Peer Effects & Differential Attrition: Evidence from Tennessee's Project STAR

Sanjay Satish

Professor Robert Garlick, Faculty Advisor

Professor Michelle Connolly, Faculty Advisor

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University

Durham, North Carolina

2022

Acknowledgements

This work would not have been completed without the constant encouragement and advice of Professor Robert Garlick. I am deeply indebted to him for his time and teaching as I embarked on this process. I thank him for his guidance and presence in my academic career from helping me write my first paper in his development class to constantly educating me on new modes of research.

I would also like to thank Professor Michelle Connolly for her comments, inspiration, and guidance. This paper would not have been completed without her, help whether it be by introducing me to research as a sophomore in her lab or always being eager to discuss questions and ideas as part of her thesis seminar. I'd like to additionally acknowledge Professor Hugh Macartney, whose Economics of Education class sparked my interest in peer effects and attrition.

Lastly, I am especially thankful for my family and friends who have uplifted me throughout this process. I am grateful to be surrounded by such a supportive community of individuals who have always pushed me to pursue my interests to the fullest extent.

Abstract

This paper explores the effects of attrition on student development in early education. It aims to provide evidence that student departure in elementary schools has educational impacts on the students they leave behind. Utilizing data from Tennessee's Project STAR experiment, this paper aims to expand upon the literature of peer effects, as well as attrition, in public elementary schools. It departs from previous papers by utilizing survival analysis to determine which characteristics of students prolonged participation in the experiment. Clustering analysis is subsequently employed to group departed students to better understand the various channels of attrition present in STAR. It finds that students who left Project STAR were more likely to be of lower income and lower ability than their peers. This paper then uses these findings to estimate the peer effects of attrition on students who remained in the experiment and undertakes a discussion of potential sources of bias in this estimation and their effects on the explanatory power of peer effects estimates.

JEL Classification: I, I21, I26, H4, J13.

Keywords: Attrition, Clustering, Economics of Education, Peer Effects, Project STAR.

1 Introduction

Imagine yourself as an elementary school teacher trying to manage an unruly group of students. Do you seat them across the classroom from each other, hoping that once separated they stop their disruptive behavior? Or does this choice mean that they'll simply find someone else to talk to, resulting in further disruption? Answering these questions may prove instrumental to the success of your students. There exists extensive empirical literature examining the effect of exogenous policy shifts on student achievement. A considerable portion of the current economics of education literature, however, is now centered around the estimation of peer effects: the unobserved, peer-to-peer spillovers from students both inside and outside the classroom. Studying these mechanisms through which student outcomes may be maximized or diminished has significant consequences for policymakers.

It is well documented in the literature that student-to-student peer effects do in fact exist and that such effects are, in part, influenced significantly by peer quality. Such effects are also likely to persist throughout a child's development. The presence of a highly engaged peer in early education may have a lasting impact on a student's achievement, even if said peer were to move or leave after just one year of school. Peer effects are not only dependent on the composition of students in one's class, but also the history of peers a student has interacted with. As such, it is important to consider the role of attrition in educational achievement. If certain peers are provided a channel to depart the public school system, it may have an adverse effect on the students they leave behind. A student who has retained high levels of peer quality may form stronger relationships with those peers than a student at a school where peers depart or drop out frequently. An understanding of which students are likely to leave public schools in addition to understanding the potential effects of such students departures may provide researchers with a better understanding of the drivers of educational attainment.

1.1 Channels of Attrition

This paper proposes that two distinct channels of attrition exist for students of United States Public Schools. One is often comprised of lower income students while the other is often comprised of high-income students. Peers leaving through both channels may have effects on the peers they leave behind.

It well known that private schools provide a channel through which wealthier, often high-ability, students are able to leave the public school system. Many private schools (in addition to magnet and charter schools) have entrance exam requirements which deliberately select for high-achieving students. One form of departure to these schools may come through private schools engaging in tuition rationing, the practice where private schools select for students with high-income backgrounds by charging for seats – disallowing lower income families from consumption without some form of subsidization. At the same time, these schools (especially magnet schools) may engage in cream-skimming, the practice through which these schools select for high-ability students by introducing entrance exam cutoffs as a requirement for admission. In other words, the highest performing students in public schools (i.e. the "cream") is "skimmed" into private schools. Given the existence of such rationing, it is perhaps the case that the act of exiting a school may have unobserved effects on peers in the classroom. That is, if the distribution of students that exit public schools through this channel consists primarily of high-achieving and/or high-income students, there may exist negative or reduced positive peer effects from such withdrawal.

At the same time, it may be the case that lower income students experience higher attrition rates than their higher-income counterparts. As such, these lower income students are less "resilient" in the public school system, often having lower school completion rates than their peers. Though there are many reasons for this including migration, loss of a home, and familial instability, among others, such departures may also have an effect on the students they leave behind. My paper aims to analyze the effects of these two channels of attrition, specifically within the context of elementary education.

1.2 Project STAR

Utilizing evidence from Tennessee's Student/Teacher Achievement Ratio Project (Project STAR) – an experiment in which elementary school students and teachers were randomly assigned to different class sizes – my paper aims to estimate the associations between student attrition on academic achievement between the kindergarten and the 3rd grade. Given that students were randomized into distinct class types within schools, this data allows for the potnetial assumption that student's background characteristics are not correlated within each classroom, allowing for a potential reduction in the bias of peer effects measures that may arise from sorting. Additionally, as part of the experiment, a robust set of student, teacher, and school-level characteristics were collected and students were tracked individually throughout the experiment. My analysis draws upon subsequent data on graduation and dropout rates of the students involved in STAR as well as the identification of students that may have potentially left the public school system in order to better understand the various channels of attrition that are present among students in the experiment.

First, I examine attrition in Project STAR writ large. Through the identification of the proportion of students who either moved across classrooms or schools within the experiment and those that left the experiment entirely, I paint a picture of differential attrition that occurred.

Second, using this identification, I introduce a model of attrition which measures the effect of both student ability and income, controlling for confounds, on the probability a particular student stayed in the experiment. I measure these associations among students in both the kindergarten and 1st grade cohorts in STAR. This model motivates the empirical strategy behind peer effects estimation in the paper.

Third, I consider how attrition across all students who left Project STAR may have impacted students by considering the effects of peer departure in kindergarten and 1st on changes in test scores the following year (1st and 2nd grade respectively). I additionally undertake a discussion of potential sources of bias in peer effects estimation and the consequences of such bias on the interpretation of results.

Finally, I perform an empirical analysis to better understand the attrition channels present within STAR. I utilize a novel clustering method for categorical variables to sort students who departed STAR after kindergarten into "similar" groups. Using the results of this analysis, I discuss potential conclusions that can be drawn regarding the nature of attrition within STAR and the generalizability of such conclusions to public schools writ large.

2 Literature Review

Since the publication of the Coleman Report in 1966[1] there has been suspicion that student unobservables have a significant impact on education production.[2] The classic education production function predicts that test scores can be represented as a function of class size and confounds[3]. Subsequent research has shown that such a relationship is likely identifying only a portion of what affects student achievement[4], especially when viewed in conjunction with the documented relationship between peer effects and achievement. Some, as first described by Lazear (2001), theorize that disruption in the classroom presents a trade-off between learning and class size – a higher proportion of "better behaved students" leads to a larger optimal class size.[5] That is, disruptive students may have a greater negative impact on smaller class-sizes than larger ones. Such negative spillovers from disruption can perhaps be best understood through the lens of peer effects. The channel through which one's own production results in spillovers or externalities on peers achievement is measured by peer effects. Subsequent literature has broadened this model of production to focus both on inputs to production (e.g. class size, curriculum, etc.) as well as inputs that cannot be

[1] A study commissioned by the US government and mandated by the Civil Rights Act of 1964 aimed at measuring educational opportunity in the United States.

[2] Dickinson, E. E. (2016).

[3] Lazear, E. P. (2001), p.777; Examples include: student's own ability, teacher experience, parental income, percentage of students in school on free and reduced lunch, demographic characteristics, etc.

[4] Hanushek (2008).

[5] Lazear, E. P. (2001), p.779.

controlled by policymakers (e.g. parental income, student's attentiveness, peer effects, etc.).[6]

There are many approaches outlined in the literature to estimate peer effects, though the standard model often regresses an individual's own outcomes on the outcomes of peers. As outlined in Sacerdote (2001), there are three main sources of bias in peer effects estimation.[7] The first is that of selection bias, peers may sort into peer groups and disentangling the effect of selection from the peer effect may be difficult. Randomized experiments including Project STAR or random roommate assignments have thus been common datasets used for peer effects estimation as they may potentially eliminate this bias. Second is the reflection problem. That is, the issue that arises when students i and j affect each other simultaneously. This is often corrected for by utilizing an exogenous instrument for peer achievement or behavior (such as parent's educational attainment) though the validity of estimates generated by this process are often hard to verify and are subject to assumptions. The third source of bias results from the inability to separate endogenous (behavioral) effects from exogenous effects, or the effects of peers' background characteristics. Sacerdote notes that controlling for the background characteristics of peers as well as their environments (e.g. the school) may solve this problem.

Although there exists a vast literature on peer effects, papers analyzing the peer effects of attrition are somewhat limited, especially given the lack of longitudinal data of student-tracking across both the private and public school systems. One notable study is that of Dills (2005) which utilized the entry of a magnet school in Virginia to estimate the effect of the exit of high-ability students on the proportion of students scoring in the top quartile on a nationally-standardized exam. Dills finds that the loss of high-achieving students to the magnet school lowers the achievement of their lower-scoring counterparts who remained in public schools.[8] Key limitations of Dills' paper, though, include an inability to track individual students' test scores and a lack of demographic controls. Randomized experiments

[6] Hanushek (2008).
[7] Sacerdote (2001).
[8] Dills (2005).

offer the best data to track peer effects given the elimination of selection bias due to peer sorting.

Within the context of Project STAR, papers such as Boozer (2001) focus on estimating the difference in peer-effects due to variation in class-size. Researchers have found that, especially for small classes, much of the effect of class-size on achievement is actually captured by the increased intensity of spillovers from peers in these settings.[9] Given missing data concerns in Project STAR, Sojurner (2011) develops a model for estimating peer effects with the caveat of missing peer data. Sojourner utilizes pre-assignment measures of student ability in conjunction with random assignment to estimate effects, finding that, on average, the effects of being around high-achieving peers are positive and that peer-ability perhaps matters more for low-achieving students than for high-achieving students.[10] Key limitations of Project STAR include differential attrition and a lack of data on where students who left the experiment attended school after leaving. Rohlfs & Zilora (2014) exploit whether or not a student had public school testing records after the conclusion of Project STAR as a proxy for identifying if a student remained in the public school system. Such developments should assist in the estimation of peer effects of attrition and reduce the limitations the data impose.

With regard to attrition, there exists a vast psychological literature that examines the characteristics of students who drop out of school. Much of the current literature on attrition in schools focuses on students at the high school level; however, the general patterns are likely applicable throughout a child's schooling. It is clear that a student's socioeconomic status and family (both in structure and home environment) are related to attrition.[11] Lower income students from disruptive families are more likely to leave schools than their peers.[12] In addition, lack of school attendance is a risk factor for attrition as well – a quantity which

[9] Boozer (2001).
[10] Sojurner (2011).
[11] Fernandez-Suarez, et al. (2016).
[12] Fernandez-Suarez, et al. (2016) – the reasons for this theory are quite complex, though some believe that students who grow up in families with a lack of rules are often less disciplined and more likely to be affected by peer pressure with regard to substance use.

is tracked in STAR.[13]

Certain papers have found that attrition in public schools is more common than in charter schools. These papers, including an analysis by the New York City Independent Budget Office find that students who stayed in the same school between kindergarten and 3rd grade performed better on standardized tests than peers who switched schools.[14] Special education students in addition to minority students are also found to leave schools at a higher rate than their peers. The effect of socioeconomic status, though, is likely a greater predictor of attrition as it may significantly influence factors related to departure including lack of access to quality schooling, substance abuse, truancy, lack of familial stability, and others.[15]

3 Data

3.1 Project STAR

Project STAR was a class-size randomization experiment across grades K-3 in Tennessee between 1985-1989. 11,600 students across 79 schools voluntarily participated in the study. Students and teachers were randomized into one of three class types at the start of the 1985 school year: small classes (13-17 students), regular-sized classes (22-25 students), and regular-sized classes with a teacher's aide (22-25 students in addition to a full-time teacher's aide). All Tennessee school systems were invited to participate in STAR and costs associated with STAR would be funded by the State. A minimum of 57 students in a school was necessary for participation, allowing for one class of each of the three class types. Most schools stayed in the experiment; however, between grades 1-3 four schools withdrew for

[13] Fernandez-Suarez, et al. (2016)

[14] *Staying or Going? Comparing Student Attrition Rates at Charter Schools with Nearby Traditional Public Schools* (2014).

[15] Stinebrickner & Stinebrickner (2003). It may also be the case that due to the high costs associated with education beyond the elementary school level including school supplies, college entrance exams, etc. that low-income families especially those that are borrowing constrained face are burdened heavily by financial factors in addition to family environment.

several reasons including not being able to maintain the randomization of the experiment.

