Big fishes in small ponds: Ability rank and human capital investment*

Benjamin Elsner[†] Ingo Isphording[‡]

March 6, 2015

Abstract

We study the impact of a student's ordinal rank in a high-school cohort on educational attainment several years later. To identify a causal effect, we compare multiple cohorts within the same school, exploiting exogenous variation in cohort composition. We find that a student's ordinal rank in high-school significantly affects educational outcomes later in life. If two students with the same ability have a different rank in their respective cohort, the student with the higher rank is significantly more likely to finish high-school, to attend college, and to complete a 4-year college degree. These results suggest that students under-invest in their human capital if they have a low rank within their cohort even though they have a high ability compared to most students of the same age. Exploring potential channels, we find that students with a higher rank have higher expectations about their future career, and feel that they are being treated more fairly by their teachers.

JEL Codes: I21, I23, J24

Keywords: Human capital, ordinal rank, peer effects, educational attainment

*We would like to thank Chris Jepsen and Ulf Zölitz for helpful comments.

†Corresponding author. Institute for the Study of Labor (IZA). Address: Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany. elsner@iza.org, www.benjaminelsner.com.

‡Institute for the Study of Labor (IZA). Address: Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany. isphording@iza.org

1 Introduction

The characteristics of classmates are often among the decisive factors for parents when choosing a school for their child. It is commonly believed that children learn and achieve more when they are surrounded by high-ability classmates. The literature on peer effects in education confirms this belief by showing that high-ability peers can significantly improve a student's academic achievement (Hanushek *et al.*, 2003; Ammermueller & Pischke, 2009; Lavy *et al.*, 2012), and increase the likelihood of going to college (Patacchini *et al.*, 2012). In this paper we explore a channel that runs counter to the positive impact of high-ability peers: a student's ordinal rank in her peer group. A student who is surrounded by classmates with a higher ability has a low rank within the class, which may give the student a distorted signal about the her actual ability. Some of the brightest people in the country may believe that their ability is low, simply because they compare themselves to even brighter classmates. Psychologists have labeled this phenomenon the big-fish-in-a-little-pond effect (Marsh, 1987).

In this study, we test whether being a "big fish" in one's high-school cohort affects the critical transition period from high-school to college and leads to better educational outcomes later in life. Consider two students, Jack and Jim, who have the same ability but a different rank position in their respective high school cohort: Jack is among the students with the lowest ability in his grade, while Jim is among the brightest students in his grade, making him a big fish in a small pond. We want to know whether Jim is more likely to finish high-school, to attend college after graduating from high-school, and to complete a 4-year college degree.

We use data from the National Longitudinal Study of Adolescent to Adult Health (AddHealth), a representative survey that tracks students in the US from high-school to their mid-30s, and contains rich information on cognitive skills as well as educational outcomes before and after graduating from high-school. The key feature for our analysis is that AddHealth covered multiple cohorts of students within the same high-school. To identify a causal effect, we exploit year-on-year changes in the cohort composition within the same school. In a thought experiment, we compare two students with the same ability in the same school, who have a different ordinal rank because they are in different cohorts. We argue that, conditional on attending a given school, being in a given cohort can be seen as quasi-random, because it is only determined by a student's birth date and cannot be influenced by the parents or the students.

We find that a student's ordinal rank has a strong positive impact on educational attainment. Conditional on a student's own ability, a one-decile increase in a student's rank position in high-school — the difference between the first- and the third-best student in a grade of 20 students — increases the likelihood of completing college by one percentage point, and lowers the likelihood of dropping out of high-school by half a percentage point. These average effects mask a great deal of heterogeneity across demographic groups and school types. We find that the ordinal rank matters more for women than for men, and for children with low-educated parents.

We further explore potential channels through which a student's rank position affects educational outcomes later in life. While we cannot directly observe the costs and expected returns to third-level education, we can exploit rich survey information on expectations, mental distress, as well as perceptions about oneself and one's environment as proxies. Applying the same identification strategy as before, we find that students with a higher rank have significantly higher expectations about their educational career, and are generally more hopeful about their future. In addition they are significantly more likely to feel that their teachers treat them fairly. Other potential channels, such as feeling depressed or happy, or feeling that teachers, parents, or friends care about the student, do not vary systematically with the ordinal rank. These results suggest that a student's rank position affects her expected returns to rather than the costs of third-level education.

This paper contributes to at least three literatures. First, it shows that having high-ability peers may come at a cost of being a small fish in a big pond, and thus have negative consequences for one's educational outcomes later in life. The literature on peer effects in education has mostly found that better peers have a positive impact on immediate educational success, ranging from primary schools (Ammermueller & Pischke, 2009) to high-schools (Calvó-Armengol *et al.*, 2009; Lavy *et al.*, 2012; Imberman *et al.*, 2012) to college (Sacerdote, 2001; Zimmerman, 2003; Carrell *et al.*, 2009; Booij *et al.*, 2015).[1] Moreover, better peers in school affect outcomes later in life. As shown by Bifulco *et al.* (2011) and Patacchini *et al.* (2012), exposure to students from higher socio-economic backgrounds increases the likelihood of going to college, while decreasing the probability of dropping out of high-school. The mechanism suggested in most of these studies is spill-overs. High-achieving students increase the performance of low-achieving students, by helping them, or by making them more ambitious. The findings in this paper run counter to most of this literature. Once we consider a student's rank position in a grade rather than the average ability of her peers, we show that a student with a lower ability than her peers has a significantly lower educational attainment in her early 30s.

Second, this paper shows that a student's ordinal rank early in life affects her educational choices *and* outcomes later in life. It has a significant influence on the critical transition from high-school to college. The rank channel has been mostly overlooked in the peer effects literature. A notable exception is the work of Murphy & Weinhardt (2014), who study the impact of rank position in primary school cohorts in the UK on test scores in secondary school, using a large administrative dataset. They find a significant impact of the rank position in primary school on test scores in secondary school. Additional survey data suggests that the effect is mainly explained by a higher self-confidence among students with a higher rank. Our study demonstrates that the consequences of having a high or low rank are more far-reaching, as students with different ranks in high-school invest differently in their human capital later on, and this difference cannot be explained by their absolute level of ability.

