Beyond Test Scores: the Rank Effect and Non-Cognitive Skills

Emilia Del Bono[1], Angus Holford[1], Tommaso Sartori[1]

[1]Institute for Social and Economic Research, University of Essex

Last Updated: June 9, 2024

Abstract

Making comparisons is key to our understanding of the world, which naturally leads to the creation of rankings. Students naturally rank themselves based on how they perform in school. While a high rank consistently predicts better future scores, the reasons are unclear. We followed children in all the children in primary school in the city of Aberdeen (Scotland) to see how non-cognitive skills were affected by rank. We find that a higher rank affects both academic performance and internalizing skills, which relate to self-esteem and academic self-concept. Our findings suggest that there is also a rank effect on boys' externalizing skills, indicating an improvement in their self-control as well.

Keywords: Peer Effects, Non-Cognitive Skills, Primary Education

JEL Classifications: I21, J24

1 Introduction

As humans, we rely on comparison to define the characteristics and properties of what surrounds us. Our understanding of the world builds on identifying similarities and differences, and how certain features relate to each other. When it comes to social interactions, comparing ourselves to others can be a double-edged sword, fueling motivation and growth but potentially leading to insecurity and self-doubt. Schools are an ideal setting to study not only interactions but social comparison itself. Students are grouped and engage with each other frequently and for a prolonged period. Moreover, they are constantly evaluated through test scores, providing an actual metric that can be used for the purpose of ranking.

Decades of research on the impact of peers on academic outcomes highlighted positive but moderate effects, regardless of the structure of the model adopted.[1] Sacerdote (2011) reviews the existing literature comprehensively. More recently, following the footsteps of the "big fish in a little pond" theory (Marsh and Parker, 1984), economists focused on a different aspect of the interplay between peer characteristics and individual test scores, their ordinal rank.

Rank is a feature that emerges naturally when people are grouped. Especially in school, as students spend a long time with their colleagues, and they are also constantly evaluated. Researchers started using test scores previously attained to construct the rank of individual students within their peer group (generally classroom or cohort) and, conditional on the test score itself (used as a baseline measure of cognitive skills), estimate the rank effect on future academic performance and other outcomes.

The findings consistently showed a positive and significant impact of ordinal rank on future achievements and educational attainments (Elsner and Isphording, 2017, 2018; Murphy and Weinhardt, 2020; Elsner et al., 2021; Pagani et al., 2021; Megalokonomou and Zhang, 2022; Carneiro et al., 2023; Denning et al., 2023). It also affects major choice (Delaney and Devereux, 2021; Goulas et al., 2022), and future earnings (Denning et al., 2023).

Naturally, estimating the rank effect conditional on the test score used to build the ranking implies that the effect does not convey any information on individual cognitive skills. If that were the case, predicting a future test score using class or cohort rank and the test score on which the ranking was built should lead to estimating an insignificant rank effect coefficient. That is because rank simply shows a student's standing compared to the others, not an actual measure of knowledge, understanding, or problem-solving skills. The fact that an impact is measured repeatedly across different schooling systems and levels of education requires us to think about the possible channels of the effect. In their excellent literature review, Delaney and Devereux (2020) highlight two: what rank could do to intrinsic beliefs and behavior, and what rank could do to factors in the external environment that respond to a student's rank.

Previous research finds overwhelming support for the first channel. Pagani et al. (2021) show that rank affects conscientiousness (out of the five personality traits). That implies that children with a higher rank, everything else equal, build skills such as persistence through hardships, self-discipline, and impulse control. Murphy and Weinhardt (2020) find a positive rank effect on self-confidence, while Elsner and Isphording (2017); Elsner et al. (2021) estimate a positive impact on student

[1] Studies moving beyond the linear-in-means model provide a richer understanding of peer effects by capturing heterogeneity within peer groups, accounting for selection effects, and exploring how peer influence might evolve. These studies often find more nuanced and sometimes stronger effects compared to the linear-in-means model. However, the differences are not large enough to justify dismissing the linear-in-means model entirely.

expectations for future grades and careers. Moreover, Carneiro et al. (2023) show that rank improves student executive function, a non-cognitive skill that is positively related to academic performance and learning, self-regulation and emotional control, as well as focus[2].

We will examine two mechanisms related to the relationship between rank and student beliefs and behavior. We name them the "self-control" and "self-concept" hypotheses, respectively. We do not consider them mutually exclusive. Our goal is to understand whether the rank effect operates mainly through one of the non-cognitive skills related to one of the two.

The self-control theory implies that the positive rank effect on future test scores is determined by how ranking higher affects non-cognitive skills such as self-discipline and impulse control. In this framework, the higher the rank, the more the students adapt their behavior, reducing those that are disruptive to learning. On the other hand, the self-concept hypothesis entails that a higher rank shapes student self-perceptions, improving their confidence in their abilities and expectations for their future.

The available evidence does not distinguish between these two hypotheses, and its findings are mixed. The positive rank effect on conscientiousness and test scores found by Pagani et al. (2021) is attributable to the self-control channel. On the other hand, the results highlighting a positive rank effect on self-confidence (Murphy and Weinhardt, 2020) and expectations (Elsner and Isphording, 2017; Elsner et al., 2021) point in the direction of the self-concept idea. The improvement in executive function shown by Carneiro et al. (2023) does not help to disentangle these two mechanisms.