After being initially assigned to a class type, students were intended to be kept in that class for the duration of the experiment (i.e. 4 years, or between K-3rd grade, although some students were placed across class types). Among participating schools, each was required to have at least one of the three class types and randomization occurred within each school. Study administrators collected a variety of data on students from observable characteristics (e.g. race, age, gender, etc.) of students, teachers, and schools as well as a variety of standardized test scores, administered at the end of each school year. Students who entered a participating school in the middle of the experiment were also randomly assigned to a class type upon entry. There are some limitations to the design, mainly that at the beginning of 1st grade, students in the regular and regular-aide classes were re-randomized across these types, with approximately one half of students in regular classes being reassigned to regular-aide classes. Secondly, approximately 10 percent of students moved across small and regular-sized classes between grades.[16] Thirdly, between grades 1 and 2 in year two of the experiment (1987) 54 second grade teachers were given additional training before assignment to class types. Researchers determined that there was no significant difference between the scores of students with trained versus untrained teachers. Lastly, previous papers have demonstrated evidence of differential attrition from the experiment by class type. I exploit this to determine the severity of the peer effects from attrition based on a student's class type. It is important to note that identification of students who left the experiment, as noted by Rohlfs & Zilora (2014), can only be separated into students who either changed class types, switched to another public school, or left the public school system entirely. Thus, I am unable to accurately identify among the students that left the public school system, who left to a private school. Further information about the experiment is well described in Krueger (1999).

[16] Kruger (1999), p.499.

3.2 Preliminary Observations of Student Ability

The primary educational attainment measure used to track students was the Stanford Achievement Test or "SAT".[17] These scores are reported as item-response-theory scores, which allow for comparison across grades. Starting in the first grade, Basic Skills First or "BSF" tests were administered. These tests corresponded to the state standards in math and reading. The BSF tests standards vary between grades, thus they are unable to be compared across grades for peer effects or attrition estimation.[18] I utilize the SAT scores as my primary measure of student ability due to this limitation.

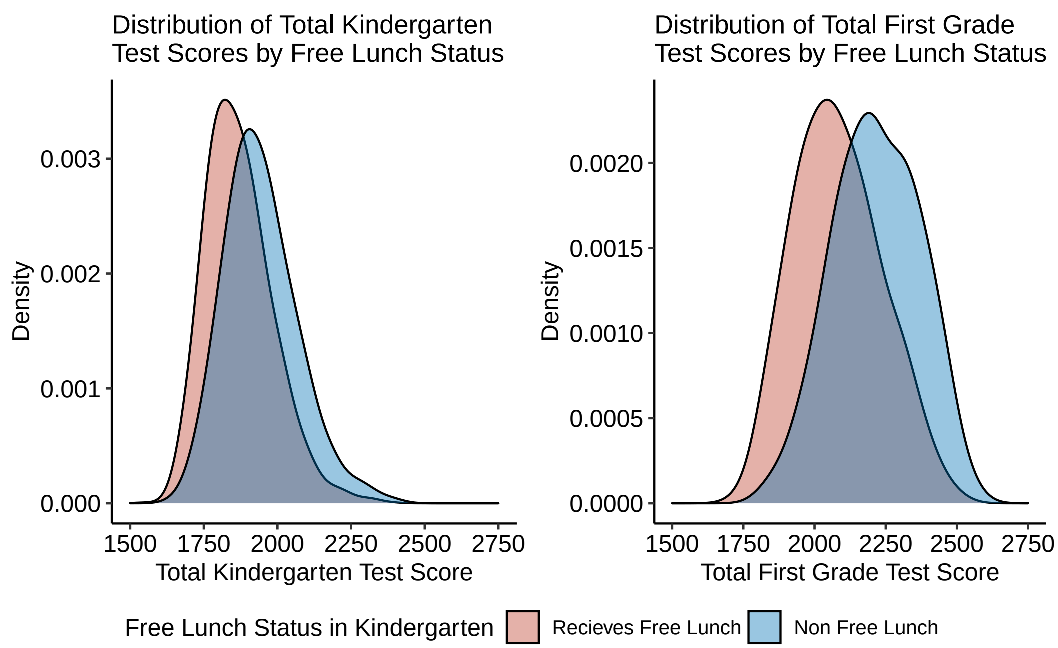


Figure 1: Histogram of ability by free lunch status. Ability defined as the sum of SAT test scores across reading, math, listening, and word skills.

Figure 1 illustrates the distribution of total test score across both kindergarten and 1st grade by income. As shown, it seems that across both grades students who received free lunch had lower mean scores than their non-free lunch counterparts. In particular, among kindergarten students who received free lunch the mean total test score was about 1876

[17]Not to be confused with the Scholastic Achievement Test, an exam taken by many college applicants in the US, also denoted by the acronym SAT.

[18]Kruger (1999) reports a high degree of correlation between the BSF and SAT scores.

whereas among students who did not receive free lunch it was about 1946. Among these two groups the standard deviation of scores was quite close – about 116 and 125 respectively. For 1st grade students the result is similar similar as students who received free lunch had a mean total test score of about 2084 and students who did not receive free lunch had a mean total test score of about 2220. As with kindergarten, the standard deviations in scores are similar - about 153 for both of these groups. In all, it seems that lower income students seem to also be of lower ability.

This conclusion is also illustrated below in Figure 2. Here, we see that among students who departed before 1st grade, the difference in mean test scores is larger – providing evidence for the two channels of attrition that were outlined in the introduction. There appears to be a high-ability and high-income group with a mean test score of 1901 and a lower-ability, lower income group with a mean test score of 1814. As such, we see that it may be the case that the two channels of attrition (private schools and low-income migration) outlined previously are reflected in the data.

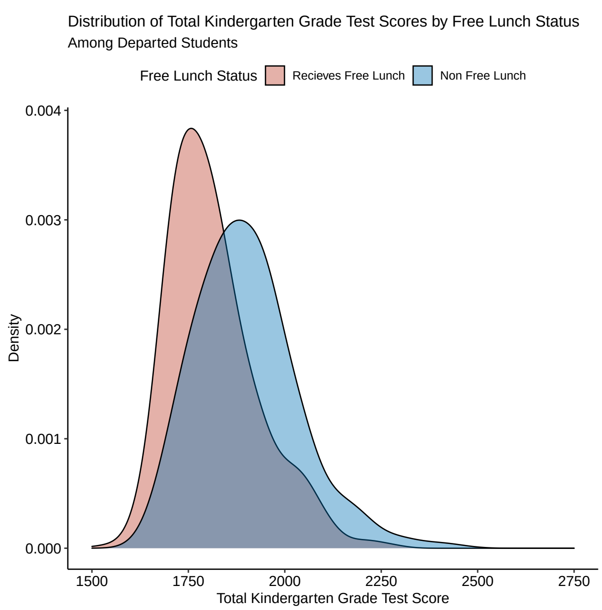


Figure 2: Histogram of ability by free lunch status among students who departed before 1st grade. Ability defined as the sum of SAT test scores across reading, math, listening, and word skills.

3.3 Summary Statistics

Table 1 is a summary table of student, teacher, and school level characteristics for students in grade 1 of Project STAR. As shown, most of the schools at this level are rural, with most students staying in Project STAR for a total of 3 school years. Table 2 is a representation of the same characteristics for students in the kindergarten entry cohort, separating these students by whether or not they left the experiment at the conclusion of the kindergarten school year. Selection criteria for departed students was determined by whether or not the student was "flagged" for being in Project STAR in kindergarten and was subsequently flagged for not being included in the experiment during 1st grade. It is unclear as to whether or not missing data may confound these flags (i.e. if a student may be flagged as not being in the experiment if there was no data collected for that student – even if they were enrolled in a Project STAR classroom). Summary statistics can be compared across both groups depicted in Table 2 to understand the differences between students who left the experiment after kindergarten and those who remained. At first glance, it seems that the students who departed seem to be of lower-ability, are lower income, have a higher probability of being in special education, and are more concentrated in inner cities.

Summary Statistics for Students in 1^{st} Grade

	Small Class (N=1925)	Regular Class (N=2584)	Regular Class with Aide (N=2320)
Experimental Characteristics			
Number of Years in STAR	3.2 ± 1.0	2.9 ± 1.1	3.0 ± 1.0
Number of Years in Small Classes	3.0 ± 1.0	0.2 ± 0.5	0.1 ± 0.3
School Urbanicity			
Inner City	19.8%	22.1%	18.5%
Suburban	23.5%	24.9%	21.2%
Rural	47.3%	45.0%	50.1%
Urban	9.4%	8.0%	10.2%
Teacher Characteristics			
Female	97.5%	100.0%	100.0%
White	81.2%	83.8%	81.6%
Years of Experience	12.2 ± 8.7	10.3 ± 8.7	12.7 ± 9.2
Class Size	15.7 ± 1.6	22.7 ± 2.3	23.4 ± 2.4
Receives Free Lunch	47.8%	51.9%	50.3%
Special Education	0.6%	1.5%	1.5%
Math SAT Scaled Score	538.7 ± 44.1	525.3 ± 41.7	529.6 ± 42.9
Reading SAT Scaled Score	530.0 ± 56.6	513.5 ± 53.5	521.3 ± 54.7
Listening SAT Scaled Score	572.7 ± 34.5	563.8 ± 32.4	567.2 ± 33.9
Word Study Skills SAT Scaled Score	523.0 ± 52.6	506.2 ± 54.0	513.7 ± 51.8

Table 1: Summary table for students in grade 1 of Project STAR.
Note: Some variables expressed as Mean ± Standard Deviation. Per the data, Inner-city schools are defined as those with more than half of students on free or reduced lunch. Schools in outlying areas of metropolitan areas are considered suburban. Urban schools are those in towns with population over 2500. All other schools are considered rural. Test score data is measured at the end of each year, on testing dates specified by the state.

Summary Statistics for Students in Kindergarten

	Students Who Departed After Kindergarten			**Students Who Stayed to 1st Grade**		
	Small Class (N=453)	Regular Class (N=603)	Regular-Aide (N=580)	Small Class (N=1303)	Regular Class (N=1425)	Regular-Aide (N=1490)
Experimental Characteristics						
Number of Years in STAR	1.1 ± 0.4	1.1 ± 0.5	1.1 ± 0.5	3.5 ± 0.8	3.5 ± 0.8	3.5 ± 0.8
Number of Years in Small Classes	1.0 ± 0.2	0.0 ± 0.2	0.0 ± 0.1	3.3 ± 1.0	0.3 ± 0.8	0.3 ± 0.7
Kindergarten School Urbanicity						
Inner City	28.3%	33.5%	30.7%	18.5%	18.0%	20.3%
Suburban	32.2%	27.4%	30.7%	21.0%	17.1%	19.7%
Rural	30.0%	31.3%	31.6%	51.0%	57.1%	49.3%
Urban	9.5%	7.8%	7.1%	9.4%	7.8%	10.7%
Kindergarten Teacher Characteristics						
Female	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
White	84.5%	71.5%	82.4%	86.6%	83.2%	85.4%
Years of Experience	8.5 ± 5.7	8.9 ± 5.9	9.8 ± 5.9	9.0±5.8	9.1±5.7	9.8±5.8
Kindergarten Class Size	15.2 ± 1.4	22.6 ± 2.1	22.9 ± 2.3	15.1 ± 1.5	22.2 ± 2.2	22.6 ± 2.3
Receives Free Lunch	54.1%	56.7%	57.8%	44.3%	43.2%	45.8%
Special Education	5.7%	3.8%	4.5%	3.1%	2.8%	2.2%
Math SAT Scaled Score	473.5 ± 51.1	467.8 ± 48.5	469.0 ± 46.8	497.1 ± 47.5	489.8 ± 45.8	488.4 ± 44.1
Reading SAT Scaled Score	431.4 ± 33.0	425.3 ± 29.5	427.3 ± 29.7	443.7 ± 31.6	438.8 ± 30.7	438.7 ± 31.7
Listening SAT Scaled Score	531.7 ± 34.7	527.6 ± 35.3	527.5 ± 33.2	542.9 ± 32.2	541.6 ± 31.6	538.9 ± 32.2
Word Study Skills SAT Scaled Score	427.9 ± 35.7	422.0 ± 33.9	425.0 ± 32.8	441.7 ± 37.3	436.2 ± 35.9	436.3 ± 37.9

Table 2: Combined summary table for students who were in Project STAR in Kindergarten but left the experiment before entering grade 1 and who were in Project STAR in kindergarten and continued in the experiment in grade 1..
Note: some variables expressed as Mean ± Standard Deviation. Values correspond to kindergarten-grade information. Does not account for students who may have re-entered the experiment in later years or distinguish between students by attrition category. Variable definitions are consistent with Table 1.

3.4 Characteristics of Attrition in STAR

Since the completion of Project STAR, subsequent studies have been commissioned to analyze the long-run impacts of the experiment. As part of these studies, data is collected on the academic achievement of students who participated in Project STAR in grades 4 through 8. At the same time, students within the experiment were tracked. If a student moved within the experiment to another Project STAR school, that student's movement can be identified through the data. Utilizing this data, Rohlfs & Zilora (2014) identify five types of attrition within Project STAR:

1. Students who were in a Project STAR school at time $t - 1$, but whose school left the experiment at time t.

2. Students who changed class type within a school.

3. Students who left to another Project STAR school.

4. Students who left to another public school.

5. Students who left the public school system (could be characterized by death, moving to a private school, moving out of state, or test absence).

Identifying students who fall into attrition types 1-4 is easily seen through the data. Both students and schools are flagged for leaving the experiment. Though there may be some missing data concerns, subsequent identification of students who fall into categories 1 and 3 can be undertaken by examining whether or not there exists Project STAR testing data for these experiments both within the experiment's timeframe (i.e. grades K-3) or in the subsequent studies tracking the students post-experiment. Rohlfs & Zilora (2014) identify students who left the public school system as those who left Project STAR in 3rd grade and test scores do not exist for them in 4th grade. It is likely the case that if a student left Project STAR for a private school, they would not re-enter the public school system until after the 5th grade. This is due to the fact that private schools in certain geographic regions may only accept students between grades K-5. Thus, if a student left for a private school, it is likely that they do not have test score data for both grades 4 and 5. Another potential point of identification is to see which of these students perhaps re-entered the public school system in grades 6-8 (typically denoted as "middle school"). Students with access to only K-5 private schools may have perhaps been forced to re-enter the public school system for middle school. In this paper, I measure the differences in attrition across class types, providing the opportunity to further identify the characteristics of students, namely ability, that are perhaps correlated with attrition.