Third, this paper also relates to the literature on relative concerns, which provides ample

[1] Peer effects may not necessarily be positive, however. Angrist & Lang (2004), for example, find virtually no causal effect of a higher share of new low-achieving students on the test scores of incumbent students. Moreover, more recent studies have shown that the average positive effect of good peers masks a significant negative impact for very low-achieving students, which can be explained by little interaction between high-and-low achieving peers (Carrell *et al.*, 2013; Feld & Zölitz, 2014).

evidence that people make choices based on comparisons with their peers, and that this comparison affects a person's choices and her well-being. Luttmer (2005), for example, finds a strong negative relationship between a person's well-being and the relative income of her neighbors: people feel worse if their neighbors earn more than them. Clark *et al.* (2010) use experimental and survey data to show that relative concerns affect a person's effort provision. Once informed about others' incomes, people with higher ranks in the income distribution provide more work effort, in a gift-exchange experiment as well as in their actual job. We show that a person's ordinal rank among their peers is equally important in the education context, and that it has long-run implications for a person's career choices and success.

2 Data and Descriptive Statistics

2.1 The Add Health data

Our data is the restricted-use version of AddHealth, a representative longitudinal dataset of American high-school students run by the Carolina Population Center. The aim of AddHealth is to study the impact of health, social conditions, education, as well as family situation and friendships of adolescents on outcomes later in life, such as high-school completion, transition to college, and transition into the labor market. To date, four waves are available. The first wave was administered among a representative sample of US high-school students in grades 7-12, when respondents were between 13 and 18 years old. Follow-ups were run in 1996, in 2000/2001 when most students had left high-school, and in 2008/2009, when most had entered the labor market.[2]

The unique feature of AddHealth for this study is the combination of extensive coverage of school grades in the first wave with a longitudinal dimension, allowing us to observe a representative sample of the high-school peers of every student in the sample, and track most students until their early 30s. We can thus observe the critical transition from high-school to tertiary education, and into the labor market. Moreover, the survey includes a standardized cognitive skills test that gives us an objective measure of cognitive ability, without having to resort to grades or self-reported measures as proxies for ability.

The sampling of AddHealth is cluster-based; in a first step, a representative sample of schools was drawn, and subsequently a random sample of students within each school was drawn from strata defined by gender and grade. The first wave included representative high-schools from all parts of the US. While the complete student population in these schools was interviewed using a basic "In-School" questionnaire, an additional and more comprehensive "In-Home" questionnaire, including the ability test, was administered to a subsample of students. Within schools, students were drawn from grades 7-12. Within each grade, a random sample of 17 boys and 17 girls was drawn, forming the core sample of the survey. Additional students were drawn from the population in order to oversample groups with certain characteristics: twins,

[2] For further information on the study design and the sampling, see Harris (2009), and Harris *et al.* (2009). Data for a fifth wave are currently being collected.

students with disabilities, blacks from well-educated families, as well as students of Chinese, Cuban, and Puerto Rican origin. In 16 schools, all students that were present on the day of the survey were included (saturated schools).

Our sample is the in-home sample of the first wave of AddHealth. We drop from the sample all schools with 20 observations or less (109 obs.), and all grades with 5 students or less (304 obs.). Moreover, we drop all students for who we do not observe the educational attainment (finished high-school, attended college, completed college) in wave IV (4,426 obs.). In total, this leaves us with 13,930 students in 130 schools and 432 grades.

2.2 Outcome variables: Educational attainment

We consider three outcome variables that measure different degrees of educational attainment: *high-school dropouts*, *attended college*, *completed college*. In wave IV of AddHealth, respondents were asked for their highest educational attainment. We consider as high-school dropouts all students who reported that their highest educational attainment was *less than high-school*. The categories *attended college* and *completed college* are nested; *completed college* only includes students who completed a Bachelor's degree, while *some college* is broader and also includes students who attended college but finished with a degree lower than a Bachelor's, or did not finish at all.

Table 1 summarizes the outcome variables for various subgroups. Among all students, 7% did not finish high-school, while 67% attended college. Around half of those who attended college finished at least with a Bachelor's degree. Women have a higher average educational attainment than men, with fewer high-school dropouts and higher college attendance and completion rates.

The educational attainment differs considerably by parental background, as well as by race. The data show a high correlation between the educational attainment of the parents and their children. Children of college-educated parents are four times as likely to complete a college degree and ten times less likely to drop out of high-school than children whose parents were high-school dropouts. There is less variation in the educational attainment of various ethnic groups. Hispanics and blacks have lower educational attainment than whites, but the raw differences are less than 10 percentage points. An exception are Asians, whose educational attainment is considerably higher than in all other groups.

Finally, we consider schools with different average ability and heterogeneity. Unsurprisingly, students from schools in the top half of all sampled schools have a higher educational attainment later in life. We also want to see whether more heterogeneous schools are more or less conducive to educational success. If schools are segregated with respect to students' ability, for example because of a tracking system or because of self-selection of students into schools, one would expect more segregated schools to have different outcomes from more integrated and therefore more heterogeneous schools. The raw data, however, do not support this conjecture. Students who attended schools with a high within-school standard deviation in their students' cognitive ability have similar shares of high-school dropouts and college graduates than more homogeneous schools.

Table 1: Educational attainment by group

Group	*High-school dropouts*		*attended College*		*college completed*		*N*
	mean	(SD)	mean	(SD)	mean	(SD)	
All	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Male	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Female	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Parental background:							
Less than high-school	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
High school	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Some college	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
College	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Race:							
White	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Asian	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Hisp	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Black	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Average school ability:							
High average ability (above median)	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Low average ability (below median)	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
School heterogeneity (within-school SD in ability)							
High heterogeneity(above median)	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645
Low heterogeneity (below median)	0.07	(0.26)	0.67	(0.47)	0.33	(0.47)	13645

Notes: This table displays the share of high-school dropouts, the share of people who enrolled in college, and the share of people who finished a college degree of 3 years or more. Standard deviations reported in parentheses. Within-grade quartiles refer to the quartile in the ability distribution of a student's high-school grade. Parental background refers to the highest level of education and the highest occupational status among both parents. Average school ability is the average ability of the entire school, and above/below median refers to the school distribution, i.e. students in the "above median" group attend schools with an above-median ability-level. The school heterogeneity is measured by the within-school standard deviation of ability.

