experience summer slide. The

32-100 and 110-480 hour levels show that students are less likely to experience summer slide in there reading scores; however, the effect is larger for the 32-100 hour level of dosage. This suggests that, while dosage has an important effect on summer learning loss in reading, students do not need as much time in tutoring to see a positive impact on summer slide for reading as they do for math.

Table 13: Marginal effect of dosage level on probability of experiencing summer slide

	Math	SE	Read	SE
1-10 hours	-0.044***	0.012	0.080***	0.016
12-30 hours	0.111***	0.025	0.154***	0.036
32-100 hours	-0.011	0.011	-0.263***	0.087
110-480 hours	-0.115***	0.029	-0.091***	0.030

****=significant at the 0.01 level, **=significant at the 0.05 level, *=significant at the 0.10 level*
Regressions included all covariates from Tables 5 and 6, not shown here.

Limitations:

Selection on unobservables – Inverse propensity score weighting is a reliable statistical approach in exploring the dosage effect of summer tutoring because it relaxes the limitations of ordinary least squares regression and can strengthen causal inference. However, the downside of using IPSW as a method to control for selection is that is does not account for selection on unobservables. By definition, the variables used in the model to produce a student's probability of receiving a certain amount of tutoring need to be observed and recorded. As such, the model used in this analysis uses common demographic characteristics that are often used in the analysis of educational interventions. Additionally, the model incorporates a variable which accounts for how many parent-teacher conferences family members attended during the school year, acting as a proxy measure for parent/family engagement which has often been conceptualized as an unobservable factor.

So, while this analysis makes as attempt to account for unmeasured factors which could affect the impact of summer tutoring on academic achievement, there are a number of factors that have been left out. Most notably, a measure of student motivation is not included in the analysis. Student motivation is an important factor when looking at the effects of tutoring because it likely has an important impact on selection into tutoring. For example, highly motivated students may select into higher levels of tutoring over the summer because they are motivated to reach higher academic levels. Conversely, we could conjecture that unmotivated students may be selected into tutoring either by parents or school staff because they need to increase their test scores. Since there was no proper proxy for student motivation, it was ultimately left out of the analysis, meaning that it is impossible to distinguish between highly motivated and unmotivated students and how these factors might influence the effect of summer tutoring.

Beyond student-level unobservable factors, there are factors of the tutoring program which have an impact on student outcomes. Most importantly, the quality of tutoring programs can strongly influence success of the intervention. As quality was not accounted for in the dataset, it was not included in this analysis. Because of this, there is no way to distinguish between high- and low-quality tutoring programs or to determine how many students received high-quality or low-quality tutoring. As such, low-quality tutoring programs included in the analysis may be dragging down the overall results of tutoring and vice versa for high-quality programs. Additionally, this complicates the question of tutoring dosage because it is unclear if high-quality programs are more likely to tutor students for longer or for less time over the summer. Also not accounted for in this analysis is the quality of the tutors. Since tutors do not go through the same rigorous certification and training process as tutors and even some traditional youth workers, there is generally a certain level of quality and skill that can be expected from them. For instance, some tutors may struggle to tutor students in some of the higher level math and reading concepts. This is not likely to have too much of an effect on this analysis, though, as most tutors can be expected to tutor basic math and reading concepts without much trouble.

Small sample size – An additional limitation of the analysis conducted for this paper was the relatively small sample size of students who received tutoring over the summer. Given that the ECLS-K is a longitudinal nationally representative dataset, most analyses can expect to utilize large sample sizes. In the case of summer tutoring, though, data was collected only during one round of data collection: the fall of first grade. Coincidentally, this round of data collection was an approximately 30% subsample of the base year respondents, so the pool of potential respondents was greatly reduced. Further, the number of students reported to have received tutoring over the summer was just under 200. Finally, when breaking down students into tutoring dosage categories, the sample size of each category was further reduced, with some categories including fewer than 15 observations. These small sample sizes pose a problem in finding statistical significance, leading to a conclusion of no effect where there may in fact be some effect on summer tutoring. Even though the limited amount of tutoring data collected by ECLS-K does produce small sample sizes, it is important to note that a dataset which collects any information on tutoring services can provide valuable information for analysis of out-of-school time interventions.

Factors affecting dosage of summer tutoring – Beyond methodological limitations, there are some limitations of the scope of this paper. Most notably, this paper does not cover the very real challenges faced by practitioners, youth, and families in reaching adequate levels of dosage. Challenges including availability of travel arrangements, money for programs to stay open all summer, child illness, family vacation plans, and more all affect the amount of tutoring that students can receive over the summer. So, while this paper makes recommendations for how much tutoring students should receive over the summer to see results, it largely ignores the factors that might affect access to and attendance in tutoring. These factors are of very real concern to out-of-school time practitioners but were outside of the score of this analysis.