In this paper, we do so by linking the self-control hypothesis to the existence of rank effects on externalizing skills, and the self-concept hypothesis to rank effects on internalizing skills. Externalizing and internalizing skills are established psychological constructs that capture two different types of non-cognitive skills (Eisenberg et al., 2001; Creemers et al., 2013). In a school context, the former refer to a student's outward expression of emotions and behaviors. They are often associated with hyperactive behavior, impulse control, difficulties in following rules and restrictions, and social conflicts. Low externalizing skills lead to behaviors that disrupt the learning environment for both the peers and the individual. The latter are related to the ability to manage emotions and thoughts internally. They encompass aspects like self-regulation (managing anxiety or frustration healthily), coping skills, and self-esteem. We argue that looking at the divide between externalizing and internalizing skills allows us to examine the two channels effectively.

We do not focus on mechanisms depending on how rank affects the external environment (possibly through shaping teacher and parental investment). Our data do not allow testing these potential channels precisely. Moreover, several studies show no evidence of teachers' responses (Elsner and Isphording, 2017, 2018; Murphy and Weinhardt, 2020; Pagani et al., 2021). Regarding parents, most of the literature does not find any influence of rank on parental investment (Elsner and Isphording, 2017, 2018; Murphy and Weinhardt, 2020; Pagani et al., 2021). Only Megalokonomou and Zhang (2022) show that this happens in a context where rank is salient among parents. However, our setting is quite different.

To get a better grasp of the mechanism behind the rank effect, we exploit a population study that gathered data on all the children in the city of Aberdeen (Scotland, United Kingdom) between grades 3 and 7 in 1962. More than 10,000 children are included in our sample, born between October 1950 and September 1955. Information on the children related to their journey through primary schools,

[2]Executive function is a set of mental skills that act like the "control center" of your brain. It helps you manage various cognitive processes to achieve goals and navigate complex situations effectively.

the standardized test scores outcomes, as well as their school and grade. Moreover, the survey required teachers to fill out the Rutter Questionnaire (Rutter, 1967), an established psychological test, used to screen children for minor behavioral issues. We exploit the questionnaire to derive measures of externalizing and internalizing skills. In addition, a follow-up survey ran in 2001, which has information on roughly 60% of the children in the original sample, allows us to look at long-term outcomes.

We use a standardized test score attained at 9 years of age to construct the rank within the school-cohort groups to which the children belong. Exploiting the month of birth cutoff that determined assignment to school-cohort groups, we estimate the rank effect on the future test scores, our two measures of non-cognitive skills, educational attainment, and mental health in 2001.

We find a positive and statistically significant rank effect on a high-stakes test at age 11. A jump of 4 positions in the ranking, if we consider the average group size (40 students), implies an increase of 6% of a unit of standard deviation in the test. The effect size is very similar to the one reported by Murphy and Weinhardt (2020) and Elsner et al. (2021). We show that externalizing skills, which we associate with the "self-control" hypothesis, are not affected by rank. On the other hand, ranking 4 positions higher leads to growth of internalizing skills of roughly 6% of a unit of standard deviation. Internalizing skills are related to the "self-concept" hypothesis. These effects increase linearly with rank. They are robust to changes in specification and other robustness exercises. We also show rank affects educational attainment, increasing the probability of obtaining the Scottish equivalent of an O-level and A-level qualification. However, there is no impact on the probability of completing higher education. The improvement in educational attainment mediates the long-run effect on mental health, which we measure almost 40 years later.

While the "self-concept" mechanism seems to be prevalent, it can work together with "self-control", at least when it comes to boys' academic performance. While ranking improves girls' internalizing skills, it seems to have an impact, albeit smaller, on boys' externalizing and internalizing skills.

Our main contribution is a deeper understanding of the channels behind the rank effect. To our knowledge, the characterization of non-cognitive skills along the externalizing/internalizing dimension has not been examined in the rank effect literature. While we recognize the drawback of being unable to look at individual non-cognitive skills, we believe our approach has the advantage of not forcing us to look at individual components of a set of intertwined elements. There is no doubt that internalizing skills are a multidimensional construct involving self-confidence, academic self-concept, and expectations for own abilities. We believe that focusing on the externalizing/internalizing divide allows for a broader view of how the rank effect works, and a better understanding of the interplay between the two mechanisms.

We contribute to the literature on the long-term influence of non-cognitive skills on life outcomes from two different sides. On one hand, we show that these non-cognitive skills can persist throughout the life cycle (Attanasio et al., 2020b,a). In fact, we show that there is an impact of the rank effect on self-reported mental health almost 40 years later, mediated mainly by additional educational attainment (which can indeed be attributed to the rank effect). This fits with those studies showing how increasing non-cognitive skills through intervention during childhood can have long-lasting benefits (Heckman and Kautz, 2012; Heckman et al., 2013; Sorrenti et al., 2024). However, we do not find any evidence supporting that there is an impact on socioeconomic status in the long run.