2.3 Measuring ability and ordinal rank

We use as ability measure the results of a standardized Peabody picture vocabulary test, of which a shortened version was included in the survey. The test scores are computed on basis of a simple task: After a set of pre-tests, participants are asked to allocate words spoken aloud by the interviewer to a set of four different simple pictures. The test proceeds through a series of rounds with increasing difficulty. Test scores are standardized by age with mean 100, and are computed automatically without being made available to the interviewer or the respondent.[3]

[3] Further information on the Addhealth Picture Vocabulary Test is available in the AddHealth documentation at `http://www.cpc.unc.edu/projects/addhealth/data/guides`

Though measuring very basic cognitive skills, the Peabody Picture Test has been shown to have a high re-test reliability and cross-validity to other intelligence tests for adolescents (Dunn & Dunn, 2007). For our purpose of analyzing relative rank concerns, the test has the advantage of measuring individual verbal ability, which is clearly visible in the classroom.

Based on the test score we rank all students of a school grade, with rank 1 being assigned to the student with the lowest ability in the grade, and a rank equal to the number of students in a grade to the student with the highest ability. Students with the same test score are assigned an equal rank. To make the within-grade ranks comparable across grades and schools with a different number of students, we standardize the absolute ranks within a grade to a relative rank position, a continuous measure that assigns value 1 to the student with the highest rank in class, and 0 to the lowest. Simply put, the third brightest student in a grade of 100 has a better rank than the third in a grade of four students, because the third in a grade of 100 is better than 97 students in her class and is thus at the 97th percentile, while the third in a grade of four is better than one person, and worse than 2, and is at the 25th percentile. The relative rank measure reflects this difference; it is calculated as

$$\text{relative rank} = \frac{\text{absolute rank} - 1}{\text{nr of students in grade} - 1}. \tag{1}$$

2.4 Summary statistics

The first panel in Table 2 summarizes the ability measures and other individual characteristics. The two columns on the right display the means of the individual-level variables for students in the bottom and top half of the ability distribution of their grade. At first glance, women and blacks are over-represented among students in the bottom half of a grade, while there is no large difference with respect to average age, and neither in the share of Asians, Hispanics, or students with a migration background. A strong correlation appears between ability and parental education. Children with highly educated parents are more likely to have a higher rank within their grade, while children from parents with lower educational backgrounds are over-represented in the bottom half of a grade.

The second panel in Table 2 summarizes the school and grade characteristics. Schools differ greatly in terms of average ability and heterogeneity. Students in the lowest-ability school scored on average 79 on the standardized test, which is three between-school standard deviations below the mean; the highest-ability school scored 116, or 2.5 between-school standard deviations above the mean. To measure heterogeneity in ability we take the within-school standard deviation of the ability distribution. The within-school standard deviation varies between 9.2 and 20.5, and is on average twice as large as the between-school standard deviation, which is 6.5.

Depending on the school, the grade size varies greatly; in the population it ranges from 5 in the smallest grade to 645 in the largest. More relevant for our study is the actual within-grade sample size. The average grade has 40 students in the sample, which is more than the 34 students drawn at random due to oversampling of minorities and the inclusion of saturated schools. On average, 22% of a grade have been sampled.

Table 2: Summary Statistics of the main variables

		All		bottom 50%	top 50%
Variable	*N*	*Mean*	*SD*	*Mean*	*Mean*
Ability					
Cognitive ability	13645	101.14	14.24	91.18	110.84
Ability rank	13645	0.50	0.29	0.24	0.75
Personal characteristics					
Age	13645	16.13	1.68	16.25	16.01
Female	13645	0.54	0.50	0.57	0.50
Ever repeated a grade	13645	0.20	0.40	0.28	0.13
Migration background (1st & 2nd gen.)	13645	0.15	0.36	0.16	0.14
Asian	13645	0.07	0.25	0.06	0.07
Black	13645	0.22	0.42	0.27	0.19
Hispanic ancestry	13645	0.14	0.35	0.16	0.13
Highest parental education					
Less than high-school	13645	0.14	0.35	0.19	0.10
High-school	13645	0.25	0.43	0.29	0.21
Some college	13645	0.25	0.43	0.24	0.26
College	13645	0.36	0.48	0.29	0.42
School characteristics	*N*	*Mean*	*SD*	*Min*	*Max*
Small (< 401 students)	130	0.22	0.42		
Medium (401-1000 students)	130	0.47	0.50		
Large (> 1000 students)	130	0.31	0.46		
Class size	128	25.86	5.18	10.00	39.00
Mean ability	130	100.31	6.46	79.19	115.80
SD ability	130	12.89	2.29	9.24	20.48
Grade characteristics					
Grade size (population)	432	184.27	131.54	5	645
Nr students in sample	432	40.63	45.27	6	545

Notes: This table displays the mean and standard deviations of the main regression variables for the whole sample, as well as the means for the students above and below the median ability of their school grade. Besides the share of Asians, all differences are statistically significant at the 1%-level. The school characteristics have been reported by the school administrator in a separate survey. In two cases, the information on the average class size was missing.

3 Identification and Estimation Strategy

The cleanest strategy to estimate a causal effect of ordinal rank on educational attainment would be through a randomized experiment. One could assign students randomly to high-schools and observe the relationship between their ordinal rank in their grade and their educational outcomes later in life. This strategy, however, would not allow for a separate identification of rank and school effects, because unobserved school characteristics could influence educational attainment. A more promising alternative would be to assign students to different classrooms within the same school, and run the same experiment in many schools, such that the results would only be driven by the difference in a student's rank position within a class, while keeping the school environment constant.