Discussion and Conclusion:

Tutoring's effect on overall math and reading scores – Overall, almost no effects were found for summer tutoring on fall first grade reading and math scores. The only statistically significant effect found from the base linear model was a negative impact of summer tutoring on reading scores after all covariates in the model were accounted for. As mentioned before, this suggests that, on average, students who receive summer tutoring perform worse than their non-tutored peers in reading. Because demographic confounders were accounted for in this model, it appears that the pure effect of summer tutoring on fall reading scores is negative, meaning that summer tutoring is an ineffective intervention for raising reading achievement. However, there are a number of factors that were not accounted for in this model, including student motivation, which could have an effect on these results. It's easy to imagine a scenario in which, students burnt out and bored from an entire year of schooling, would be unmotivated to practice reading over the summer, which could result in lower test scores in the fall.

Tutoring's effect on proficiency levels – The results regarding summer tutoring's effects on reading and math proficiency levels tell much the same story. No effect was found for tutoring on any of the first three reading and math proficiency levels. This suggests that summer tutoring is not effective in changing achievement not just overall in reading and math, but also in specific skillsets. However, again, these results should not be taken as the final say on summer tutoring's effects on academic achievement. As the discussion on tutoring dosage will show, a more granular exploration of tutoring's effects reveals that there is more to summer tutoring than this basic analysis would suggest.

Dosage effect on overall math and reading scores – The dosage of tutoring matters for academic outcomes. When looking at overall reading scores, it is clear that in order for tutoring to have a positive effect, students must receive at least 32 hours over the summer.

In fact, the IPSW model shows that the 32-100 hour range is optimal for raising reading test scores. So, why isn't even more tutoring better in this case? It could be that the positive effects of tutoring only start to appear around 32 hours and fade away as a student receives more than 100 hours, a pattern which suggests diminishing returns of more tutoring.

As for overall math scores, the opposite appears to be true for the 32-100 hour level of dosage, as the IPSW model estimates a large negative effect on math scores at this level. Further, no other level of dosage is found to have any effect on overall math scores. When one considers the grade level (first grade) of the students in this analysis, it becomes clearer why there is such a marked difference in the dosage effect of tutoring on reading and math achievement. At the kindergarten and early first grade levels of school, instruction is based heavily around building the foundations of literacy and numeracy, both of which require a certain amount of memorization and mechanistic skill development. Numeracy and literacy differ, however, in that numeracy requires more memorization and functional skill development than literacy. Thus, these skills can be harder to retain with less instruction. Additionally, we have to remember that this model does not account for the effects of summer slide, which posits that reading and math scores will drop over the summer months. So, while the model estimates a negative impact of 32-100 hours on overall math scores, we cannot separate this effect from the overall effect of summer slide. All we can say is that it appears that no amount of summer tutoring is "enough" to mitigate the effects of the summer slide.

Dosage effect on proficiency levels – Dosage's effect on mastery of specific reading and math proficiency levels is less clear than its effect on overall reading and math test scores. This is because there does not appear to be a clear pattern of effect on proficiency levels as one moves across dosage categories. For example, the lowest level of dosage shows negative effects on the first two levels of reading proficiency and a positive effect on the third level. Moving into the next dosage category, this pattern is reversed, with positive effects on the first two levels of proficiency and a negative effect on the third level. Again, moving up to the 32-100 hour dosage category, this pattern reverses itself, and again when moving into the highest level of dosage. This pattern suggests that the optimal level of dosage for any particular student is dependent on the particular skills that student needs to improve on. For example, if a student struggles with recognizing and naming the letters of the alphabet, that student would be best served by at least 110 hours of tutoring over the summer.

The pattern with math proficiency levels is clearer than that of reading proficiency levels. As dosage increases, we see a positive effect for the first two levels of math proficiency, though the positive effect reduces from 12-30 hours to 32-100 hours. Additionally, a negative effect is found for these two dosage levels at the third level of math proficiency, though there is a smaller negative effect at the 32-100 hour mark. This, paired with the fact that there is no observed effect of the 110-480 hour category on math proficiency, suggests diminishing returns of tutoring dosage on math achievement. This is also borne out by the above discussion of tutoring and tutoring dosage's effects on overall math scores.