2 Data

2.1 The Aberdeen Children of the 1950's Survey

The "Aberdeen Child Development Survey"[3] involved the children between grade 3 and grade 7 in the city of Aberdeen, Scotland, in December 1962[4]. It aimed to understand the relationship between anthropometric measures and reading disabilities. While the survey focus was not school *per se*, the available information allows us to deduct the children's path through primary school.

The data collection for this "Reading Survey" involved different stages. In December 1962, teachers helped the children provide detailed demographic information (date of birth, father's occupation, family size, etc). Each school provided attendance records of the children for the previous two years and the test scores available at the time (those missing were added later). Moreover, the hospital records allowed access to obstetrics and social information collected during the mother's pregnancy.

At the end of the same school year, in July 1963, schools provided additional information on measures (for example, height and weight) collected in medical exams held when children entered school (approximately at 5 years of age), and then again when they were 9 and 12 years old.

In the final stage of the survey, in March 1964, teachers filled out questionnaires to pin down the children's behavior. The first was the "Rutter Questionnaire (scale b)" (Rutter, 1967), an established psychological test used to screen for minor behavioral disorders.

A random sample of 25% of the children in the Reading Survey was selected for a "Family Survey". Overall, the response rate was 80%. Their mothers (or a substitute) were asked about the children's health and behavioral conditions, time allocation between leisure and school activities, and further information on the parent's demographic characteristics.

Finally, in 2001 the children participating in the original survey were mailed a set of follow-up questions, on their health, psychological, and socioeconomic status. Roughly 60% of those who received the questionnaire returned it fully completed.

2.2 The Grade Progression Mechanism

Since the cross-sectional structure of the data does not allow us to observe how students progressed from one grade to the next, we need to rely on the available historical information to reconstruct the mechanism[5].

We know that children born between the beginning of April and the end of September would enter primary school in the August intake[6]. Those born between the beginning of October and the end of March would start school in January[7]

Children entering primary school in different intake windows would remain separated from each other until the end of primary school. In other words, two different cohorts of students would enter

[3] Now known as "The Aberdeen Children of the 1950's"

[4] All the children enrolled in the Primary Schools of the Aberdeen Municipality that respect the grade cutoff are included in the sample.

[5] Based on the statement by the Director of Education to a Town Council of Aberdeen meeting that took place on 3rd October 1960 (Lawlor et al. (2006)) we know that schools had two or more admission dates every academic year. All the students who became five years old before the next admission date had to attend school from the first school day following that admission date.

[6] Some of the October-born were admitted to make classes of a reasonable size.

[7] The children born between the beginning of January and the end of March would be admitted with their parent's permission. Still, the rest of those born between the beginning of January and the end of March would join those same classes that started in January in April.

primary school each calendar year. Table 1 illustrates how, despite being assigned to the same grades, students were separated based on their intake. Focusing on grade and not on assignment based on the combination of grade and intake would imply mixing children at different moments of their primary school journey. We call the combination of the grade and intake variable **cohort**. We calculate students' rankings within groups defined by the combination of school and cohort.

2.3 Sample Definition

The sample size of the dataset consists of 12,151 observations, one for each student born in Aberdeen between the beginning of October 1950 and the end of September 1955. We exclude all children enrolled in private schools or special schools, consisting of 2.88% and 9.5% of the sample, respectively[8]. We also exclude all children enrolled in schools that ceased to exist before 1964, and those in a specific institution that has very few students in its cohorts. Finally, we keep only those students for whom we can classify their cohort in December 1962. That leaves us with 10,073 observations.

We are left with a total of 28 schools in our sample. In Table 2 we list the characteristics of schools, groups, and children in the sample. The median size of the schools is 300 students. Roughly 20% of these schools enroll less than 200 students, while two of them, on the other side of the spectrum, have around 700. Still, the information we are more interested in is group size. As we stated, that is defined as the combination of school, grade, and intake. On average, there are 40 students in each group. Some school-cohort groups involve up to 100 students, while the smaller ones include slightly more than 10. The median size is around 35 children per group.

We also exclude certain children from our estimation sample based on the outcome variable examined and the expected cohort in December 1962. To capture the rank effect on the outcome of the 11-plus Test we consider all the students born between the beginning of October 1950 and the end of September 1955. In other words, 10 cohorts (9,985 children). We also exclude those children in a school-cohort group with less than 10 students, leaving 9,969 observations in our sample.

When we study the rank effect on non-cognitive skills, we drop the three oldest cohorts, as they were already either in Secondary School or Junior Secondary School in March 1964 (6,779 children). In March 1964, when teachers completed the Rutter Questionnaire, those two cohorts were already in secondary school.