Although we have no such experiment available, our identification strategy replicates the within-school experiment. Instead of assigning students to classrooms, we compare students from different age cohorts, exploiting the variation in the skill distribution across grade within the same school.

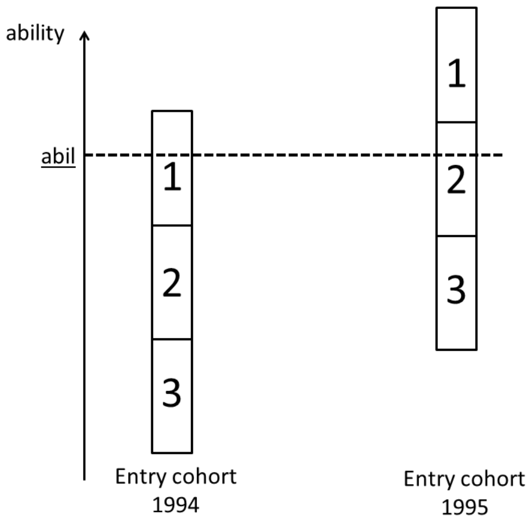

Figure 1: A student with ability=abil has rank 1 in the entry cohort in 1994 but rank 2 in 1995.

Figure 1 illustrates a simple example for the identifying variation. It shows two entry cohorts in the same school, each cohort consisting of 3 students. Consider a student with a given level of cognitive ability *abil*; if she entered the school in 1994, she would have the first rank in her cohort. If she entered the school one year later, she would only be on the second rank, simply because there is one student with a higher cognitive ability in the same cohort. The main identifying assumption is that the year of entry into a school is quasi-random, and cannot be influenced by the students or their parents. Obviously, in this set-up students are not randomly assigned, but given that the entry year into middle/high-school is mainly determined by a student's age, it is plausible to assume that the variation in the ability distribution across cohorts works like random assignment and produces a causal estimate. In the following, we will explain the estimation strategy, and discuss the plausibility of the identifying assumptions in greater detail.

3.1 Estimating equation

The following regression setup relates the educational attainment in wave IV of the survey (in 2008) to a student's ordinal rank in high-school measured in wave I (in 1994/1995):

$$\begin{aligned} \text{Educ. attainment}_{ijk} &= \gamma \text{ ordinal rank}_{ijk} + g(\text{cog. ability}_{ijk}) + \boldsymbol{X'_{ijk}\beta} \\ &+ \text{School FE}_j + \text{Grade FE}_k + \varepsilon_{ijk}. \end{aligned} \tag{2}$$

We consider the three outcome variables in separate regressions. The outcome variable of person i who attended high-school j and grade k is a dummy variable that takes value one if a person has achieved a certain educational attainment — dropped out of high-school, attended college, or completed college — and zero otherwise. The coefficient of interest is γ, which measures the impact of a marginal increase in the relative rank of a student within a high-school cohort on educational attainment.

Given that a person's ordinal rank is determined by her cognitive ability, the ordinal rank could be seen as a mere proxy for cognitive ability, in which case γ could be interpreted as the marginal effect of cognitive ability and not of ordinal rank. To ensure that γ exclusively measures the marginal effect of ordinal rank, we control for a person's cognitive ability with a fourth-order polynomial $g(\text{cog. ability}_{ijk})$, which captures the potential non-linear relationship between ability and educational attainment.

As shown in Table 1, the outcome variables differ considerably between demographic groups. For example, men have lower educational attainment than women, blacks have lower educational attainment than whites, and children of highly educated parents have a higher educational attainment. The vector of individual control variables $\boldsymbol{X_{ijk}}$ accounts for these differences and ensures that in our regression we compare students with the same observable characteristics. The controls include a dummy for gender, dummies for race (asian, black, hispanic), a dummy for migration background (1 if a person is a first- or a second-generation migrant), dummies for the highest level of education of both parents (less than high-school, high-school, some college, college degree), and dummies for the highest occupational status of both parents (not working, blue collar, white collar low-skilled, white collar high-skilled).

We also control for age, which is important because age effects could confound the estimate of γ, for example if older students within a cohort are at an advantage and therefore have a higher educational attainment later. Given that we have the exact date of the interview as well as the month and year of birth, we compute the age in months, allowing for variation in age within a birth year. Finally, we include a dummy that equals one if a student has ever repeated a grade until wave I of the survey. As shown in Table 2, repeaters are concentrated in the lower half of the ability distribution of their grade. If they also have lower educational attainment, not controlling for repeaters would lead to an upward-bias in the estimate of γ.

The inclusion of separate school and grade fixed effects ensures that we compare the outcomes of students within the same school across different grades. The school fixed effects remove the mean differences between schools in educational attainment, cognitive ability, as well as the demographic composition of schools. The grade fixed effects remove the mean differences in all variables between the six grade levels in our sample. Grade fixed effects are particularly important because of the mechanical positive correlation between grade in wave I and the likelihood of completing college in wave IV, which would bias the estimate of γ. Students who were in grade 12 in wave I are on average 30 in wave IV, and therefore have a higher likelihood of college completion than students who were in grade 7 in wave I and who were 25 in wave IV.

Finally, ε_{ijk} is an i.i.d error term that captures all unobservable factors that affect educational attainment. Due to the heteroskedasticity inherent in linear probability models, we use robust standard errors.

In a second set of estimates, we will replace the separate school and grade fixed effects by an interaction of both fixed effects. This accounts for changes in cohort quality within a school over time, or differences in the average peer ability across grades within the same school. The estimating equation then becomes

$$\begin{aligned} \text{Educ. attainment}_{ijk} &= \gamma \text{ ordinal rank}_{ijk} + g(\text{cog. ability}_{ijk}) + \boldsymbol{X'_{ijk}\beta} \\ &+ (\text{School} \times \text{Grade FE})_{jk} + \varepsilon_{ijk}. \end{aligned} \tag{3}$$

This set-up is more restrictive, as the identification of γ now relies only on differences in the dispersion of the ability distribution across grades within a school.