4 Survival Analysis of Attrition

4.1 Empirical Framework

To quantify the effect of student ability on attrition, I turn to survival analysis. Within the context of Project STAR, attrition can be considered a survival observation of interest. That is, the act of a student leaving the experiment entirely (i.e. they fall into attrition categories 1,4, and 5) can be considered a "failure." Survival analysis allows for the identification of factors which may be correlated with the amount of time a student stayed in the experiment (in other words, their time to failure). Here, I represent $T_{i,g,c,s}$ as the failure time (i.e. the number of years a student stayed in the experiment until they left) for each student i in grade g in classroom c at school s. It may be the case that due to missing data concerns or students leaving the experiment in one year and coming back in another, $T_{i,g,c,s}$ is unobserved (or, perhaps, not "completely" observed). Such observations are considered to be censored, with censoring time C_i. I model the data under the assumption that $T_i > C_i$ (i.e. the "survival time" is longer than the "censoring time"). This assumption is consistent with empirical evidence of certain students leaving and returning to the experiment. I model the survival time through the following accelerated failure time (AFT) model:

$$(1) \qquad log(T_i) = \beta_0 + \beta_1(TS_{i,g,c,s}) + \beta_2(CT_{i,g,c,s}) + \beta_4(FL_{i,g,c,s}) + \beta_5(X_{i,g,c,s}) + \epsilon_{i,g,c,s}$$

Where $TS_{i,g,c,s}$ is the test score for student i in grade g in classroom c in school s, $CT_{i,g,c,s}$ is a categorical variable for the class type student i was enrolled in upon entry to the experiment (e.g. small, regular-sized, or regular sized with teachers aide), $FL_{i,g,c,s}$ is a categorical variable for the free-lunch status of student i in grade g, and $X_{i,g,c,s}$ is a vector of controls for each student including observable student, teacher, and school data. Examples of this include student attendance, teacher's highest education level, student's race, percentage of school on free and reduced lunch, school geography (e.g. urban, rural, etc.), and others. The coefficients of this model have a multiplicative effect on time. For example, if $\beta_1 = 1$, then $exp(\beta_1) \approx 2.718$. Holding all other variables constant, an individual with $TS_{i,g,c,s}$ one unit greater than another would be expected to stay in the Project STAR experiment 2.718 times longer than the other. In other words, the probability this individual "survived" to time $2.718t$ would be the same as the probability that another individual has "survived" to time t.

The AFT model is a fully parametric model. Thus, I must specify a probability distribution for $log(T_i)$, or a distribution of attrition probabilities throughout the experiment. Based on the nature of the experiment, attrition was more likely in early years (i.e. kindergarten and 1st grade) than later on in the experiment, as schools who were unable to continue randomization dropped out, or students who were unhappy with their class assignment left STAR. Thus, it is reasonable to model the attrition probabilities in the experiment as monotonically decreasing with time. The Weibull distribution allows for the estimation of monotonically increasing, decreasing, or constant hazards and has relevant analytical advantages in the presence of right-censoring. Given my assumptions regarding the attrition probabilities over time, I represent the hazard function ($\lambda(t)$) of $T_{i,g,c,s}$ using the Weibull distribution:

$$\lambda(t) = p\lambda^p t^{p-1}$$

Here, p is the scale parameter and a value of p less than 1 indicates that attrition rates decrease over time in STAR. An illustration of a Weibull distribution with p less than 1 (monotonically decreasing) is depicted below:

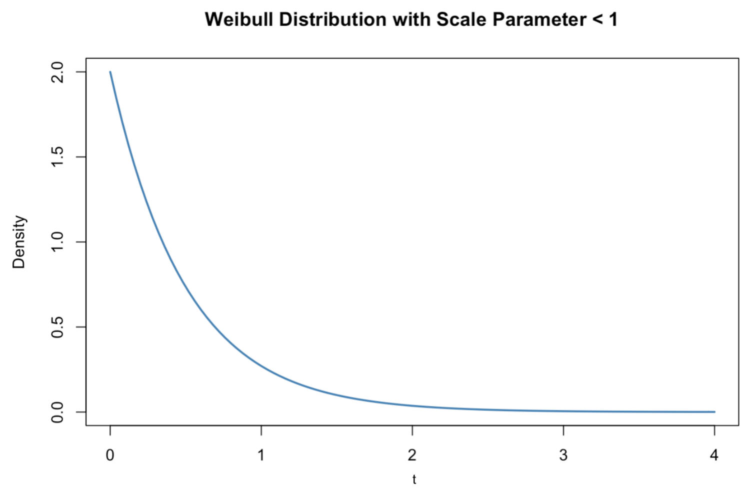


Figure 3: Illustration of monotonically decreasing Weibull Distribution with shape parameter $\lambda = 1$ and scale parameter $p = 0.5$.

4.2 Results - Survival Analysis of Attrition

4.2.1 Preliminary Survival Estimates – Kindergarten Entry Cohort

Kaplan Meier estimates of survival probabilities for students in the Kindergarten Entry Cohort of Project STAR are depicted in the ensuing four figures. In each of these plots, the dependent variable is "time until attrition" (i.e. Years in star until attrition). Here, students who left are first identified by whether they spent the full four years in STAR (K-3 grade) and subsequently identified by their first exit from the experiment. There are a number of students who leave and come back into the experiment and are re-randomized upon entry. I consider consider these students right censored ($T_i > C_i$). That is, after their first time leaving the experiment, they are considered "left" even though they may have re-entered at a later date. This assumption allows for better parameterization of survival estimates and better represents the monotonically decreasing nature of attrition in STAR. Other papers in the literature make use of "composite duration" and "composite class type" variables which subjectively sort STAR students into different class types and re-calculate years in the experiment. I believe that these measures, although useful for assessing randomization conditions in the experiment, bias estimates of attrition due to the classification of students who re-enter the experiment multiple times as having never been in the experiment. In addition, I remove students who leave STAR due to the fact that their school left the experiment. Thus, Figures 1-4 are illustrations of what I consider "unforced" attrition (i.e. it was the student's or student's guardian's choice to leave STAR)

These initial estimates depict three interesting patterns. First, with regard to attrition based on class type - Figure 4 shows that differential attrition based on class type is likely overstated in the literature. Although there is a statistically significant difference in attrition probabilities across class types at the 5% level, it seems that the effect of being in a regular or regular-aide class does not increase the probability of leaving the experiment drastically compared to being in a small class. It is plausible that this result is due to the fact that students who leave and re-enter the experiment more than once were preferentially re-assigned to smaller classes, thereby providing them an incentive to re-join the experiment. These effects of increased retention in smaller classes are muted in the figure, which displays student's "first" chance to leave the experiment.

Second, it is clear that lower income, lower ability students (as shown in Figures 1, 2, and 3) leave the experiment much faster than their higher-income, higher-ability counterparts. This is shown both by the significant reduction in survival probability for students who are in the bottom 25th percentile of test scores or on free-lunch. This provides evidence that the main channel for attrition in public schools may not be students moving to private schools, but rather drop-outs or significant migration of low-income families. These effects seem independent of class size, potentially leading to the conclusion that such attrition is constant regardless of the random assignment of the student to a class – leading non-experimental forces to be contributing to this attrition.

Third, the attrition probabilities seem to be monotonically decreasing across the course of the experiment. As shown in Figures 1-4, most of the attrition occurs in the first year of the experiment, with less and less students leaving as time goes on. This is consistent with the re-randomization and re-entry that occurred between grades K-1. As well, it seems there is more variation in the students who left in this first year than otherwise. Such results justify the use of a Weibull distribution to parametric the survival estimates, which allows for monotonically decreasing attrition probabilities. This result is elaborated upon in the following section and confirmed through the distributional convergence chart located in the Appendix.

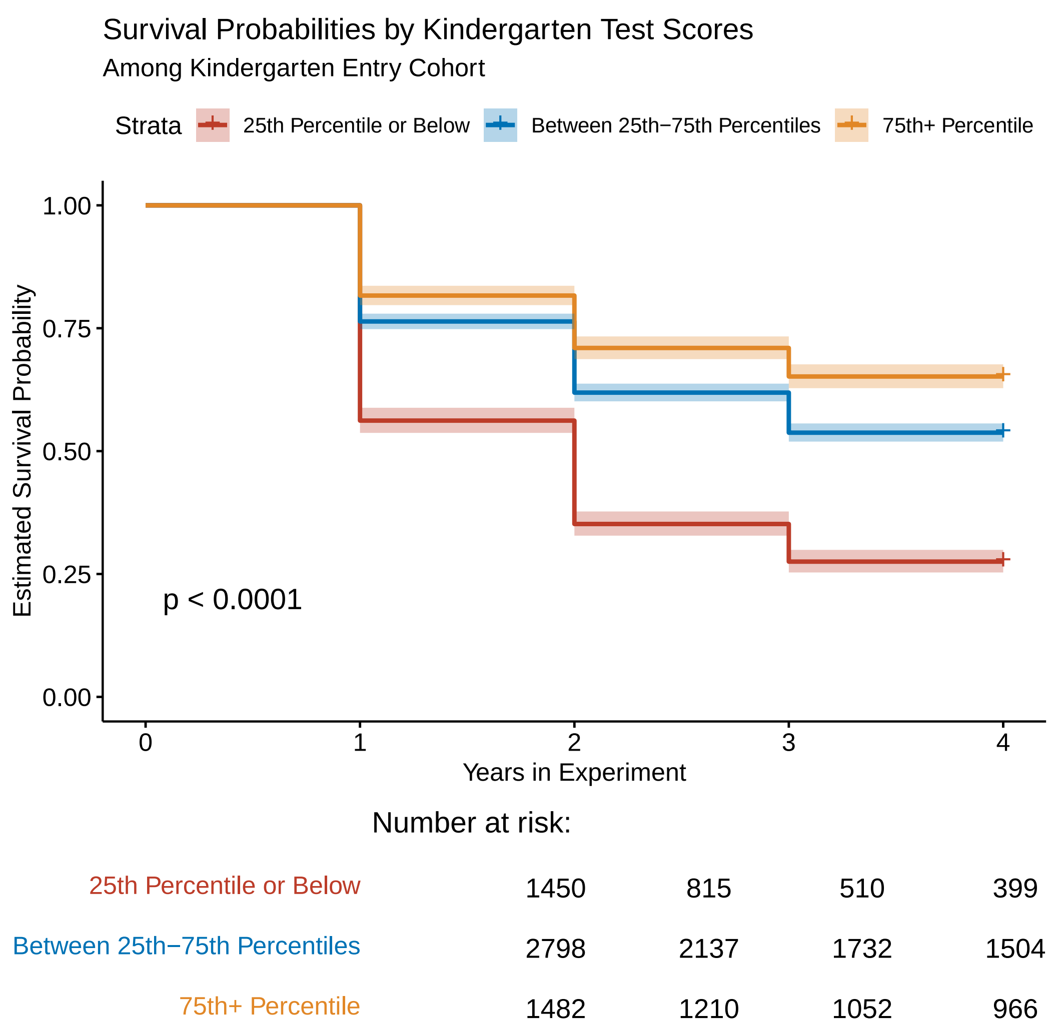


Figure 4: Test scores calculated as the sum of kindergarten math, word skills, listening, and reading SAT scaled scores. Percentiles calculated across all students in the kindergarten entry cohort. Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals.