2.4 Standardized Tests: an Outline

Scottish schools administered a standardized test to students within 6 months of their 9th birthday. The goal of the test was both to offer a way of comparing students across different areas of the country and screen for poor reading ability. That was achieved through the "Schonell and Adams Essential Intelligence Test (Form B)" (Schonell and Adams, 1940). The test consists of a battery of 100 questions, focusing on both verbal reasoning and arithmetic. It was administered such that children in different cohorts would take the test at the same relative age[9]. We will refer to this as the "**Age 9 Test**", and use it to build the within-cohort ranking and as a baseline measure of cognitive skills. We show the standardized distribution of the Age 9 Test by cohort in Figure 1. The

[8]Private schools refused to have their teachers administer the Rutter Questionnaire, which is problematic for our analysis. We do not know enough about how special schools worked to understand if they could be compared to the other public schools in the sample.

[9]Students who entered primary school in January would take the test in November, while those who started school in August, would take the test in May.

test scores do not follow a standard normal distribution. The Kolmogorov-Smirnoff test shows that the two cohorts born between April 1951 and 1952 have different distributions, while the t-test is insignificant for the others.

Another standardized test characterized the end of primary school. Its outcome determined admission to grammar schools. Being able to enter a grammar school had a substantial impact on educational attainments, as well as on women's labor market outcomes (Clark and Del Bono, 2016). The test was held in November, for kids aged between 11.1 to 12.0 years old. Scores were then standardized using national norms and for the city of Aberdeen, by year. Students would take four different tests: two verbal reasoning tests, one arithmetic test, and one English test[10]. In the rest of the paper, we will refer to this as the "**11-plus Test**". We show the standardized distribution of the Verbal Reasoning Test by cohort in Figure 2. The Verbal Reasoning Test was available for all cohorts in our data, so use that as our main outcome variable[11]. The distributions are very similar to a standard normal and do not differ much from one another. We run a Kolmogorov-Smirnov test and observe that all the cohorts are statistically similar, except the one involving students born between October 1954 and April 1955.

2.5 Non-cognitive Skills: The Rutter Children's Behaviour Questionnaire (scale b)

In March 1964 teachers completed the Rutter Children's Behavior Questionnaire (Rutter, 1967) for each child in their class[12]. The Rutter scale was developed to identify children with minor behavioral issues. The questionnaire consists of 26 questions (or items) that aim to isolate features of the children's behavior[13]. The teachers had to state whether a specific behavior "Does not Apply"/"Somewhat Applies"/"Definitely Applies" to each of the children. Since its development, the questionnaire has been one of the most widespread behavioral assessment tools in a classroom environment (Behar and Stringfield, 1974; Boyle and Jones, 1985; McGee et al., 1985; Iloeje and Meme, 1992; Klein et al., 2009; Narusyte et al., 2017). We discuss the details of the Rutter Questionnaire extensively in section 4.

2.6 Defining Rank

We use a standardized test score to construct the ranking of the students within their school-cohort group. We follow Murphy and Weinhardt (2020) approach and normalize rank by group size to make them comparable:

$$RANK_{isgk} = (n_{isgk} - 1)/(N_{sgk} - 1) \tag{1}$$

Where n_{isgk} is the ordinal rank of individual i enrolled in school s, in grade g, and who entered in

[10] The teacher's estimate of the student's score in the test is also reported, and it is scaled to match the mean and standard deviation of the class's actual performance

[11] We will also show the rank effect on outcomes that are not harmonized across years.

[12] Two versions, or scales, exist a and b. They are questions asked and how answers are recorded is nearly identical. However, the former is meant to be filled out by the mother of the children, the latter by their teacher.

[13] The 26 items in the questionnaire are the following: "Restless", "Truant", "Fidgety", "Destroys Belongings", "Fights other Children", "Disliked", "Anxious", "Solitary", "Irritable", "Often Unhappy", "Tics", "Sucks Fingers", "Nail Biting", "Trivial Absences", "Disobedient", "Short Attention Span", "Fearful", "Fussy", "Often Lies", "Stealing", "Wet/Soiled Themselves", "Often Aching/in Pain", "Tearful", "Stutters", "Other Speech Difficulty", "Bullies other Children".

intake k. while N_sgk is the size of the school-cohort group to which the student belongs. When we have ties between students, the mean rank is assigned. This ensures that the average rank of a group with ties is the same as the average rank of a group with no ties, while not breaking ties arbitrarily.

3 Empirical Strategy

We aim to estimate the rank effect on future academic achievements (the 11-plus Test) and non-cognitive skills (externalizing and internalizing skills). Our empirical strategy is informed by the recently developed literature on the rank effect (Murphy and Weinhardt, 2020; Elsner et al., 2021; Carneiro et al., 2023; Denning et al., 2023). We start by describing the identification strategy, clarifying our assumptions, and illustrating the identifying variation we exploit.