3.2 Identifying assumptions

The most basic confounding factor in estimating the effect of ordinal rank in high-school on educational attainment is self-selection into schools, be it because parents or students deliberately choose a certain school, or because students with similar characteristics live in the same neighborhood and go to the same school. The school fixed effects in Equation 3 account for systematic differences across schools that would lead to a spurious relationship between a person's rank and her educational attainment.

Within a given school, the identification of a causal effect rests on the assumption that being in a certain cohort is as good as random. This assumption only holds if at least two conditions are fulfilled:

1. Conditional on having chosen a specific school, neither parents nor students can manipulate the student's cohort.

2. Within a school, there is no systematic correlation between cohort characteristics and educational attainment.

The first assumption would be violated if parents could manipulate the rank by delaying their children's entry into a given school. While possible, how feasible would this be in reality?

Consider a parent whose child will soon reach the age of going to high-school. Within our quasi-experiment, their choice would be to send their child to high-school in the next school year, or to wait for another year. In order to know a child's rank in a *future* school grade of — on average — 184 students, the parents would need to know the ability of most other children that will also enrol in the same school. And even if they could perfectly foresee their child's rank, they would have to weigh the benefit of having a higher rank one year later against the cost of waiting for one more year, which would include one year of foregone earnings due to delayed access to the labor market, one more year of having to provide for the child, etc.

A more realistic scenario that could invalidate the first assumption is grade retention. If students repeat a grade, either because they did not have sufficient marks to progress, or because they repeated a grade voluntarily, they may end up in a different rank simply because they are older than the rest of the grade. We address this concern by controlling for a student's age in months, which absorbs the difference between older and younger students within a grade, and we include a dummy that equals one if a student has ever repeated a grade up to wave I of the survey.

The second assumption could be violated if the school quality changes over time. For example, if a school is "in decline", younger cohorts have a lower average ability in grade 7 than older cohorts had. A student with a given ability would then mechanically have a higher rank the younger she is. In that case, γ would be biased unless the cohort quality changes at the same rate in each school, in which case the grade fixed effects would capture the change. The model in Equation 3 would not account for school-specific changes in cohort quality. To demonstrate that the second assumption is not violated, we will estimate Equation 4, which includes *school* $\times$ *grade* fixed effects instead of separate school and grade fixed effects and show that the results are unchanged.

Including *school* $\times$ *grade* fixed effects addresses a further potential violation of assumption 2), namely that the results could be explained by differences in mean cohort ability, as suggested by most of the peer effects literature. Consider the two cohorts in Figure 1. A student with ability level $\underline{abil}$ has a higher rank if he enters the school in 1994. However, the entry cohort in 1994 had on average a lower ability, which is mechanically linked to a higher rank, and leads to a downward-bias in the estimate of γ.

A further bias could be due to measurement error in the relative ability rank. One source of measurement error is the over-sampling of minorities. If minorities are in the bottom half of the within-grade ability distribution, and if they are over-sampled compared to white Americans, then white Americans would be assigned a higher relative rank than under random sampling. The survey design offers an opportunity to assess the size of the measurement error through the sequencing of the sampling. First, a random sample was drawn and labeled as the *core sample*, and second, additional students were drawn from given minorities. Hence, we observe for each student in the sample the rank with and without over-sampling. The correlation in the relative ranks in both samples is 0.9867, which indicates that measurement error from oversampling is negligible.

4 Results

4.1 Ordinal rank and human capital investment

We now turn to the estimation of Equation 3. Table 3 displays the basic results for separate regressions of each of the three outcome variables — dummies for being high-school dropouts, having attended college, and having completed college — on the ordinal rank of a student in her high-school cohort. We begin with the unconditional relationship in Column (1), and gradually introduce fixed effects and control variables. Our preferred specification represents Equation 3, and is displayed in Column (5).

Table 3: OLS regression results: the importance of rank position

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
High-school dropout	-0.133***	0.053***	0.002	-0.054**	-0.034	-0.048*
	(0.008)	(0.013)	(0.022)	(0.023)	(0.023)	(0.026)
Attended College	0.386***	-0.106***	0.091**	0.139***	0.103***	0.112***
	(0.013)	(0.024)	(0.039)	(0.039)	(0.038)	(0.043)
Completed College	0.364***	-0.266***	0.073*	0.121***	0.102***	0.082*
	(0.013)	(0.024)	(0.040)	(0.041)	(0.039)	(0.044)
Controls:						
Individual ability (quartic)	No	Yes	Yes	Yes	Yes	Yes
School FE	No	No	Yes	Yes	Yes	No
Grade FE	No	No	Yes	Yes	Yes	No
School $\times$ Grade fixed effects	No	No	No	No	No	Yes
Individual controls	No	No	No	No	Yes	Yes
Goodness of fit:						
R^2 High-school dropout	0.02	0.04	0.08	0.09	0.15	0.17
R^2 Attended College	0.06	0.10	0.15	0.15	0.23	0.25
R^2 Completed College	0.05	0.12	0.18	0.18	0.26	0.28

Note: This table displays the results of separate OLS regressions of the dependent variables *high-school dropout*, *attended college*, and *completed college* on the relative rank. From left to right, more controls and fixed effects are being introduced. Robust standard errors are displayed in parentheses, with significance levels * $p<0.10$, ** $p<0.05$, *** $p<0.01$. The R^2 measures the goodness of fit for each regression.

The unconditional relationship in Column (1) strongly suggests that a higher within-grade rank is associated with higher educational attainment. An increase in the relative rank by one decile, that is, the difference between the second- and the third-best student in a grade of ten students, or the difference between the second- and the fourth-best in a grade of 20, is associated with a decrease in high-school dropout rates by 1.3 percentage points, which is 19% of the overall high-school dropout rate. The association with attending college and completing college is even larger. A one-decile increase in the relative ability rank increases the likelihood of going to college by 3.8 percentage points, which is 5% of the mean rate of college attendance, and increases the likelihood of completing college by 3.6 percentage points, which is more than 10% of the college completion rate in the sample.