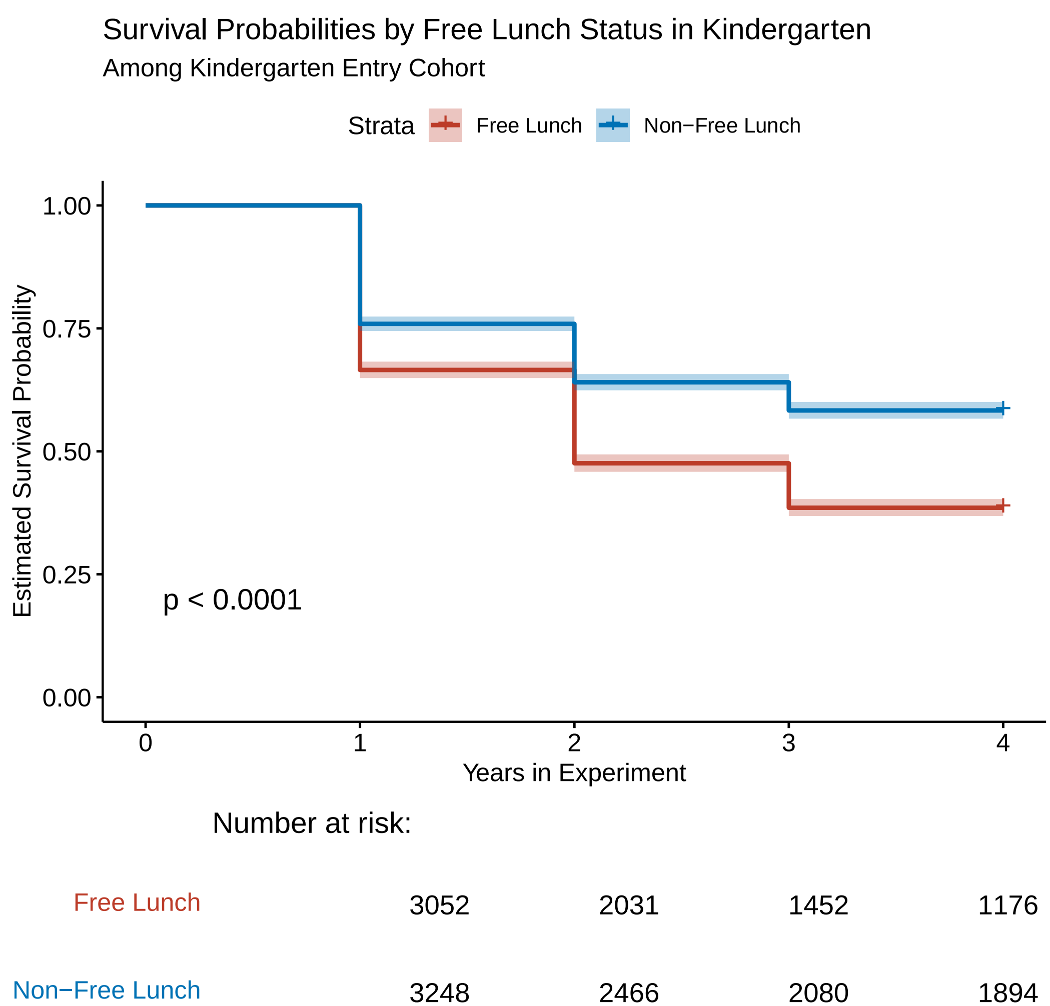
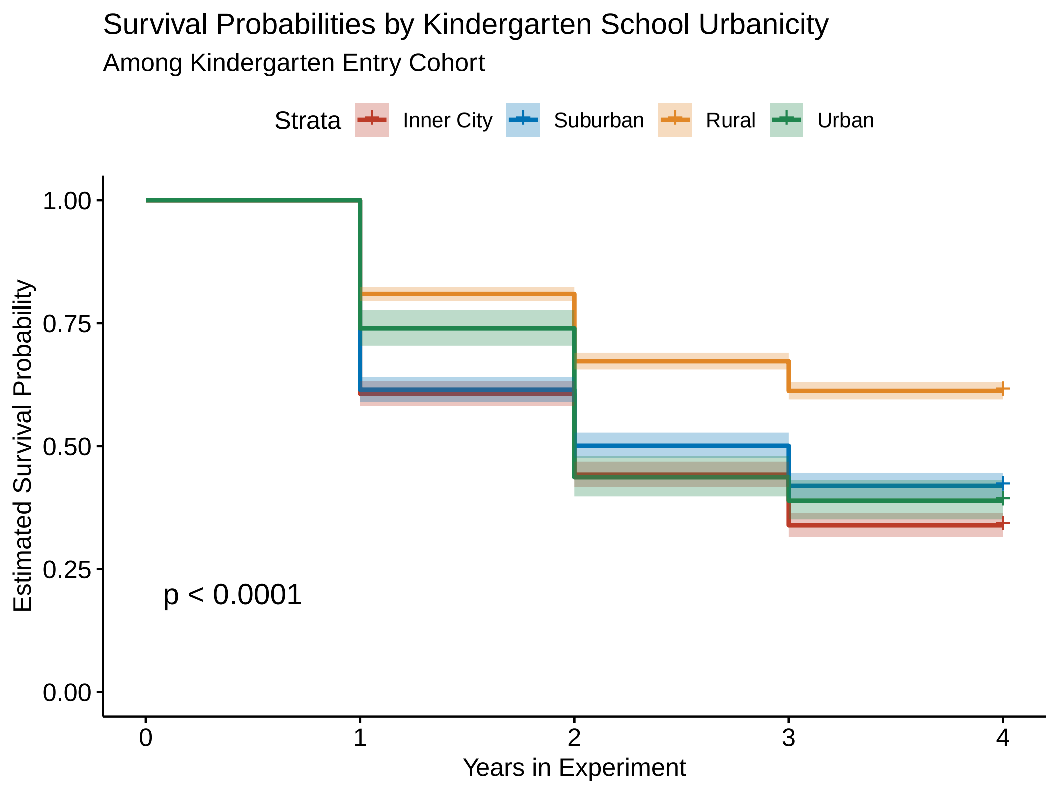


Figure 5: Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals.



Number at risk:

Inner City	1428	866	631	484
Suburban	1412	868	707	592
Rural	2917	2361	1962	1786
Urban	568	420	248	221

Figure 6: Students may have switched to a different school during their final year in STAR. Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals.

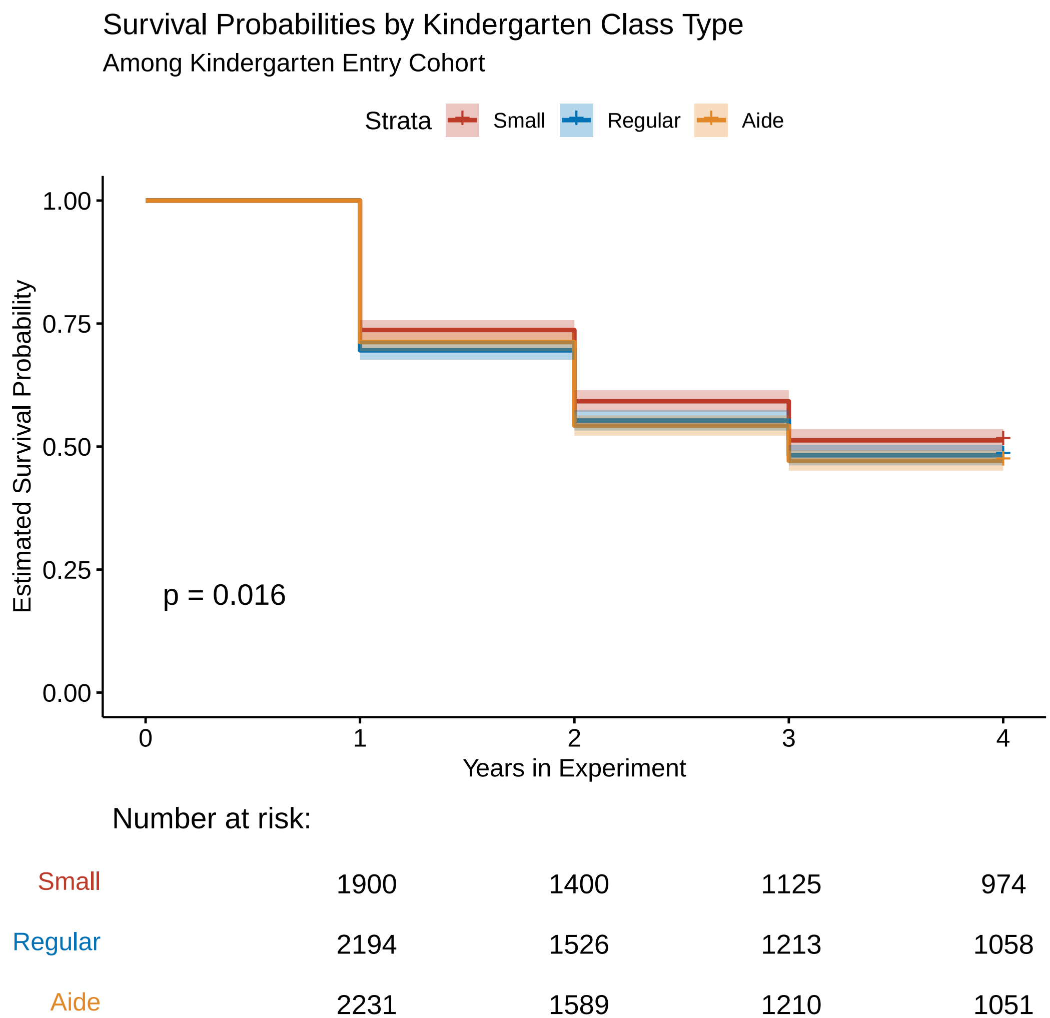


Figure 7: Students may have switched to a different class type during their final year in STAR. Survival probabilities based on non-parametric Kaplan Meier estimates. P-value for difference in survival probabilities across groups calculated using log-rank test. Risk table shows number of students in STAR at each time interval. Highlighted regions represent confidence intervals.

4.2.2 Survival Analysis of Attrition - Grades K to 1

The below table shows regression results for the AFT model of student attrition in STAR among the Kindergarten Entry Cohort. Specification 1 is fitted on all students in the kindergarten cohort, while specification 2 is fitted on the subset that stayed beyond 1st grade (hence they have the opportunity to switch schools or class types). Specification 1 allows us to see the total effects of attrition across all years in the experiment and specification 2 aims to reduce the effects of attrition that arose from re-randomization or school constraints after the first year of the experiment.

Similar to the survival curves, I decline to fit this model on any students whose schools dropped out of the experiment, once again focusing on "unforced" attrition. Additionally, I drop observations where a student in the kindergarten cohort switched to a different school in 1st grade and came back to the same kindergarten school in 2nd grade. Such students would not be considered "departed" in my experiment; however their school-to-school movement might bias my results. In total, there are 7 students who fall into this category, a very small amount. I estimate that less than .15% of total observations exhibit a similar phenomenon (multiple school switches back to original school) across all years of the experiment.

The results from specification 1 are consistent with those found in the non-parametric estimates in the previous section. Specifically, students who are not on free lunch are expected to survive $e^{0.287} \approx 1.332$ times longer than students who are on free lunch, holding all else constant. Students in rural schools are expected to survive $e^{0..341} \approx 1.406$ times longer than students in inner-city schools and similarly students in urban schools are expected to survive $e^{0.141} \approx 1.151$ times longer than students in inner-city schools, holding all else constant. With regard to ability, I find that students with kindergarten test scores between the 25th and 75th percentiles of all students in the kindergarten cohort are expected to survive $e^{0.506} \approx 1.657$ times longer than those in the 25th percentile or lower, holding all else constant. Similar results are observed for students in the 75th percentile or higher who are expected to survive $e^{0.741} \approx 2.098$ times longer than those in the 25th percentile or lower, holding all else constant.

The results for both income and ability are also observed in specification two, although with lower effect sizes. Across both specifications, these effects are significant at the 1% level. It is also important to note that I do not find significant effects from class size on attrition among this cohort. Diagnostic plots in the appendix show that the assumption of a Weibull distribution to model attrition is well-founded. Attrition appears to monotonically decrease as the experiment goes on and the residuals appear to converge.

It seems that there is a strong association between being of low-ability and of low-income and attrition in Project STAR. Thus, these preliminary estimates would lead me to conclude that much of the attrition is taking place through the lower income channel I described previously rather than to private or charter schools among the Kindergarten Entry Cohort.

Table 3: Coefficient-Level Estimates for AFT Model of Attrition for Kindergarten Entry Cohort

	Dependent variable: Years Until First Exit	
	All Students in K-Cohort	Students Who Stayed Past First Grade
	(1)	(2)
Non-Free Lunch	0.287***	0.249***
	(0.037)	(0.035)
School Suburban	-0.069	-0.066
	(0.079)	(0.077)
School Rural	0.341***	0.103
	(0.082)	(0.080)
School Urban	0.141*	-0.060
	(0.086)	(0.084)
Switched Schools		0.097
		(0.066)
Regular Class	-0.048	-0.016
	(0.036)	(0.038)
Regular Class w/ Aide	-0.030	-0.025
	(0.036)	(0.036)
Switched Class Types		0.024
		(0.031
Kindergarten Test Score Between 25th-75th Percentile	0.506***	0.307***
	(0.034)	(0.032)
Kindergarten Test Score Above 75th Percentile	0.741***	0.482***
	(0.045)	(0.044)
Not Special Education	-0.022	-0.020
	(0.080)	(0.080)
Not Pulled Out for Special Instruction	0.189***	0.150**
	(0.067)	(0.063)
Constant	0.544***	1.299***
	(0.161)	(0.153)
Controls	Yes	Yes
Parametric Form	Weibull	Weibull
Observations	5,504	3,970
Log Likelihood	-6,816.844	-3,625.752
χ^2	1,163.570*** (df = 26)	456.399*** (df = 27)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Controls include: Student Gender, Student Race, Kindergarten Teacher Experience (yrs.), Days Absent in Kindergarten, Kindergarten School Grade Range,% Students in Kindergarten School Receiving Free Lunch, % Students in Kindergarten School Bused. Student characteristics determined based on kindergarten information unless otherwise specified. Coefficient estimates for controls reported in Appendix.

4.2.3 Survival Analysis of Attrition - Grades 1 to 2

The below table shows regression results for the AFT model of student attrition in STAR among all students in 1st grade. This includes new students who entered in the 1st grade cohort as well as those that stayed from kindergarten. Similar to the previous section, specification 1 is fitted on all students in 1st grade, while specification 2 is fitted on those who stayed for at least 1 year. I drop observations where students switch back and forth between schools during the experiment.

Among this cohort, the results are mixed. With regard to ability, the results from specification 1 are consistent with those in the previous section. Students above the 25th percentile are expected to stay in the experiment longer than those below the 25th percentile based on 1st grade test scores. These effects are insignificant among those who stayed past 2nd grade. This is potentially due to the fact that attrition in the final year of the experiment was quite small, with little variation among students that left.

It is important to note that in specification 1, I find that students in inner-city schools are expected to survive longer than those in their suburban, rural, or urban counterparts. This is consistent with the finding that being on free lunch is no longer a significant factor in attrition, with it in specification 2 actually contributing to an increase in survival time. Class size effects are also significant in both specifications, with students in regular-aide classes more likely to survive than their small-class counterparts.

It is most likely the case that attrition among the kindergarten students was more pronounced along income lines; however, students with low-ability were less likely to stay in the experiment, regardless of their entry cohort. This pattern of attrition for low-ability and low-income students most likely began to wane as the experiment went on. One potential explanation for this is that within Project STAR, about 6% of students in 1st grade were recommended not to be promoted to grade 2. This percentage dropped to about 2% in both 2nd and 3rd grade. A primary driver of attrition may have been the holding back of low performing students, the number of which diminished as the experiment went on.