3.1 The Identification Problem

The rank of an individual i within a group g, R_{ig} can be considered a function of the student's cognitive skills A_{ig} and group characteristics $\overline{W}_g$:

$$R_{ig} = f(A_{ig}, \overline{W}_g) \tag{1}$$

Assuming that the two components are additively separable[14], we can estimate the rank effect on an outcome Y_{ig}, exploiting the variation in rank that occurs when two identically able students end up in groups with different $\overline{W}_g$ distributions, conditional on their ability. In this situation R_{ig} and A_{ig} are not perfectly collinear, making α identifiable:

$$Y_{ig} = \alpha R_{ig} + \beta A_{ig} + \epsilon_{ig} \tag{2}$$

The identifying variation comes from comparing students with similar cognitive skills, but different within-group ranking. To account for the potential bias that unobserved group effects could cause, we could include group-level dummies, λ_g, and estimate:

$$Y_{ig} = \alpha R_{ig} + \beta A_{ig} + \lambda_g + \epsilon_{ig} \tag{3}$$

The rank effect in Equation 3 is identified if rank is uncorrelated with unobserved features of the group and of the student, which is equivalent to the following conditional independence assumption: $E[\epsilon_{ig}|A_{ig}, R_{ig}, \lambda_g] = 0$. It exploits the comparison of students with different ranks within the same group, accounting for the impact of their cognitive skills. In other words, all the common variation for students in a given group is accounted for.

The rank effect literature widely adopts this solution, as it allows to partial out some of the variation spurring from students sorting into groups. However, it also implies that Equation 3 is identified only out of functional form assumption if λ_g accounts for all the variation in $\overline{W}_g$. The α coefficient will not be identified unless higher-order individual cognitive abilities and group effects which are not constant for all the groups we are comparing exist. In other words, we cannot expect to identify a rank effect if we are looking at identical groups, as we would not find any difference in rank.

We believe it is more interesting, at least in our context, to compare students across groups. Our identification strategy exploits exactly this kind of variation.

[14] Without forcing a structure in the relationship between the individual- and group-driven effects we would not be able to estimate the rank effect. Such an assumption is required whenever we are studying peer effects, as features of the group distribution cannot arbitrarily interact with others.

3.2 Our Set Up

We adapt the model to our research setting. We have individual children i, enrolled in school $s \in [1, ..., 28]$, in grade $g \in [3, ..., 7]$ (as in, all the children born over the 12 month-period which defines grade assignment, hence from October of year t until October of year $t+1$), that enter school with intake $k \in [1, 2]$ (where 1 stands for the January intake and 2 stands for the August intake). The group now is given by each intake, within a grade, within a school. We refer to the combination of grade and intake as cohorts $c \in [1, ..., 10]$, which we will use as our group unit. We rewrite Equation 1 accordingly:

$$R_{isc} = f(A_{isc}, \overline{W}_{sc}) \tag{4}$$

Now rank is defined as R_{isc}, and characterizes student i, in school s, and cohort c. It is a function of individual cognitive skills A_{isc} and group characteristics $\overline{W}_{sc}$. We could estimate the within-cohort rank effect by relying on the following adaptation of Equation 3:

$$Y_{isc} = \alpha R_{isc} + \beta A_{isc} + \lambda_s + \lambda_c + \epsilon_{isc} \tag{5}$$

Where the variation that allows us to identify the rank effect has two sources:

A) Children with identical cognitive skills, who are in the same school, but belong to different cohorts.

B) Children who are in the same cohort, but are enrolled in different schools.

For the rank effect to be identified, we need the following conditional independence assumption to hold: $\mathbb{E}[\epsilon_{isc}|R_{isc}, A_{isc}, \lambda_s, \lambda_c] = 0$. The assumption could fail if the outcomes that we study are correlated with factors that are specific to a group (meaning a school-grade-intake group), or if peer characteristics that affect our outcomes and are different from rank differed across grade-intake groups within the same school. We will show that this is not the case in our sample. An advantage of maintaining this stronger assumption is that our identification will not rely on a within-group comparison, which we believe reveals more of what happens to two similarly capable students assigned to different groups.

Our final specification adds a few more elements. We want to make sure that the rank effect coefficient is not capturing any higher order peer or ability effect[15]. We add a quadratic polynomial of individual cognitive skills $g(A_{isc})$. Moreover, we include the mean and standard deviation of the group cognitive skills distribution, $\overline{A}_{sc-i}$ and σ_{sc-i} (calculated as the mean and standard deviation of the children's test score within their school-cohort group). We also include a vector of individual characteristics X_i (which accounts for sex, socioeconomic status, and month of birth). The equation we estimate is the following:

$$Y_{isc} = \alpha R_{isc} + \beta g(A_{isc}) + \gamma_1 \overline{A}_{sc-i} + \gamma_2 \sigma_{sc-i} + X_i \delta + \lambda_s + \lambda_c + \epsilon_{isc} \tag{6}$$

[15] Denning et al. (2023) provide an insightful discussion on this issue. They compare the estimation of the rank effect to an exclusion restriction. Once the common impact of being in a certain group is accounted for, and the groups we are comparing are sufficiently similar, the remaining variation in student outcomes (given their cognitive skills) spurs from peer effects related to rank. While we can corroborate this claim with several robustness checks, it remains an assumption that needs to be discussed on a case-by-case basis.

The choice to include quadratic polynomials of individual cognitive skills is not only to account for higher-order effects. If these were mapped to future outcomes differently our model would be misspecified. We want to avoid estimating biased coefficients due to this specification error. The polynomials introduce flexibility in the specification and allow accounting for potential nonlinearities. Including the mean and standard deviation of the group cognitive skills distribution forces a comparison between those groups that are similar to each other, reducing the chance that the rank coefficient will capture group-specific variation attributable to cognitive skills.