While pointing to a strong association, the information we can get from Column (1) is limited, because the ability rank is based on the score on the ability test, and merely is a proxy for ability due to the strong positive correlation between rank and ability. In Column (2) we control for a fourth-order polynomial in individual ability, in which case the sign of the marginal effect of ordinal rank gets reversed. Taken at face value, the results suggest that the ordinal rank negatively affects educational attainment. This result may seem surprising, but it reflects a mechanical correlation between rank and school quality. At a given level of ability, a student in a school with a low average ability has a higher rank than in a school with a high-ability, but students from schools with a high average ability have a higher educational attainment.

In Column (3) we control for unobserved heterogeneity across schools by introducing school fixed effects. Identification now only comes from within the schools. Compared to the model without fixed effects, the R^2 is considerably higher, confirming the importance of unobserved heterogeneity across schools. The coefficient for high-school dropouts suggests that dropping out of high-school is not influenced by ordinal rank, while there is a positive association between ordinal rank and college attendance and completion.

Column (4) makes a leap towards a causal effect. While in Column (3) we estimated an average effect across all students within a school, in Column (4) we compare students with the same ability across different cohorts within the same school. We introduce grade fixed effects, which absorb the mean difference between different cohorts across the sample. If students who were in 7th grade in 1995 were on average different from those in 8th grade, this difference is accounted for in this specification. Compared to Column (3), the effects are larger and more precisely estimated. For all three outcome variables, these effects are substantial. An increase in the within-grade rank by one decile decreases the likelihood of dropping out of high-school by half a percentage point (7% of the mean), and increases college attendance and completion by 1.4 and 1.2 percentage points (2% and 3.6% of the mean), respectively.

In Column (5) we introduce individual control variables to take into account the differences in observable characteristics, and their potential effect on the outcome. There are two reasons for including control variables. First, as shown in Table 1, the outcome variables differ significantly across ethnic and parental backgrounds. Second, as indicated by the increased R^2 in Column (5), the control variables have additional explanatory power and ensure a better model fit. The inclusion of individual controls, however, has no statistically significant impact on the point estimates, which lends further credibility to our claim that cross-cohort variation in the ability distribution within the same school is quasi-random. The point estimates in Column (5) are slightly smaller than in Column (4), but the difference is not statistically significant.

Finally, we address the concern that the average grade ability, which is omitted in Equation 3, biases the estimates. In Column (6) we include $school \times grade$ fixed effects, taking into account school-specific average differences across grades. In this specification, we identify the impact of rank position only from differences in the dispersion of the ability distribution across grades within the same school. It is reassuring that the results are similar to those in the estimation with separate sets of fixed effects in Column (5). The differences between the coefficients in

Columns (5) and (6) are not statistically significant.

4.2 Heterogeneous effects

While the regression results in Table 3 show positive impact of high-school rank position on educational attainment, the strength of this impact differs across groups. In Figure 2 shows the marginal effects for various subgroups, which we obtained by re-estimating Equation 3 and including interaction terms of the relative rank with dummy variables for different subgroups, along with the subgroup dummies themselves.

We first consider the differences between male and female students. For high-school dropouts and college attendance, the effect is significantly stronger for men than for women. For college completion, in contrast, the result is the opposite; the impact of rank position on college completion among women is three times as strong for women than for men.

Next we check whether the impact differs along the ability distribution. The baseline estimate is a linear effect, which means that the rank difference between the second and third-best student has the same average effect on educational attainment as the difference between the 12th and 13th-best student. However, the effect could vary along the ability distribution within a grade. The ranking within a grade could work like a tournament, in which case the difference between the first and the second would matter more than between second and third, and so on. In the second row of Figure 2, we compare students with an ability in the top half of their grade to those in the bottom half. The effect for high-school dropouts is small and statistically insignificant. For college attendance and completion, we find larger effects in the top half. A higher rank is twice as important in the top half than in the bottom half.

The importance of ordinal rank may also differ by parental background. Children with highly educated parents may see their parents as role models and potentially know more about the benefits of going to college than children from parents with a low education. In that case, we should observe a larger impact for children from low-educated parents. The results in Panel D confirm this conjecture. The impact of ordinal rank on completing high-school is large among children from low-education households, while it is close to zero for children with at least one tertiary-educated parent. A similar pattern can be observed for college attendance. Both results are not surprising, given that among children with tertiary-educated parents only 2% drop out of high-school, while 85% attend some college. There is a difference, however, with respect to completing at least a Bachelor's degree. Here the effect is largest for children with tertiary-educated parents. These results suggest that for children from highly educated households the ordinal rank does not matter for the decision to attend college, but it matters for the type of degree program they choose, and the success in these programs.

4.3 Potential channels

The baseline results show a significant impact of ordinal rank in high-school on human capital investment later in life. The question remains which mechanisms can explain this reduced-