Table 4: Coefficient-Level Estimates for AFT Model of Attrition for 1st Grade Entry Cohort

	Dependent variable: Years Until First Exit	
	All Students in 1st Grade	Students Who Stayed Past 2nd Grade
	(1)	(2)
Non-Free Lunch	0.010	-0.004**
	(0.013)	(0.002)
School Suburban	-0.081***	-0.001
	(0.029)	(0.005)
School Rural	-0.067**	-0.003
	(0.031)	(0.005)
School Urban	-0.075**	-0.001
	(0.032)	(0.005)
Switched Schools		0.001
		(0.005)
Regular Class	0.017	0.005**
	(0.013)	(0.002)
Regular Class w/ Aide	0.044***	0.007***
	(0.014)	(0.002)
Switched Class Types		0.004
		(0.003)
1st Grade Test Score Between 25th-75th Percentile	0.187***	0.003
	(0.014)	(0.003)
1st Grade Test Score Above 75th Percentile	0.188***	0.001
	(0.017)	(0.003)
Not Special Education	-0.144	-0.023
	(0.102)	(0.019)
Not Pulled Out for Special Instruction	-0.021	-0.009***
	(0.016)	(0.003)
Constant	1.132***	1.145***
	(0.115)	(0.021)
Controls	Yes	Yes
Parametric Form	Weibull	Weibull
Observations	5,354	4,077
Log Likelihood	-6,478.014	-326.741
χ^2	299.507*** (df = 25)	66.860*** (df = 27)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Controls include: Student Gender, Student Race, 1st Grade Teacher Experience (yrs.), Days Absent in 1st Grade, 1st Grade School Grade Range, % Students in 1st Grade School Receiving Free Lunch, % Students in 1st Grade School Bused. Student characteristics determined based on 1st Grade information unless otherwise specified. Coefficient estimates for controls reported in Appendix.

5 Identification of Peer Effects From Attrition

5.1 Empirical Framework

Exploiting the panel nature of Project STAR data, I use a fixed effects model to determine the peer effects from attrition across all students who left Project STAR, controlling for class size. The dependent variable of interest is a measure of a student's test score, given by the testing data in Project STAR. I make use of lagged covariates to measure the effect of attrition in time $t - 1$ on test scores of remaining students in time t. Essentially, I am estimating the strength of the relationship between the test scores of student i and the background characteristics of student i's peers, allowing for the possibility that such a relationship may not exist. For example, student i's learning is likely to be improved if student j's test scores are high. I include school fixed effects to account for unobserved differences across schools. There are likely to be school-specific, time-invariant effects which this fixed effect picks up on (e.g. location, funding characteristics, etc.) which should aid in the robustness of my results.

As explained in the literature review, Sacerdote's 2001 paper, *Peer Effects with Random Assignment: Results for Dartmouth Roommates*, describes many of the issues that arise when estimating peer effects. Among these, one particular issue within the context of my analysis is that of the reflection problem. Student i's contemporaneous test score may have influenced student j's contemporaneous test score. Thus, I am unable to make use of contemporaneous measures of ability on the right hand side of my equation. To eliminate potential bias that may arise from this problem, Sacerdote lays out a reduced form model for peer effects estimation which I adapt to include measures of departed peer ability.

5.1.1 Case for Two Students i, j

The model for two students can be described as follows:

$$TS_i^t = \delta + \alpha(TS_i^{t-1} + \mu_i) + \beta(TS_j^{t-1} + \mu_j) + \gamma(TS_j^t) + \epsilon_i \tag{2}$$

$$TS_j^t = \delta + \alpha(TS_j^{t-1} + \mu_j) + \beta(TS_i^{t-1} + \mu_i) + \gamma(TS_i^t) + \epsilon_j \tag{3}$$

Here, TS^t is the contemporaneous test score, TS^{t-1} is the lagged test score and μ represents the measurement error that arises from the inability to directly observe true ability.[19] We also assume that μ is uncorrelated with TS^t, otherwise the reduced-form estimates will be biased due to classical measurement error. Substituting equation (3) into equation (2) allows us to find the following reduced form equation:

$$TS_i^t = \frac{1}{1-\gamma^2}[(1+\gamma)\delta + (\alpha+\gamma\beta)TS_i^{t-1} + (\beta+\alpha\gamma)TS_j^{t-1} + (\alpha+\gamma\beta)\mu_i + (\beta+\alpha\gamma)\mu_j + \gamma\epsilon_j + \epsilon_i] \tag{4}$$

Sacerdote simplifies this expression as the following, where π_0, π_1, π_2 are the reduced form coefficients and η is the error term:

$$TS_i^t = \pi_0, +\pi_1 TS_i^{t-1} + \pi_2 TS_j^{t-1} + \eta \tag{5}$$

Here, we know that π_2 is not influenced by selection bias due to the randomization of students in Project STAR and we would interpret coefficients π_1 and π_2 as the total effect of observed peer and own lagged test scores on own contemporaneous test scores. We consider the full set of parameters (δ, α, β, and γ) under-identified. However, under the assumption that there are no unobserved background characteristics of students i and j that have an effect on TS^t we may observe the following: if $\pi_2 \neq 0, \beta + \alpha * \gamma \neq 0$ and thus $\beta \neq 0$ or $\gamma \neq 0$. As such, testing the estimate of the coefficient π_2 is a valid test for the presence of peer effects.[20]

[19] Sacerdote describes measures of ability including test scores as "noisy."

[20] If this model were fully identified and we were able to solve for δ, α, β, and γ, β and γ can be interpreted as casual estimates of the effect of lagged and contemporaneous peer test scores on own academic achievement.

5.1.2 Case for Three Students i, j, k Where k Departs

Expanding Sacerdote's model to include the effect of departed peer, I describe the following:

$$(6) \qquad TS_i^t = \delta + \alpha(TS_i^{t-1} + \mu_i) + \beta(TS_j^{t-1} + \mu_j) + \theta(TS_k^{t-1} + \mu_k) + \gamma(TS_j^t) + \gamma'(TS_k^t) + \epsilon_i$$

$$(7) \qquad TS_j^t = \delta + \alpha(TS_j^{t-1} + \mu_j) + \beta(TS_i^{t-1} + \mu_i) + \theta(TS_k^{t-1} + \mu_k) + \gamma(TS_i^t) + \gamma'(TS_k^t) + \epsilon_i$$

Here, since student k has departed by time t we not only find TS_k^t is unobserved, but assume that $\gamma' = 0$ due to the fact that departed peers who have left STAR have no effect on their peers who remained in the experiment. As such, we can once again substitute equation (7) into equation (6) to yield a reduced form model including the effect of a departed peer:

$$(8) \qquad TS_i^t = \frac{1}{1-\gamma^2}[(1+\gamma)\delta + (\alpha+\gamma\beta)TS_i^{t-1} + (\beta+\alpha\gamma)TS_j^{t-1} + (\theta+\theta\gamma)TS_k^{t-1} + (\alpha+\gamma\beta)\mu_i + (\beta+\alpha\gamma)\mu_j + (\theta+\theta\gamma)\mu_k + \gamma\epsilon_j + \epsilon_i]$$

Which we can simplify as:

$$(9) \qquad TS_i^t = \pi_0, +\pi_1 TS_i^{t-1} + \pi_2 TS_j^{t-1} + \pi_3 TS_k^{t-1} + \eta$$

In this equation, I assume that the departed peer k is not identical to the peers i, j which remain in STAR.[21] Similar to the previous case, we find the set of structural parameters δ, α, β, γ, and θ to be under-identified.[22] However, by testing the estimate of π_3, we may find

[21] Relaxing this assumption would mean that θ is equivalent to β which would result in the representation of the effect of peer's lagged achievement on own test scores as a linear combination of these two quantities (i.e. $\beta(TS_j^{t-1} + TS_k^{t-1})$).

[22] If this were fully identified, we could solve for θ which would reflect the causal effect of departed peer ability on contemporaneous peer ability, assuming student k's ability is measured without error.

that if $\pi_3 \neq 0$, $\theta * (1+\gamma) \neq 0$, thus $\theta \neq 0$ - so testing $\hat{\pi}_3$ is a valid test for the presence of peer effects of departed peers.[23]

Thus, I interpret the coefficients π_1, π_2, and π_3 together as the total effect of lagged departed peer, remaining peer, and own test scores on own contemporaneous test scores where the departed and remaining peers are assumed to be non identical.

5.1.3 Final Model

Utilizing this theoretical framework, my main model is outlined in equation (10):

$$
\begin{aligned}
TS_{i,g,c,t,s} = \beta_0 &+ \beta_1(L_{i,g,c,t-1,s}) + \beta_2(A_{i,g,c,t-1,s}) + \beta_3(\overline{CA}_{i,g,c,t-1,s}) \\
&+ \beta_4(\overline{PA}_{i,g,c,t-1,s}) + \beta_5(CT_{i,g,c,t-1,s}) + \beta_6(\sigma^L_{i,g,c,t-1,s}) \\
&+ \beta_7(\overline{PA}_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}) + \beta_8(\sigma^L_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}) \\
&+ \beta_9(X_{i,g,c,t-1,s}) + \alpha_s + +\epsilon
\end{aligned}
\tag{10}
$$

Here, $TS_{i,g,c,t,s}$ is the test score for student i in grade g in classroom c in school s at time t. $L_{i,g,c,t-1,s}$ is the proportion of students in student i's class that left the experiment at time $t-1$. $A_{i,g,c,t-1,s}$ is student i's ability, measured by their test score in time $t-1$. $\overline{CA}_{i,g,c,t-1,s}$ is classmate ability for student i. That is, the average test score of the students in student i's class (not including student i) that don't leave the experiment, measured by their test score in time $t-1$. $\overline{PA}_{i,g,c,t-1,s}$ is the departed peer ability for student i. That is, the average test score of the students in student i's class that left the experiment in $t-1$. This is my main coefficient of interest as I am interested in how the leavers' ability affects students. $CT_{i,g,c,t-1,s}$ is a categorical variable for the class type student i was enrolled in during time $t-1$ (e.g. small, regular-sized, or regular sized with teachers aide). $\sigma^L_{i,g,c,t-1,s}$ is the variance in departed peer ability (i.e. the variance in test scores across all departed peers in student i's class). $\overline{PA}_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}$ is an interaction term between the departed peer

[23]It may be the case that $\gamma = -1$, in which case we would not observe $\theta \neq 0$; however, this is unlikely.

ability and the proportion of students who leave the experiment and $\sigma^L_{i,g,c,t-1,s} \times L_{i,g,c,t-1,s}$ is an interaction between the departed variance in departed peer ability and the proportion of leavers. $X_{i,g,c,t-1,s}$ is a vector of controls for each student including observable student, teacher, and school data. Examples of this include student attendance, teacher's highest education level, student's race, and others. Lastly, α_s is the school fixed effect and ϵ is the regression error term.

5.1.4 Potential Limitations in Estimation

There are additional limitations not previously mentioned that may influence the interpretability of my coefficients. First, since test scores are measured at the end of the year, all measures of lagged test scores are correlated. The following table shows how the OLS estimates of various lagged measures change when they are subsequently included in a model:

Table 5: OLS Estimates for First Grade Test Score on Lagged Achievement Measures

	Dependent variable: Total 1st Grade Test Score				
	(1)	(2)	(3)	(4)	(5)
Total Own Kindergarten Test Score	0.911***			0.960***	0.963***
	(0.023)			(0.021)	(0.022)
Peer Average Test Score Kindergarten (Leave-One-Out Mean, Excluding Leavers)		0.140**		-0.435***	-0.335***
		(0.058)		(0.059)	(0.074)
Mean Test Score of Leavers in Kindergarten Class			0.121***		-0.111**
			(0.039)		(0.046)
Mean ± SD of Response	2163 ±169	2163 ±169	2157 ±169	2163 ±169	2159 ±169
School Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,739	2,739	2,708	2,739	2,509
R^2	0.618	0.270	0.271	0.631	0.631
Adjusted R^2	0.607	0.250	0.251	0.621	0.619

Note: *p<0.1; **p<0.05; ***p<0.01
Constant not shown. Standard errors clustered at the kindergarten peer-group level

As shown, when lagged own test scores are included, the sign of the coefficient on the lagged leave-one-out mean peer test scores remains negative (even though based on correlation and intuition it should be positive). This is also evident in the change of the sign on the coefficient of lagged departed peer mean test score. In addition, there appears to be a significant negative correlation with the proportion of student i's class that left in time $t-1$ and test scores in both time t and $t-1$ and kindergarten test scores. These initial relationships are also depicted in the below scatterplots.

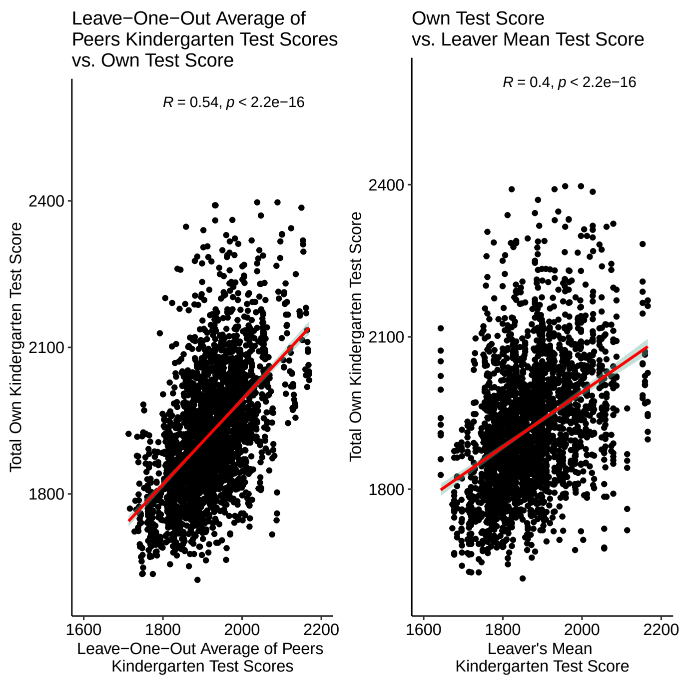


Figure 8: Relationship between proportion of student i's class that left and test scores. Red line indicates line of best fit, shaded region indicates confidence interval. P-value for regression coefficient and correlation coefficient shown.