As we expect the variation in rank to occur at the group level, all our equations are estimated clustering the standard errors at the school-grade-intake level (or group level).

3.3 Evidence on the Validity of the Identifying Assumption

Believing that our model can disentangle the rank effect from other types of peer effects is essential, but not sufficient, for our identification strategy to be successful. A key fact we need to establish is that assignment to groups was quasi-random. The month of birth cutoff is insufficient to ensure that is the case. It makes the variation of cognitive skills across cohorts idiosyncratic but does not account for students potentially sorting into schools. The literature separates between active and passive sorting. The former involves children (or, more realistically, parents) selecting their peer group based on their rank preference. The latter implies that children with certain characteristics could be more likely grouped (if, for instance, school choice is non-random, as in our setting). Considering that children are assigned to a certain intake based on their month of birth, we can safely rule out active sorting. On the other hand, passive sorting remains a concern Denning et al. (2023).

We use two tests to diagnose whether passive sorting exists in our setting. First, we estimate the relationship between individual characteristics and rank in the school-cohort group. We regress individual characteristics on percentile rank, conditioning on a quadratic polynomial of cognitive skills, and school and cohort fixed effects. The dependent variables we use are sex at birth (male or female), socioeconomic status (high or low), height and weight at the time of the first medical exam, birth weight, and number of siblings.

$$X_i = \alpha R_{isc} + \beta g(A_{isc}) + \lambda_s + \lambda_c + \epsilon_{isc} \tag{7}$$

The top panel of Table 3 shows the estimated coefficient α for Equation 7. We do not find any conditional relation between individual characteristics and rank. However, the coefficient for weight is significant at the 10% level, while the one for height is just barely insignificant. These coefficients are not necessarily problematic, and controlling for the two in our main equation would address the sorting concerns. An additional test can provide more information on whether sorting is happening.

Ranking is a function of the distribution of cognitive skills within the group. As such, an association between individual characteristics and features of the distribution of peer cognitive skills $f(\overline{A}_{sc-i})$ can inform us on whether children with certain characteristics are more likely to end up in certain groups. In particular, we compute the mean, standard deviation, and quartiles of the student-specific leave-out peers' outcome in the Age 9 Test. Again, we include a quadratic polynomial of individual cognitive skills and school and cohort fixed effects as controls.

$$X_i = \gamma f(\overline{A}_{sc-i}) + \beta g(A_{isc}) + \lambda_s + \lambda_c + \epsilon_{isc} \tag{8}$$

The panels of Table 3 referring to Equation 8 report estimates of γ. These results rule out any (conditional) relationship between peer cognitive skills and individual characteristics, suggesting that passive sorting is not an issue in our setting. It corroborates our assumption of quasi-random assignment to school-cohort groups.

4 Non-Cognitive Development

4.1 Externalizing and Internalizing Skills

We exploit the Rutter Questionnaire (Rutter, 1967) for teachers to derive our measures of non-cognitive skills. That is based on 26 items (questions), chosen to capture two types of behaviors: externalizing and internalizing. The distinction comes from a well-known categorization used in child psychiatry to describe behaviors associated with psychological disorders (Achenbach and Edelbrock, 1978). While externalizing behaviors are directed outwardly and reflect on actions in the physical environment, internalizing disorders are directed inward and indicative of a child's psychological and emotional state (Eisenberg et al., 2001). The former include problems of inadequate regulation, such as the inability to inhibit behavior, control attention, and cognitive processing. Aggressive and hyperactive behaviors are an example. The latter are such problems as high levels of rumination, sadness, anxiety, and depression, but involve low impulsivity as well.

Moving away from the cases of severe behavioral disorders we can use the scale to measure the counterpart of those and consider externalizing and internalizing skills (Attanasio et al., 2020b). Externalizing skills are related to the ability to engage in interpersonal relationships and focus attention on a task. On the other hand, internalizing skills refer to the ability to channel your skills to perform a task. To put the latter in the context of school, it can be intended as being aware of your skills and using them proficiently. As we discussed in subsection 2.5, all the items refer to questions describing how often children exhibit certain behaviors[16].

4.2 Extracting Measures of Non-Cognitive Skills: Factor Analysis

We rely on principal (or common) factor analysis to isolate the two underlying constructs. That statistical technique allows capturing the joint variation observed with different variables and estimating the values of the latent constructs that better approximate said variation. As the items in our questionnaire are discrete, we rely on a polychoric correlation matrix to estimate unbiased correlations between the items (Olsson, 1979). To decide on the number of factors that provide the best description of the variation in our data, we use the theoretical insights offered by the psychology literature (Behar and Stringfield, 1974; Boyle and Jones, 1985; McGee et al., 1985; Iloeje and Meme, 1992; Klein et al., 2009; Narusyte et al., 2017) and the data themselves.

The scree plot of the eigenvalues of the first iteration shows which factors are relevant, based on their eigenvalue. A commonly used rule of thumb is to exclude factors with eigenvalue below one (using the criterion set by Kaiser (1960)), as that implies those factors explain less variation than an individual item. Figure 3 shows that, while there are four factors with an eigenvalue above one, two stand out in this first iteration. That reconciles with our theoretical framework.