Dosage, proficiency, and subgroups – The starkest difference in predicted probabilities was observed between white and black students for the third level of reading proficiency. Though white and black students different predicted probabilities across proficiency probabilities for both reading and math, the third level of reading showed a concerning amount of difference between the two groups. There may be a number of reasons why black students' predicted probability of mastering the second level of reading was higher than white students and yet their predicted probability of mastering the third level was much lower than white students. For instance, it could be that white students select into more academically rigorous tutoring programs. Alternately, black students may be spending more of their time in tutoring on the lower levels of reading skills than do white students. Whatever the case may be, this analysis suggests that summer tutoring is not an effective intervention for reducing the gap between white and black students.

The story is much different for low-income and high-income students. While this subgroup does not see as marked a difference in predicted probabilities for mastery of any proficiency level at any level of dosage, there are still some differences between the two. Most notably, low-income students are predicted to have a higher probability of mastery of the third reading and math proficiency levels than high-income students, even within the same dosage category. Again, we can conjecture a number of reasons why low-income students might be receiving more instruction in higher-level math and reading skills than high-income students. Perhaps these students are attending higher-quality tutoring programs or are spending more instructional time on high-level skills. Whatever the reason, the implication is that summer tutoring could play an integral role in reducing gaps between students at different income levels. Also, in comparison to the differences between white and black students, this suggests that racial gaps may be more pernicious and hard to close than gaps based on income.

Policy recommendations –Since summer tutoring's effects vary greatly by subject matter and proficiency level, policy implications should be considered based on whether improved reading scores or improved math scores are the goal for a particular population of students. For instance, many policy efforts focused on younger students (K-3) will tend to focus on reading skills whereas policy efforts for older students (4-8) will often focus on math skills. In terms of improving reading scores, this analysis shows that, on average, more tutoring will result in a greater positive impact of mitigating summer learning loss. In response to this, it is important to fund summer literacy programs so that they can run all summer. As for tutoring's effects on math scores, it is not clear that more is better. As such, policy that addresses summer learning interventions for math should focus not on how long a program runs, but on what skills and content are being taught in such programs. So, policy should focus on supporting high-quality, rigorous programs that emphasize instruction on higher levels of proficiency. The gap between white and black students in reading proficiency mastery also presents an opportunity for policy intervention. Policies that emphasize the importance of high quality programs by funding such programs and supporting them in a process of continuous quality improvement can help ensure that all students, no matter their race, attend high quality tutoring programs.

Year-round schooling – Another potential approach to combatting the summer slide that has been gaining popularity is year-round school calendars. In this model of schooling, the school does not take off for a solid three month chunk of time in the summer and rather has smaller breaks spread throughout the year. Students who attend year-round schools still attend school for the same amount of days as their non-year-round peers, they just do not have the long summer break. This would suggest that students who do not have a summer break will not experience summer slide and thus may have higher test scores than students who attend regular school year schools. However, research on this subject does not present a positive case for year-round schooling. One study found that disadvantaged students in California were actually negatively affected by year-round schooling.[29] Another study using a natural experimental design in North Carolina public schools found that the year-round school calendar had no impact on average students or on subgroups of students.[30] Further, there is some evidence that year-round schooling is more an approach to solving school over-crowding and budgetary issues rather than academic achievement gaps.[31] This suggests that, while summer learning loss continues to be a problem, the solution is likely not "more of the same" in terms of school day instruction. Out-of-school time programs like camp, youth development, and tutoring could be an important piece in solving the summer slide.

Further research – There are some areas pertaining to summer tutoring that were outside the scope of this analysis but that also deserve further exploration. For instance, summer tutoring's effects on non-academic outcomes like self-regulation, metacognition, and perseverance could prove to be very illuminating. Research in this area could provide a more holistic view of how tutoring impacts the whole student, not just academic test scores. Additionally, any further research in this area should explore tutoring at different grade levels, as it is likely that different impacts would be observed for students as different stages in their academic careers. While data for tutoring of students at more grade levels was not available for this analysis, future research may wish to invest in collecting such data for multiple age and grade groups. Finally, this analysis only covers a small portion of the tutoring that a student may receive during their time in school. Much more could be learned from an analysis that encompasses tutoring during the school year as well as over the summer. This approach would provide full dosage information rather than a small chunk of tutoring dosage, and it could raise some further insights into the dosage effects of tutoring.

References:

[1] Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review. *Review of Educational Research* , p. 227-268.

[2] Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review. *Review of Educational Research* , 227-268.

[3] Cohen, Peter A., Kulik, James A., & Kulik, Chen-Lin C. (1982) Educational Outcomes of Tutoring: A Meta-Analysis of Findings. *American Educational Research Journal*, Vol. 19, No. 2 p. 237-248

[4] Harvard Family Research Project. (2004). Understanding and Measuring Attendance in Out-of-School Time Programs. *Issues and Opportunities in OST Evaluation*, p. 1-12.