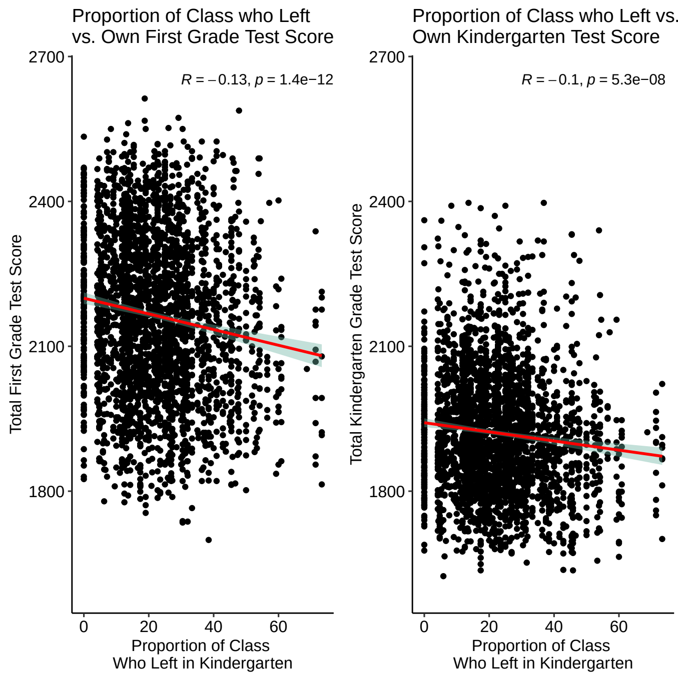


Figure 9: Relationship between proportion of student i's class that left and test scores. Red line indicates line of best fit, shaded region indicates confidence interval. P-value for regression coefficient and correlation coefficient shown.

As such, the inclusion of all three lagged measures of achievement in addition to the proportion of leavers results in what appears to be collinearity - there is a high level of correlation among these variables and their inclusion seems to change the magnitude of their effect.[24] In particular, this inclusion results in a downward shift on the coefficients of lagged leaver's means, and lagged peer means as well as upward bias in lagged own test score. In addition, the positive correlation between leaver's mean test scores and contemporaneous own test score is surprising. These correlations also continue with varying degrees between the 1st and 2nd grade years as well. A discussion of the reason for this relationship is outlined at the end of the paper. I subsequently fit two additional models, one which does not include the mean and variance of lagged peer test scores and one which does not include the proportion of leavers or the variance in leavers' test scores. I utilize these specifications to better understand the nature of the these effects in the model.

[24]This effect, although not shown in the paper, is also present with the inclusion of interaction terms between own lagged test scores and both departed and non-departed peer test scores.

5.2 Results - Peer Effects from Attrition

5.2.1 Grades K-1

Table 6 below depicts the estimates from the peer effects from attrition between grades K-1. The regression is fit over all students that stayed between grades K and 1 and did not switch schools. Specification (2) is that with school fixed effects and is my preferred specification. Specification (3) is a model without the inclusion of the proportion of leavers in a class and the variance of leavers' test scores. Specification (4) is a model without the inclusion of leaver mean test score. The purpose of specifications (3) and (4) are to examine the effects of the previously explained limitations on my coefficients. I find that, even with the inclusion of school fixed effects, a student's own test score is a strong predictor of future test scores. Additionally, with the inclusion of the proportion of departed peers I find that the coefficient on kindergarten peer average test scores (leave-one-out mean) is shifted downward. As examined before, due to the reflection problem among my lagged measures of achievement, none of my estimates can be interpreted as causal. Additionally, I find that the estimates of the effect of lagged peer achievement are negative. Correlations in the data point to a positive association between lagged peer test scores and contemporaneous test scores. Given the existence of this discrepancy, it is not feasible to interpret these estimates as causal. We can only conclude that the composite effect of peer and own lagged achievement on own contemporaneous achievement is positive, with much of this effect attributed to own achievement and a smaller portion attributed to lagged achievement. Once controlling for the proportion of peers who left a class, we find that the coefficient on leaver's mean test scores is no longer significant.

As well, students who were not on free lunch had first grade test scores approximately 40 points higher than those on free lunch, holding all else constant. This further speaks to the association between income and ability previously mentioned. Given the general insignificance of the leaver coefficient and the positive correlation previously identified between leavers' test scores and contemporaneous test scores it may be the case that both channels of attrition are present in this experiment, though the lower income channel dominates the higher-income channel. We would expect a positive correlation between mean leaver's test score and contemporaneous test score if it meant that the loss of these students increased achievement (i.e. these students were potentially disruptive to learning) under the assumption that the presence of higher-ability peers have a positive association with learning. Given the insignificance and downward bias of this effect, it may be the case that the effect of attrition from high-ability students is muted by the effect of attrition from lower-ability students. Such a result corresponds with the findings regarding attrition.

A potential reason for the negative coefficient on non-departed peer average test score in kindergarten across specifications (3) and (4) is that of exclusion bias - the phenomenon whereby when peers are randomly assigned a negative correlation between peer outcomes may be present.[25] Exclusion bias occurs due to the fact that when peers are randomized into classrooms, they are done so without replacement. As such, since randomization within STAR occurred at the school level, when school fixed-effects are included in the estimation of peer effects we observe a negative bias in estimates of peer effects generated by these models. [26] Exclusion bias is reduced when the size of each pool (in this case the number of students within each Project STAR school) is large enough.[27] Since peer departures are perhaps non-random, such bias cannot explain the negative coefficient on leaver mean test score in specification (3) (which is a surprising result in and of itself); however, it may be one reason for why we observe a negative effect of lagged peer test scores on own contemporaneous achievement.

[25] Caeyers & Fafchamps (2020).

[26] Caeyers & Fafchamps (2020) describe that for a student i, if their test scores are above-average then the average of their "pool" (i.e. the set of students whom they could possibly be randomized with) have scores lower than that of student i. The converse is true if i's test scores are lower than average. As such, even after accounting for the average test score of the pool at the school-level through fixed effects, student i's test score "are negatively correlated with the expected value of the remaining peers in the pool."

[27] Given the high number of "rural" schools who participated in STAR, it is likely that many of these selection pools were comparatively "small" thereby increasing the possibility that these coefficients are potentially somewhat influenced by exclusion bias.

Table 6: Coefficient-Level Estimates for Peer Effects among Kindergarten Entry Cohort

	Dependent variable: Total 1^{st} Grade Test Score			
	(1)	(2)	(3)	(4)
Total Kindergarten Test Score	0.905***	0.897***	0.889***	0.880***
	(0.024)	(0.024)	(0.023)	(0.022)
Proportion of Departed Peers in Kindergarten Class	12.834	6.955		0.483
	(8.243)	(7.397)		(0.330)
Mean Test Score of Departed Peers in Kindergarten Class	-0.074	-0.118	-0.131***	
	(0.119)	(0.090)	(0.041)	
Variance of Departed Peers Test Score in Kindergarten Class	0.00000	0.0002		
	(0.0003)	(0.0002)		
In Small Classes Kindergarten	12.279	7.366	-0.127	-1.831
	(11.808)	(11.228)	(11.088)	(10.578)
In Small Classes 1^{st} Grade	28.105***	33.782***	31.687***	33.088***
	(10.728)	(11.222)	(10.843)	(10.191)
Non-Departed Peer Average Test Score Kindergarten	-0.287***	-0.314***	-0.359***	-0.470***
	(0.083)	(0.070)	(0.066)	(0.056)
Not Free Lunch	39.369***	39.962***	40.000***	39.962***
	(6.950)	(6.194)	(5.802)	(6.194)
Proportion of Leavers in K : Mean Test Score of Leavers in K	-0.007	-0.003		
	(0.005)	(0.004)		
Proportion of Leavers in K : Variance of Test Score of Leavers in K	-0.00002	-0.00002		
	(0.00002)	(0.00002)		
Constant	942.469***			
	(290.964)			
Mean ± SD of Response	2159 ± 170	2159 ± 170	2159 ± 169	2159 ± 170
Fixed Effects	No	Yes	Yes	Yes
Observations	2,228	2,250	2,481	2,710
R^2	0.606	0.674	0.662	0.663
Adjusted R^2	0.599	0.659	0.648	0.651
Residual Std. Error	107.989 (df = 2189)	99.445 (df = 2151)	100.422 (df = 2385)	99.705 (df = 2614)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Controls include: Student Gender, Student Race, Kindergarten Teacher Experience (yrs.), Days absent in Kindergarten, Receiving Free Lunch, School Grade Range, % Students in School Bused, 1^{st} Grade Teacher Experience (yrs.), Days Absent in 1^{st} Grade, Received Special Education in K or 1^{st} Grade, Received Special Instruction in K or 1^{st} Grade, Kindergarten Teacher Education, 1^{st} Grade Teacher Education, 1^{st} Grade Teacher Experience (yrs.), Days Absent in 1^{st} Grade, 1^{st} Grade Teacher Race. School Urbanicity. Student characteristics determined based on Kindergarten Grade information unless otherwise specified. Standard errors clustered at the kindergarten peer-group level. Coefficient estimates for controls reported in Appendix.

5.2.2 Grades 1-2

Table 7 below depicts the estimates from the peer effects from attrition between grades 1-2. The regression is fit over all students that stayed between grades 1 and 2 and did not switch schools. Similar to the previous set of models, specification (2) is that with school fixed effects and is my preferred specification. Specification (3) is a model without the inclusion of the proportion of leavers in a class and the variance of leavers' test scores. Specification (4) is a model without the inclusion of leaver mean test score. As shown, similar to the previous table, a student's test scores are a product of their own test scores as well as the lagged test scores of their peers; however, these lagged peer coefficients seem to be similarly shifted downward. Exclusion bias may explain some of this; however, it is unclear what the true reason is for the negative estimated relationship between these outcomes and own contemporaneous test scores. The effect of income is also replicated here, as students who were not on free lunch had second grade test scores approximately 21 points higher than those on free lunch in first grade, holding all else constant. The effect of leavers is once insignificant among this cohort, potentially providing evidence of the attrition profile previously described. Given that income was still a significant predictor attrition between the 1st and 2nd grades, it makes sense why we may subsequently see this insignificance.

Table 7: Coefficient-Level Estimates for Peer Effects Among All 1st Grade Students

	Dependent variable: Total 2nd Grade Test Score			
	(1)	(2)	(3)	(4)
Total 1st Grade Test Score	0.816***	0.818***	0.817***	0.811***
	(0.015)	(0.014)	(0.012)	(0.012)
Proportion of Departed Peers in 1st Grade Class	5.732	1.909	0.164	0.393
	(5.780)	(4.341)	(0.249)	(0.246)
Mean Departed Peer Test Score in 1st Grade Class	-0.019	-0.031	-0.062***	
	(0.083)	(0.054)	(0.024)	
Variance of Departed Peer Test Score in 1st Grade Class	-0.0004	-0.0003		
	(0.0004)	(0.0003)		
In Small Classes 1st Grade	3.797	12.751	5.443	8.563
	(16.093)	(12.769)	(12.534)	(12.597)
In Small Classes 2nd Grade	14.836	7.630	7.096	4.355
	(15.217)	(12.614)	(12.644)	(12.653)
Non-Departed Peer Average Test Score 1st Grade	-0.166***	-0.234***	-0.241***	-0.280***
	(0.051)	(0.050)	(0.044)	(0.043)
Not Free Lunch	23.133***	20.994***	21.041***	21.739***
	(4.639)	(3.820)	(3.502)	(3.417)
Proportion of Leavers in 1st : Mean Test Score of Leavers in 1st	-0.003	-0.001		
	(0.003)	(0.002)		
Proportion of Leavers in 1st : Variance of Test Score of Leavers in 1st	0.00002	0.00000		
	(0.00002)	(0.00001)		
Constant	904.818***			
	(199.156)			
Mean ± SD of Response	2354 ± 153	2355 ± 153	2354 ± 153	2362 ± 153
Fixed Effects	No	Yes	Yes	Yes
Observations	3,039	3,085	3,620	3,855
R^2	0.695	0.748	0.744	0.741
Adjusted R^2	0.692	0.740	0.738	0.734
Residual Std. Error	85.001 (df = 3005)	77.997 (df = 2989)	78.492 (df = 3525)	78.998 (df = 3763)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Controls include: Student Gender, Student Race, 2nd Grade Teacher Experience (yrs.), Days absent in 2nd Grade, Receiving Free Lunch, School Grade Range, % Students in School Bused, 1st Grade Teacher Experience (yrs.), Days Absent in 1st Grade, Received Special Education in 2nd or 1st Grade, Received Special Instruction in 2nd or 1st Grade, 2nd Teacher Education, 1st Grade Teacher Education, 1st Grade Teacher Experience (yrs.), Days Absent in 1st Grade, 1st Grade Teacher Race, 2nd Grade Teacher Race. School Urbanicity. Student characteristics determined based on 1st Grade information unless otherwise specified. Standard errors clustered at the 1st grade peer-group level.