We also need to consider which items are worth retaining for the following round of factor analysis. To do so, we look at the rotated factor loadings and unique variances after each iteration. Rotation involves a linear transformation of the factor loadings and helps to pin down each item to a single factor. It is performed using oblique (quartimin) rotation, allowing the factors to correlate. We drop all items with loads below 0.4 and/or uniqueness above 0.8. Table 4 shows the outcome of the

[16] We code the variables to have the lowest value ("0") to represent children to which the description "Definitely Applies", the middle value ("1") to represent children to which the description "Somewhat Applies", and the highest value ("2") to represent children to which the description "Does not Apply". In this way, we can interpret a positive coefficient as a positive effect.

first and second iteration. 19 items load clearly on factors representing externalizing or internalizing skills. A second iteration of factor analysis does not lead to dropping any further items. The remaining items load on a single factor and allow us to distinguish clearly which ones characterize externalizing or internalizing skills. The two factors explain 80% of the total variance.

We use the Bartlett method (Hershberger, 2005), which exploits maximum likelihood estimation to calculate unbiased estimates of the true factor scores. We obtain a value for externalizing and internalizing skills for each child in the sample. We then standardize the measures by cohort.

Figure 4 illustrates the distribution by cohort of externalizing skills, while Figure 5 shows the distribution of internalizing skills (again, by cohort). As we would expect from a measure derived from an index to detect small behavioral issues, we see that both the distributions of non-cognitive skills are concentrated in a fairly narrow range, capturing the majority of children with a few minor behavioral difficulties. Both have a long left tail characterizing the comparatively few children with more severe behavioral disorders.

4.3 Non-Cognitive Skills within the Rank Effect Literature

Isolating externalizing and internalizing skills allows us to test the hypotheses of self-control and self-concept mechanisms for the rank effect. The distinction between the two is subtle, as both view rank as a tool providing students with information on their academic abilities. In other words, the information received in terms of relative cognitive skills changes their expectations.

The self-control theory implies that students will exert the effort necessary to stay in that area of the distribution. In doing so, the children rely on externalizing skills such as the ability to concentrate and control impulsive behaviors that could disrupt learning. For example, Pagani et al. (2021) show that conscientiousness, the personality trait connected with being responsible and diligent behavior improves because of rank, supporting this type of mechanism.

On the other hand, the self-concept theory is based on the idea that students with a higher rank do not necessarily put more effort. They develop a different image of themselves, becoming more confident in their abilities and having higher self-expectations. Indeed, Murphy and Weinhardt (2020) show that there is a positive rank effect on self-confidence, while Elsner and Isphording (2017); Elsner et al. (2021) find that student expectations for future grades and careers are affected by rank. All these characteristics belong to internalizing skills. There is evidence from the psychology literature of the positive relationship between internalizing issues and damaged self-esteem (Creemers et al., 2013).

While expectations are a key part of both mechanisms, how they are involved is significantly different. When it comes to externalizing skills, we think of expectations as a feature affecting the investment put into an activity. As for internalizing skills, expectations are a process of self-confirmation, where the students receive feedback on their skills and what their realistic achievement could be based on how they fare among their peers. That impacts their self-image and how they approach learning and being tested.

Contrasting these two dimensions is our major contribution. Externalizing and internalizing skills are composite measures. Studying the rank effect on uni-dimensional concepts such as self-esteem, expectations, or conscientiousness, has the clear advantage of pinning down a unique mechanism. However, a phenomenon such as the rank effect is likely to impact different types of non-cognitive skills, and we want to emphasize the importance of not looking at it from a limited perspective.

4.4 Adapted Specification for the Estimation of the Rank Effect on Non-Cognitive Skills

A concern in our setting is the absence of a baseline measure of non-cognitive skills when we estimate the rank effect. While quasi-randomization into groups solves the possible selection problem, using cognitive skills as the only control for non-cognitive skills may not adequately address any correlation or causal relationship between non-cognitive skills on school entry and the formation of rank by age 9, which would lead to bias.

We look at the psychology literature to find a more robust way of controlling for the baseline of non-cognitive skills. In particular, psychologists (Santrock and Feldman, 2020; Berk, 2023) distinguish three dimensions of child development: cognitive, non-cognitive, and physical. These domains are all closely related, and extremely difficult to disentangle. However, their high correlation is helpful, as we can exploit the two dimensions available to us (cognitive and physical) as a baseline for the one we lack. Moreover, Duckworth et al. (2019) show that there is a stronger relationship between physical development and non-cognitive skills than there is between the latter and cognitive skills.