[5] Harvard Family Research Project. (2004). Understanding and Measuring Attendance in Out-of-School Time Programs. *Issues and Opportunities in OST Evaluation*, p. 1-12.

[6] Rowley, R. (2005). *No White Child Left Behind: the Academic Achievement Gap between Blacks and Whites.* Fort Hayes State University, Dept. of Sociology.

[7] Gershenson, S. (2013). Do Summer Time Use Gaps Vary by Socio-economic Status? *American Educational Research Journal*, p. 1219-1248.

[8] APA. (n.d.). Ethnic and Racial Minorities and Socioeconomic Status. Retrieved March 29, 2016, from American Pyschological Association: http://www.apa.org/pi/ses/resources/publications/factsheet-erm.aspx

[9] APA. (n.d.). Ethnic and Racial Minorities and Socioeconomic Status. Retrieved March 29, 2016, from American Pyschological Association: http://www.apa.org/pi/ses/resources/publications/factsheet-erm.aspx

[10] Kerry, T., & Davies, B. (1998). Summer learning loss: The evidence and a possible solution. *Support for Learning* , p. 118-122.

[11] Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review. *Review of Educational Research* , p. 227-268.

[12] Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review. *Review of Educational Research* , p. 227-268.

[13] Keith Zvoch & Joseph J. Stevens (2011) Summer School and Summer Learning: An Examination of the Short- and Longer Term Changes in Student Literacy, *Early Education and Development*, 22:4, 649-675

[14] Gershenson, S. (2013). Do Summer Time Use Gaps Vary by Socio-economic Status? *American Educational Research Journal*, p. 1219-1248.

[15] Gershenson, S. (2013). Do Summer Time Use Gaps Vary by Socio-economic Status? *American Educational Research Journal*, p. 1219-1248.

[16] Mousley, W. (1973). Testing the Summer Learning Loss Argument. *The Phi Delta Kappan*, p. 705.

[17] Harvard Family Research Project. (2004). Understanding and Measuring Attendance in Out-of-School Time Programs. *Issues and Opportunities in OST Evaluation*, p. 1-12.

[18] Harvard Family Research Project. (2004). Understanding and Measuring Attendance in Out-of-School Time Programs. *Issues and Opportunities in OST Evaluation*, p. 1-12.

[19] Penuel, William R., & McGhee, Raymond Jr. (2010). 21st Century Community Learning Centers Descriptive Study of Program Practices. U.S. Department of Education.

[20] Penuel, William R., & McGhee, Raymond Jr. (2010). 21st Century Community Learning Centers Descriptive Study of Program Practices. U.S. Department of Education.

[21] Penuel, William R., & McGhee, Raymond Jr. (2010). 21st Century Community Learning Centers Descriptive Study of Program Practices. U.S. Department of Education.

[22] Wilder Research. (2007). Evaluation of the East Side Learning Center tutoring program.

[23] Foster, E. M. (2003). Propensity Score Matching: An Illustrative Analysis of Dose Response. *Medical Care*, p. 1183-1192.

[24] Foster, E. M. (2003). Propensity Score Matching: An Illustrative Analysis of Dose Response. *Medical Care*, p. 1183-1192.

[25] Arteaga, I., Humpage, S., Reynolds, A. J., & Temple, J. A. (2014). One year of preschool or two: Is it important for adult outcomes. *Economics of Education Review*, p. 221-237.

[26] NCES. (n.d.). Early Childhood Longitudinal Study Program (ECLS). Retrieved March 1, 2016, from National Center for Education Statistics: http://www.nces.ed.gov/ecls/index.asp

[27] NCES. (n.d.). Early Childhood Longitudinal Study Program (ECLS). Retrieved March 1, 2016, from National Center for Education Statistics: http://www.nces.ed.gov/ecls/index.asp

[28] Harvard Family Research Project. (2004). Understanding and Measuring Attendance in Out-of-School Time Programs. *Issues and Opportunities in OST Evaluation*, p. 1-12.

[29] Graves, Jennifer. Effects of year-round schooling on disadvantaged students and the distribution of standardized test performance. *Economics of Education Review,* 30(6), pp. 1281-1305.

[30] McMullen, Steven C. & Rouse, Kathryn E. The Impact of Year-Round Schooling on Academic Achievement: Evidence from Mandatory School Calendar Conversions, *American Economic Journal: Economic Policy*, (4)4, pp. 230-52.

[31] Graves, Jennifer, McMullen, Steven C. & Rouse, Kathryn E. Multi-track Year-round Schooling as a Cost Saving Reform: More than a Matter of Time. Association for Education Finance and Policy.