6 Identification of Attrition Channels

6.1 Empirical Framework

Given that I am unable to directly classify the students who left Project STAR into those who dropped out, left for private schools, or migrated out of state I estimate the potential groupings of departed students through a two part procedure. First, utilizing clustering, I separate students who left the public school system entirely into distinct groups. I subsequently test the sensitivity of these groupings by clustering on different sets of controls, following a procedure similar to that described in Emily Oster's 2016 paper: *Unobservable Selection and Coefficient Stability: Theory and Evidence.* Such methods allow me to see, under the assumptions of my model, how "non-randomly" would the data have to be selected in order to observe the clusters I identify. Different sets of controls may induce different selections for which students may resemble, for example, those likely to leave for private schools, so understanding how sensitive my model is to these assumptions would be beneficial to understand the limitations of this analysis.

6.1.1 Clustering Analysis

To group students who left the public school system into "clusters," I fit a variation of the k-means clustering algorithm. The usefulness of clustering is that it allows me to separate students into groups that "look" alike thus allowing me to perhaps infer or make assumptions about which students went to private school without the data needed to classify these students as private school students. K-means clustering is a technique through which for a given set of observations, these observations are separated into k sets such that some notion of "closeness" within the points in the set is minimized. To explain more rigorously:

Imagine we have a collection of N students who we know left the public school system $x_1, \ldots, x_n$ denoted by the set X. For each of these students there exists a $d \times 1$ vector $\vec{v}_i \in R^D$ where each row of the vector represents the value of each covariate for student x_i (i.e. row one is race, row two is gender, etc.). We denote the set of these vectors as $V = \{\vec{v_1}, \ldots, \vec{v_n}\}$. First, we take k random points in X and generate a set $M = m_1, \ldots, m_k$ of the mean Euclidean distances between all of the points and each of the randomly selected initialization points. Then, we assign each x_i to the cluster with the closest mean: that is the cluster with the least squared Euclidean distance.

We end up with k clusters which represent points that are "close" to one another. We then iterate this process multiple times, recalculating the means of the observations assigned to each cluster. We thus end up with k clusters of observations in X which we believe to be "similar" in some way.

6.1.2 Beyond K-Means

K-means, although a widely used model, is only applicable for clustering around numeric variables. For ordinal or categorical variables (e.g. free-lunch status) it is hard to discern what the "difference" between two levels of a variable may be. For example, if one observation's Ecludiean distance from another is .5, what does this mean for a variable that takes values of 0 or 1? How is one able to discern what a free-lunch status of 0.5 is? Thus, other techniques are required for mixed numerical and categorical data. I utilize a variation of that outlined by Filaire (2018).

In order to measure distances or, more accurately, dissimilarities between categorical features, I utilize the Gower distance:

$$d(i,j) = \frac{1}{p}\sum_{i=1}^{p} d_{ij}^{(f)} \tag{11}$$

Here, f corresponds to a given variable between two observations i, j and $d(i,j)$ is the dissimilarity between these two observations along that feature. For numerical variables, dissimilarity is observed according to the following formula:

$$d_{i,j}^{(f)} = \frac{| x_{if} - x_{jf} |}{R_f} \tag{12}$$

This is calculated as the absolute value difference between the two values for the variable f across observations i, j divided by R_f, the maximum range observed in the data between two variables of this given variable.

For categorical variables, differences are either 1 or 0, where 0 indicates two different values and 1 indicates the same value.

6.1.3 K-Medoids

In conjunction with the Gower distance, I utilize the K-Medoids algorithm, a variation of K means. K-Medioids breaks the data into groups around medioids and minimizes, in this case, the Gower distance between observations. Initialization points in K-Medioids, unlike K-Means, are actual data points, rather than random points within the N dimensional space the observations may occupy. This lends for better interpretation regarding the "center" of each cluster. Additionally, K-Medioids when combined with the Gower distance is more robust than K-Means to noise and outliers due to its minimization of dissimilarities rather than Euclidean distances. To explain more rigorously:

Say we have data $X : x_1, x_2, \ldots, x_n$ and a distance metric d. First, the algorithm randomly selects k of the n total data points as medoids where k is an exogenously selected parameter. A medoid of a dataset is defined as the following:

$$M = \underset{y \in X}{\operatorname{argmin}} \sum_{i=1}^{n} d(y, x_i)$$

Among these medoids, we calculate the distance between each datapoint x_i and the each of the randomly selected initial medioids. Then, we assign each x_i to the medoid to which it is least dissimilar: that is the cluster it is "closest" to as determined by our distance metric d.

We end up with k clusters which represent points that are most "similar" to one another. We then iterate this process multiple times, recalculating the dissimilarities of the observations assigned to each cluster. We thus end up with k clusters of observations in X which we believe to be "similar" in some way.

6.1.4 Assessing Optimiality

In order to determine the optimal number of clusters K, I utilize the silhouette coefficient. This value determines how similar an observation is to its own cluster when compared to others. The value ranges from -1 to 1, where higher values indicate more optimal conditions for clustering. To explain more rigorously:

Once the data has been clustered by K-Medoids into K distinct clusters C_k, for any data point $i \in C_k$, we can denote:

$$a(i) = \frac{1}{\mid C_k \mid} \sum_{j \in C_k, i \neq j} d(i,j)$$

as the mean distances between each point i and all other points in the cluster C_k. Here, we subtract 1 from $\mid C_k \mid$ due to the fact that we do not want to account for the distance between i and itself $(d(i,i))$. The smaller $a(i)$ is the closer i is to the points in its cluster. We can thus consider the silhouette value of a data point i to be the following:

$$s(i) = \frac{\text{argmin}\{a(i)\} - a(i)}{\max\{\text{argmin}\{a(i)\}, a(i)\}} \tag{13}$$

where $\text{argmin}\{a(i)\}$ represents the smaller minimum distance between i and all points in clusters for which i is not a member. Thus, a value of $s(i)$ close to 1 indicates that observations are well matched (i.e. the level of dissimilarity $a(i)$ between data points is minimized. We can subsequently plot the mean of these silhouette values for different numbers of clusters and determine what the optimal cluster size is based on the maximal interpretive silhouette value.

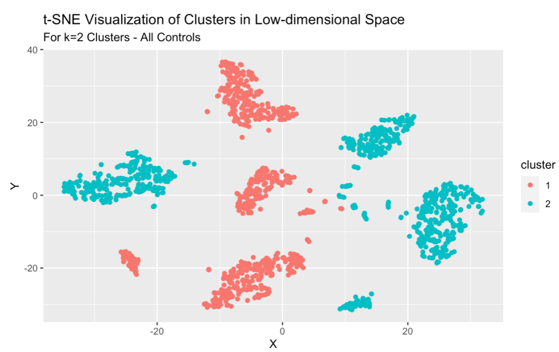
6.2 Results - Clustering Analysis of Attrition

6.2.1 Baseline Analysis for Two Groups

The following clustering analysis is undertaken among departed students in the kindergarten cohort. Students may have left immediately after kindergarten or later on in the experiment. I utilize a number of observed background characteristics to construct clusters of similar students, examining how sensitive group selection is to the choice of background characteristics used in clustering.

(a) Clusters for Data Without Controls for Income

(b) Clusters for Data Without Controls for Achievement



(c) Clusters for Data With all Controls

Figure 10: Visualization of Clusters for K=2

The above figure depicts a visualization of different clustering arrangements for $K = 2$ clusters across datasets with various controls. In this context, I defined controls as the variables utilized to construct the clusters. These figures represent a low dimensional illustration of the clusters created. As such, their interpretability is simply to visualize how the clusters are formed. If the clusters seem to be distinct with little overlap, it indicates that the choice of clusters k is correct. It appears, at least initially, that these clusters are non-overlapping with clear delineations and high levels of similarity across observations. The dataset shown in Figure 6(a) includes controls for: race; gender; and kindergarten test scores in math, reading, listening, and world skills. The dataset shown in Figure 6(b) includes controls for: race; gender; free-lunch status; special education; special instruction; recommendation to repeat kindergarten; and days absent in kindergarten. The dataset shown in Figure 6(c) includes controls for: race; gender; kindergarten test scores in math, reading, listening, and world skills; free-lunch status; special education; special instruction; and days absent in kindergarten. A summary table of the clusters is provided below:

	All Controls		No Income Controls		No Test Score Controls	
Cluster:	(1)	(2)	(1)	(2)	(1)	(2)
Number of Observations:	726	836	729	838	725	835
Race						
White	59.2%	59.9%	59.3%	60%	59.2%	60%
Black	39.9%	39.4%	39.7%	39.1%	40%	39.3%
Asian	0.15%	0.36%	0.12%	0.3%	0.13%	0.36%
Hispanic	0.15%	0.11%	0.12%	0.2%	0.13%	0.11%
Native American	0.3%	0%	0.24%	0%	0.27%	0%
Other	0.3%	0.23%	0.24%	0.3%	0.27%	0.23%
Mean Kindergarten Test Scores						
Math SAT Scaled Score	476.3	465.1	475.7	464.7	476.3	465
Reading SAT Scaled Score	431.6	424.7	431.3	424.5	431.6	424.7
Listening SAT Scaled Score	532	525.4	531.9	525.9	532	525.4
Word Study Skills SAT Scaled Score	428.7	422.2	428.1	421.83	428.7	422.2
Male	0%	100%	0%	100%	0%	100%
Free Lunch	57%	56.5%	56.8%	56%	56.7%	56.4%
Special Education	3.3%	5.2%	3.7%	5.2%	3.3%	5.3%
Receives Special Instruction	5%	6.2%	4.4%	5.7%	5%	6.2%
Mean Days Absent	11.3	11.6	11.7	12	11.3	11.6

Table 8: Summary table of characteristics for students in clusters.
Note: all values correspond to kindergarten-grade information unless otherwise specified. All students represented in the table are those that left the experiment. Variable definitions are consistent with Table 1.

However, when examining the summary table it becomes clear that for $K = 2$ clusters the algorithm is merely separating students by gender. As shown, these clusters look almost identical. If two distinct channels of attrition were present within the data, we would find that when $K = 2$ one cluster would be comprised of low-income, low-ability individuals whereas the other would be comprised of high-income, high-ability individuals. As shown, this is clearly not the case. It is likely that attrition within Project STAR is due to other factors which aren't distinguishable in two clear groups.

6.2.2 Attrition Pathways From Optimal Group Selection

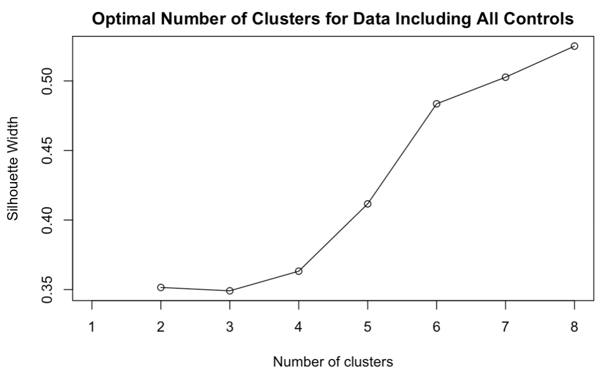


Figure 11: Silhouette Plot for Data Including All Controls

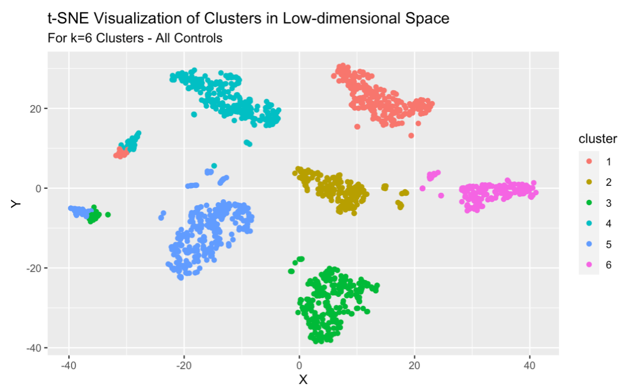


Figure 12: Visualization of Optimal Clustering for Data Including All Controls

The above figure depicts the optimal number of clusters for the dataset including all controls. As shown, it appears from Figure 7 that a value of $K = 6$ seems optimal. Figure 8's depiction of these clusters shows some distinct regions, though clusters 1 and 4 as well as clusters 3 and 5 do overlap.

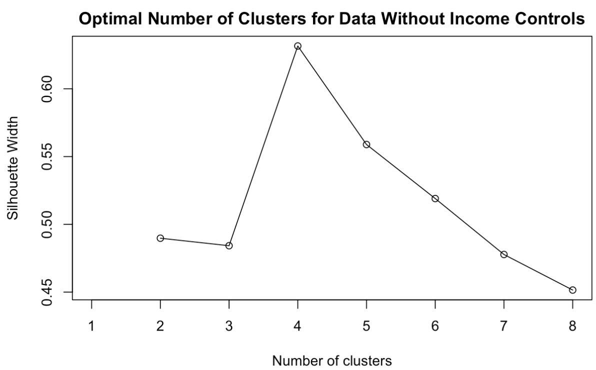


Figure 13: Silhouette Plot for Data Without Controls for Income

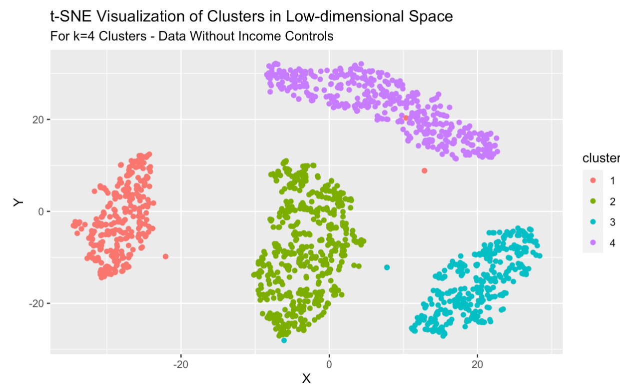


Figure 14: Optimal Clustering for Data Without Controls for Income

The above figure depicts the optimal number of clusters for the dataset without controls for income. As shown, it appears from Figure 9 that a value of $K = 4$ seems optimal. Figure 10's depiction of these clusters shows clear distinctions between clusters, indicating high levels of convergence in the algorithm and high levels of similarity between points in clusters.