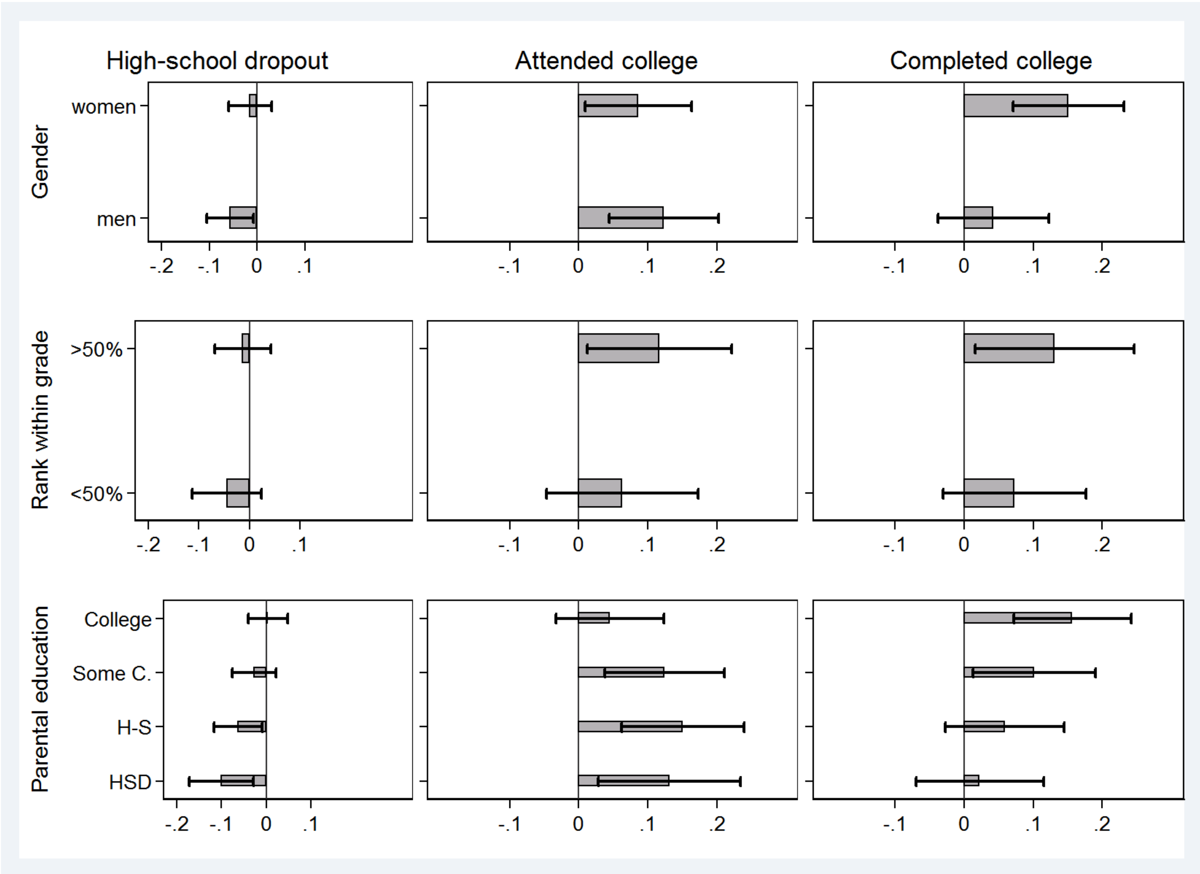

Figure 2: OLS results: heterogeneous effects

Note: Each graph displays the point estimates and 95%-confidence intervals for the marginal effects of rank on educational attainment for various subgroups.

Table 4: Regression results: rank position and intermediate outcomes, wave I

Dependent Variable	Coefficient	SE
Grades		
Grade Point Average (1-4)	0.178***	(0.055)
Self-concept		
1(I am more intelligent than the average)	0.090**	(0.044)
Expectations		
1(I want to go to college)	0.029	(0.040)
1(I will likely go to college)	0.082*	(0.043)
1(I will have a college degree by the age of 30)	0.106**	(0.044)
Questions on mental distress		
1(I was often depressed last week)	0.029	(0.028)
1(I was often fearful last week)	0.023	(0.018)
1(I was often hopeful last week)	0.110**	(0.045)
Support from others		
1(I was often happy last week)	-0.002	(0.039)
1(I feel that teachers care about me)	0.085*	(0.047)
1(I feel that parents care about me)	0.003	(0.033)
1(I feel that friends care about me)	-0.003	(0.017)
Perceptions of school environment		
1(I feel I am part of the school)	0.027	(0.042)
1(I am happy at school)	0.036	(0.044)
1(Teachers treat students fairly)	0.075	(0.047)
1(I was absent at school without excuse)	-0.116***	(0.040)

Note: This table displays the results for separate OLS regressions of the outcomes listed in the first column on relative ability rank within a school grade. Each outcome is a dummy variable with value 1 if an event occurred often or was very likely, and zero otherwise. All regressions include school fixed effects, grade fixed effects, and control for individual ability, age, minority dummies, and parental characteristics. Robust standard errors are displayed in parentheses, with significance levels * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

form relationship. If we think of a student making a rational choice, she invests in further education if the expected net gains from further are greater than zero. Students with a higher rank potentially have higher expected returns to education, because their rank within their peer group gives them a positive signal about their own ability. While we cannot directly measure the individual returns to education, we can explore the relationship between a student's rank and variables that proxy her expected returns to tertiary education. AddHealth contains several survey questions on a student's perceptions, expectations, and feelings which could indicate whether she expects to do well later on. In Table 4 we re-estimate Equation 3 with answers to these questions as outcome variables.

One potential channel highlighted in the psychology literature is self-concept (Marsh, 1987). Students with a higher rank in their peer group perceive themselves as good, even though they may be on the lower end of the overall ability distribution. We use as outcome variable a dummy that equals one if a student has a high self-concept, that is, if she agrees or strongly agrees to the statement "*I am more intelligent than the average.*" As expected, we find a positive and statistically significant effect of the ordinal rank on a student's self-concept.

Another useful proxy for expected gains from education could be a person's expectations in the first wave of the survey. Students were asked whether in the future they want to go to college, whether they think they will likely go to college, and whether they expect to have a college degree at the age of 30. We find that students with a higher rank have indeed higher expectations about going to college, and this effect is statistically significant for two out of three outcome variables.

The survey also includes questions on mental distress, which ask how often a student had a certain feeling in the week before the survey. To be sure, these questions primarily measure short-term feelings, but they could also pick up more longer-term feelings such as optimism or pessimism. We find no impact on feeling fearful, depressed, or happy, but we find a strong positive and statistically significant impact on feeling hopeful. This could indicate that students with a higher rank are more optimistic, which carries over to higher investment in education.

Students with a higher rank may also receive more or less support from their teachers, parents, and friends, be it actual or perceived support. If students perceive that they receive good support from their environment, they may be more likely to go to college later on. We find a large but imprecisely estimated effect on perceived support from teachers, while we find no effect for the perceived support of friends and parents.