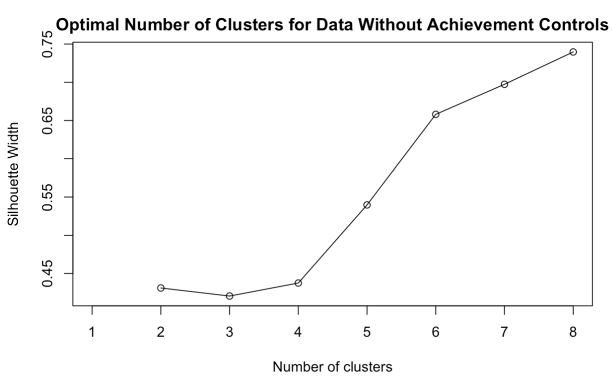


Figure 15: Silhouette Plot for Data Without Controls for Achievement

Figure 16: Visualization of Optimal Clustering for Data Without Controls for Achievement

The above figure depicts the optimal number of clusters for the dataset without controls for student achievement. As shown, it appears from Figure 11 that a value of $K = 6$ seems optimal. Figure 12's depiction of these clusters shows some distinct regions, though clusters 1 and 4 as well as clusters 3 and 6 do overlap. In addition, I include a summary table below of the clusters generated by using $K = 6$ on the data with all controls.

Cluster:	(1)	(2)	(3)	(4)	(5)	(6)
Number of Observations:	257	198	300	291	337	175
Race						
White	0%	100%	0%	87%	90%	100%
Black	100%	0%	100%	13%	8%	0%
Asian	0%	0%	0%	0%	1%	0%
Hispanic	0%	0%	0%	0%	1%	0%
Native American	.6%	0%	0%	0%	0%	0%
Other	0%	0 %	0%	0%	0%	0%
Mean Kindergarten Test Scores						
Math SAT Scaled Score	460.6	460.9	449.1	494	481.7	470
Reading SAT Scaled Score	420	421.1	414.3	440	436.1	426
Listening SAT Scaled Score	513.5	525.2	507.7	549.2	560	530
Word Study Skills SAT Scaled Score	415	417.4	411.1	444.5	435	422.7
Male	0%	100%	0%	100%	0%	100%
Free Lunch	92%	100%	91%	0%	0%	100%
Special Education	1.6%	8.6%	3.3%	2.4%	5%	7.4%
Receives Special Instruction	0%	12.1%	2.7%	3.8%	6%	14.3%
Mean Days Absent	11.5	16.5	10.1	9.8	10.1	13.5

Table 9: Summary table of characteristics for students in clusters for $K = 6$.
Note: all values correspond to kindergarten-grade information unless otherwise specified. All students represented in the table are those that left the experiment. Variable definitions are consistent with Table 1.

The above silhouette plots and visualizations show that there are potentially more than 5 distinct groups of students that are leaving Project STAR. As such, it is most likely the case that there is a high association between low-income and low-ability and attrition; however, among this group of students there is potentially more granular channels of attrition. In total, it seems that the clustering is sensitive to the inclusion of controls for income. This is further illustrated in the similarity between Figures 7 and 11. This result corresponds with the results for attrition among the kindergarten cohort. As shown in the summary table, clusters (4) and (5) seem to be a higher-ability, higher-income cluster and clusters (1), (2), (3), and (6) seem to be lower ability, lower income and broken up by race and gender. This in fact does provide evidence for a distinct channel of attrition for higher-income and higher-ability groups. It seems that primarily low-income and low-ability students in the kindergarten cohort were the ones that departed STAR, and there were varying groups of these students which can be classified based on race and gender. However, there is also a group of higher-ability, higher-income students who left star as evidenced by groups (4) and (5) although this group is much smaller in size as compared to the others.

7 Discussion

7.1 Attrition in Project STAR

There exists some evidence of two main channels of attrition for students in Project STAR. The first is a channel through which low-income and low-ability (defined as in the bottom 25th percentile) students exit the public school system. Although my paper does not identify the mechanism through which this occurs, the identification of this channel is consistent with other analyses in the literature.[28] In general, low-income and low-ability students are more likely to exit the public school system than their counterparts. This result is depicted not only in the results from the AFT model of attrition but also from the non-parametric Kaplan-Meier estimates. That being said, it seems that the presence of this channel begins to diminish as the experiment went on. Reasons for this are unclear but one possibility is a reduction in the number of students that were recommended to repeat a grade after the 3rd year of the experiment. Differential attrition based on class size seems to not be significant after controlling for income. I interpret these findings as strong evidence that attrition of low-income and often low-ability students was prevalent in Project STAR. An alternative explanation though may be that there was a financial shock unique to these neighborhoods or Project STAR schools during the first two years of the experiment that caused a significant number of low-income students to leave the public school system. Since I am unable to control for economic conditions across schools beyond the school-level controls collected within Project STAR such a shock may be possible however it is unlikely given

[28]This is consistent with findings in Fernandez-Suarez, et al. (2016) and other papers.

other analyses of STAR in the literature.

At the same time, as evidenced by the clustering analysis, there appears to be evidence of a pathway of attrition among non low-income and higher-ability students. Such students, though they only comprise about 40% of the total students that left STAR among the kindergarten entry cohort, are still starkly different from their lower income counterparts. The existence of these students is further evidenced by the insignificance of and positive association between leaver mean test score and student achievement. If these departed students were of varying ability it would make sense that peer effects from such attrition would be insignificant and its correlation would be positive, under the assumption that higher-ability peers are associated with gains in achievement. In all, it seems that the two channels of attrition I identify seem to be the dominant channels of attrition among students in STAR and among these the low-income channel seems to dominate. That being said, the generalizeability of these results to public schools writ large must be taken with caution. There are many things unique to Project STAR that may have influenced the existence of these attrition pathways including voluntary school selection into the program, attrition due to assignment in an "unfavorable" class or other issues with measurement.

7.2 Peer Effects from Leavers

Peer effects estimation is especially tricky and many models of estimation may potentially be subject to some bias. Due to the reflection problem as well as my inability to measure true ability, the peer effects estimates I have outlined cannot be interpreted as causal. However, given that students were randomly assigned to classrooms, the null hypothesis of my models would predict no relationship between peer and own outcomes and it seems that the data within Project STAR reject this null. As such, there is a relationship between own and peer outcomes and this relationship is most likely positive. Peer effects estimation must take into account the outcomes of departed peers as it is likely that these peers, although not interacting with students contemporaneously, potentially have an association with their achievement. It seems that the total effect of peer background on own achievement is a combination of the effect of contemporaneous peer achievement, lagged peer achievement, departed peer achievement. That being said, the introduction of controls such as the proportion of peers that left a student's class - although related to departed peer achievement - likely bias estimates of the association between lagged peer achievement and contemporaneous achievement downward.

8 Conclusion

There exists evidence of two channels of attrition within Project STAR: one which is primarily comprised of low-income, low-ability students and another comprised of higher-income, higher-ability students. The low-income channel appears to dominate the high-income channel; however, attrition in general seemed to decrease as the experiment went on. This may indicate that within elementary-education students who are most likely to leave the public school system are those from low-income families. The effects of this attrition on the students they leave behind are unclear. Better measures of pre-randomization achievement are required to properly measure the peer effects of attrition. There is some evidence that peer characteristics influence student outcomes positively, however the magnitude and significance of this relationship is unclear. Further research is required to better understand the impact of attrition on achievement in early education as well as the generalizability of this findings across public schools. If there exists a negative association between departed peer ability and students own test scores the following year, policymakers may be inclined to introduce targeting measures in order to reduce these effects; however, better empirical methods and data are required to have a full understanding of these effects themselves.

9 References

Boozer, M. A., & Cacciola, S. E. (2001). *Inside the "black box" of Project STAR: Estimation of peer effects using experimental data.* New Haven, CT: Yale University, Economic Growth Center.

Caeyers, B., & Fafchamps, M. (2016). *Exclusion bias in the estimation of peer effects.* National Bureau of Economic Research.

Dickinson, E. E. (2016). Coleman report set the standard for the study of Public Education. *The Hub.* https://hub.jhu.edu/magazine/2016/winter/coleman-report-public-education/.

Dills, A. K. (2005). Does cream-skimming curdle the milk? A study of peer effects. *Economics of Education Review*, 24(1), 19–28. https://doi.org/10.1016/j.econedurev.2004.01.002

Fernández-Suárez, A., Herrero, J., Pérez, B., Juarros-Basterretxea, J., & Rodríguez-Díaz, F. J. (2016). Risk factors for school dropout in a sample of juvenile offenders. *Frontiers in Psychology*, 7, 1993. https://doi.org/10.3389/fpsyg.2016.01993

Filaire, T. (2018, July 16). *Clustering on mixed type data.* Towards Data Science. https://towardsdatascience.com/clustering-on-mixed-type-data-8bbd0a2569c3

Hanushek, E. A. (2008). Education Production Functions. In *The New Palgrave Dictionary of Economics* (pp. 1–5). Palgrave Macmillan UK.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114(2), 497–532. http://www.jstor.org/stable/2587015

Lazear, E. P. (2001). Educational Production. *The Quarterly Journal of Economics*, 116(3), 777–803. http://www.jstor.org/stable/2696418

New York City Independent Budget Office. (2014). *Staying or going? comparing student attrition rates at charter schools with nearby traditional public schools.* Retrieved April 7, 2022, from https://www.ibo.nyc.ny.us/iboreports/2014attritioncharterpublic.html

Preud'homme, G., Duarte, K., Dalleau, K., Lacomblez, C., Bresso, E., Smaïl-Tabbone, M., Couceiro, M., Devignes, M.-D., Kobayashi, M., Huttin, O., Ferreira, J. P., Zannad, F., Rossignol, P., & Girerd, N. (2021). Head-to-head comparison of clustering methods for heterogeneous data: a simulation-driven benchmark. *Scientific Reports*, 11(1), 4202. https://doi.org/10.1038/s41598-021-83340-8

Rohlfs, C., & Zilora, M. (2014). Estimating parents' valuations of class size reductions using attrition in the Tennessee STAR experiment. *The B.E. Journal of Economic Analysis & Policy*, 14(3), 755–790. https://doi.org/10.1515/bejeap-2013-0024

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, 116(2), 681–704. https://doi.org/10.1162/00335530151144131

Sacerdote, Bruce. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education* (pp. 249–277). Elsevier.

Sojourner, A. (2013). Identification of peer effects with missing peer data: Evidence from project STAR: Identification of peer effects with missing peer data. *Economic Journal (London, England), 123*(569), 574–605. https://doi.org/10.1111/j.1468-0297.2012.02559.x

Stinebrickner, R., Stinebrickner, T. R. (2003). Understanding educational outcomes of students from low-income families: Evidence from a liberal arts college with a full tuition subsidy program. *The Journal of Human Resources*, 38(3), 591. https://doi.org/10.2307/1558769

10 Appendix

10.1 Diagnostic Convergence – AFT Model of Kindergarten Cohort

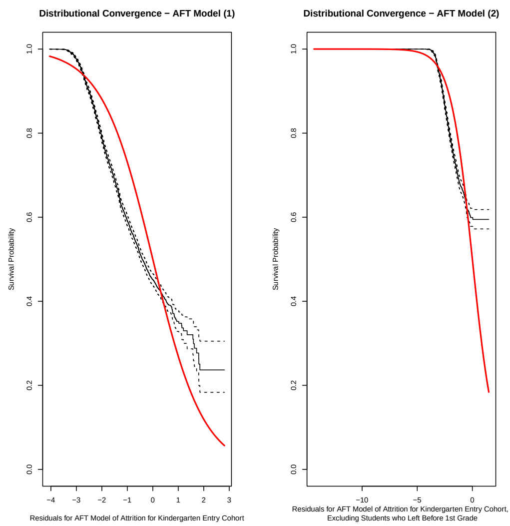


Figure 17: Plot of convergence for AFT Model of Kindergarten Cohort Attrition. Red line represents Weibull distribution, black line represents attrition as modeled by the residuals.

10.2 Diagnostic Convergence – AFT Model of 1-2 Attrition

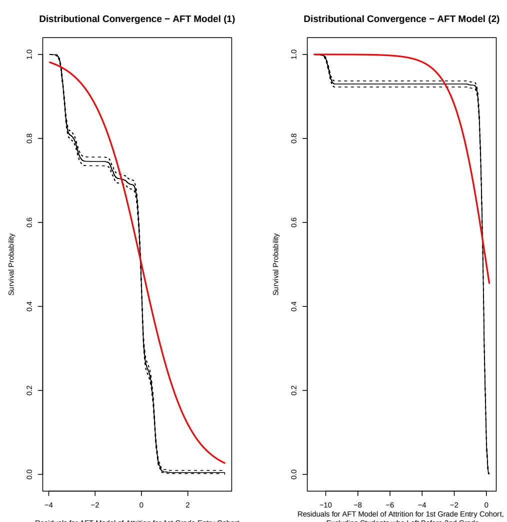


Figure 18: Plot of convergence for AFT Model of 1st Grade and Beyond Attrition. Red line represents Weibull distribution, black line represents attrition as modeled by the residuals.