Finally, we consider the broader school environment. If students feel that they are part of the school, that they are treated fairly by their teachers, and that they are generally happy at school, they may consider it worthwhile investing more in further education. Moreover, students are asked whether they were absent without excuse during the previous year, which indicates how seriously they are taking their school. We find a strong positive relationship between a student's rank and the perception of being treated fairly by the teachers, and a strong negative relationship between rank and absences without excuse. Students with a higher rank put in more effort, which could materialize in their educational career later on. We find no effect, however, that students with a higher rank are happier at school, or that they feel more that they are part of the school.

In sum, these results suggest that a student's ordinal rank within a grade affects her expected returns to education. The comparison with immediate peers provides students with a signal about their own ability. If a student is doing well compared to her peers, she seemingly has higher expected returns to education, which can explain why we find a positive impact of ordinal rank on educational attainment.

5 Conclusion

This paper shows that a student's rank in the ability distribution within a high-school cohort is an important determinant for educational attainment later in life. If Jack and Jim have the same ability, but Jim is the brightest student in his grade, while Jack ranks in the middle of his grade, our results predict that Jim is more likely to get a college degree and less likely to drop out of high-school than Jack. This effect runs counter to most of the literature on peer effects, which finds that being exposed to high-achieving peers has a positive effect on educational attainment.

The results should concern parents and policymakers alike. Parents could derive from this study that it is better to send their child to a school in which he or she has a higher rank, that is, it is better to send a child to a school with lower-ability peers. However, our results reflect local effects, which we obtained by comparing students within the same school but in different cohorts. If parents chose schools based on their children's rank, such a behavior would be problematic, however, because the positive effect of having a higher rank could be compensated by a lower peer quality, and generally a lower school quality. Moreover, if all parents choose schools according to their children's rank, this would result in a difficult choice problem, and the the general equilibrium outcomes would be far from clear.

Policymakers should be concerned as well, because the results suggest that the selection into schools and the transition into college leads to a low-level equilibrium: if parents try to send their kids to the best possible schools, and if a child's rank within the school is important for educational attainment, potentially fewer students complete college than would be optimal given their ability. For a government, this underinvestment in human capital is not optimal. Given that the ordinal rank depends on the mean ability as well as on the ability distribution within a class, it is difficult to think of an effective algorithm that changes the ability composition of schools in order to encourage more investment in human capital. A potentially more efficient policy would be to give more support to students at lower ranks of the ability distribution, in order to compensate for the negative impact of their rank.

References

Ammermueller, Andreas, & Pischke, Jörn-Steffen. 2009. Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, **27**(3), 315–348.

Angrist, Joshua, & Lang, Kevin. 2004. Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *American Economic Review*, **94**(5), 1613–1634.

Bifulco, Robert, Fletcher, Jason M., & Ross, Stephen L. 2011. The Effect of Classmate Characteristics on Post-Secondary Outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, **3**, 25–53.

Booij, Adam, Leuven, Edqin, & Oosterbeek, Hessel. 2015. Ability Peer Effects in University: Evidence from a Randomized Experiment. *IZA Discussion Paper*, **8769**.

Calvó-Armengol, Antoni, Patacchini, Eleonora, & Zenou, Yves. 2009. Peer Effects and Social Networks in Education. *Review of Economic Studies*, **76**, 1239–1267.

Carrell, Scott E., Fullerton, Richard L., & West, James E. 2009. Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, **27**(3), 429–464.

Carrell, Scott E., Sacerdote, Bruce I., & West, James E. 2013. From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica*, **forthcoming**.

Clark, Andrew, Masclet, David, & Villeval, Marie-Claire. 2010. Effort and Comparison Income: Experimental and Survey Evidence. *Industrial and Labor Relations Review*, **63**(3), 407–426.

Dunn, Lloyd M., & Dunn, Leota M. 2007. *The Peabody Picture Vocabulary Test*. 4th edn. Bloomington, MN: NCS Pearson, Inc.

Feld, Jan, & Zölitz, Ulf. 2014. On the Nature, Estimation and Channels of Peer Effects. *University of Gothenburg Working Papers in Economics*, **596**.

Hanushek, Eric A., Kain, John F., Markman, Jacob M., & Rivkin, Steven G. 2003. Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics*, **527-544**.

Harris, Kathleen Mullan. 2009. The National Longitudinal Study of Adolescent to Adult Health. *doi: 10.3886/ICPSR27021.v9*.

Harris, K.M., Halpern, C.T., Whitsel, E., Hussey, J., Tabor, J., Entzel, P., & Udry, J.R. 2009. The National Longitudinal Study of Adolescent to Adult Health: Research Design [WWW document]. *URL: http://www.cpc.unc.edu/projects/addhealth/design*.

Imberman, Scott, Kugler, Adriana D., & Sacerdote, Bruce. 2012. Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees. *American Economic Review*, **102**(5), 2048–2082.

Lavy, Victor, Silva, Olmo, & Weinhardt, Felix. 2012. The Good, the Bad and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics*, **30**(2), 367–414.

Luttmer, Erzo F. P. 2005. Neighbors as Negatives: Relative Earnings and Well-Being. *The Quarterly Journal of Economics*, **120**(3), 963–1002.

Marsh, Herbert W. 1987. The big-fish-little-pond effect on academic self-concept. *Journal of educational psychology*, **79**(3), 280.

Murphy, Richard, & Weinhardt, Felix. 2014. Top of the Class: The Importance of Ordinal Rank. *CESifo Working Paper*, **4815**.

Patacchini, Eleonora, Rainone, Eduardo, & Zenou, Yves. 2012. Student Networks and Long-Run Educational Outcomes: The Strength of Strong Ties. *CEPR Discussion Paper*, **9149**.

Sacerdote, Bruce. 2001. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, **116**(2), 681–704.

Zimmerman, David J. 2003. Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics & Statistics*, **85**(1), 9–32.

A Disclaimer

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.