With that in mind, we augment Equation 6 by adding two quadratic polynomials. $j(H_{isc})$ accounts for the height measured in the first medical exam[17]. $k(B_{isc})$ is the birth weight (measured in pounds). We estimate the following equation:

$$Y_{isc} = \alpha R_{isc} + \beta g(A_{isc}) + \gamma_1 \overline{A}_{sc-i} + \gamma_2 \sigma_{sc-i} + \theta_1 j(H_{isc}) + \theta_2 k(B_{isc}) + X_i \delta + \lambda_s + \lambda_c + \epsilon_{isc} \quad (9)$$

[17] The first medical exam had children measured and weighted at the beginning of primary school. Absentees would undergo the exam in the following intake. Therefore, the age in months at which children were examined could differ by even more than 12 months. To have a measure of height that is "standardized" by age, we use residualized height. We consider a range of children that underwent the first medical exam over a 12-month age range (dropping 1.5% of the sample). We then regress height over a quadratic polynomial of their age (both at the time of their first medical exam), and then take the residuals as our height variable.

5 Results

5.1 The Rank Effect on the 11-plus Test

We estimate the rank effect on the 11-plus Test using Equation 6. When the two youngest cohorts sat the test, they were asked to complete only the two verbal reasoning tasks. Therefore, the score given by the four components is available only for four grades out of five. Our main performance indicator will be the outcome of the Verbal Reasoning Test, available for all the children in the sample[18]. We also restrict the sample to include only the children who took the overall 11-plus test and re-estimate the rank effect.

Table 5 shows that a higher rank positively affects subsequent test scores. For both the dependent variables, we first estimate the rank effect without including controls for peer group and individual characteristics (columns 1 and 4), then add individual characteristics (columns 2 and 5), and finally estimate the rank effect controlling for peer group characteristics (columns 3 and 6). The estimated coefficients represent the difference in performance between the top and the bottom of the ranking, once all other effects have been accounted for. We characterize the effect using the average group size (40 children) as a reference, to facilitate interpretation. A 10% difference in the ranking is equivalent to a change of 4 positions. Because of that, moving up or down in the ranking by 4 positions will imply, on average, a rank effect that is 10% of the size of the estimated coefficient. Therefore, ranking 4 positions higher will cause an increase in the 11-plus test score equivalent to 6% of a unit of standard deviation (the standard deviation of the 11-plus test score was standardized to equal one). The effects align with those estimated by Murphy and Weinhardt (2020) and Elsner et al. (2021). Moreover, the coefficients remain stable when we include individual characteristics and features of the peer group characteristics (namely, mean and standard deviation).

We have assumed that the rank effect is linear and included a quadratic polynomial of individual cognitive skills to avoid potentially biased estimates due to non-linearities. We break down the ranking in deciles and plot the rank effect in Figure 6. The path we observe supports our initial assumption on the linearity of the effect and the idea that the rank effect is self-preserving. That is the fact that children's test scores increase due to rank so future rank is an outcome of the previous rank, making us indifferent to the timing of the test used to measure cognitive skills. The effect on the bottom decile tapers down compared to the overall trend, showcasing a larger confidence interval. That reconciles with Megalokonomou and Zhang (2022) findings, where students at the bottom of the ranking are less aware of their actual position, leading to noisier estimates of the rank effect.

5.2 The Rank Effect on Non-Cognitive Skills

Finding a positive and stable rank effect is ideal for studying its potential channels. We test the "self-control" and "self-concept" hypotheses, linking them to externalizing and internalizing skills, respectively. A positive rank effect on externalizing skills would imply that rank affects students' self-control, particularly behaviors that are disruptive to learning. This behavioral adaptation spurs from their attempt to maintain their position within the group. On the other hand, a positive rank effect on internalizing skills points to changes in those non-cognitive skills related to the individual

[18] Two of the four components of the 11-plus test were Verbal Reasoning Tests. We take the best result achieved by each child as the observed outcome.

ability to exploit their skills to perform a certain task (for example, self-efficacy, academic self-concept, self-confidence).

We estimate the rank effect on externalizing and internalizing skills. We include only children who were still in primary school at the time of the Rutter Questionnaire, so all those born between April 1953 and October 1955 (7 cohorts in total). Table 6 shows no impact of rank on externalizing skills, while the effect on internalizing skills is positive and statistically significant. Again, we show the coefficient estimated including only controls for cognitive skills, as well as school and cohort fixed effects (columns 1 and 5) first. We then add controls for individual characteristics (columns 2 and 6), and for peer cognitive skills (columns 3 and 7) as well. Finally, we augment the equation with physical development controls (columns 4 and 8), estimating the complete Equation 9.

Again, we characterize the effect using the average group size (40 children) as a reference. An improvement of 4 positions out of 40 impacts internalizing skills by roughly 6% of a unit of standard deviation (non-cognitive skills measures have been standardized to have unit standard deviation) in our preferred specification. The lower bound of the effect, estimated when we leave out peer distribution characteristics, is significant and does not drop in magnitude. Our result adds to the few papers showing that internalizing skills such as self-confidence (Murphy and Weinhardt, 2020) and academic self-concept (Elsner and Isphording, 2017; Elsner et al., 2021) are affected by rank. Conversely, we do not find evidence of a rank effect on externalizing skills such as conscientiousness and executive function, as in Pagani et al. (2021) and Carneiro et al. (2023).

We want to verify that the rank effect is linear also on non-cognitive skills. Figure 7 shows the breakdown of the rank effect by rank decile. On the left-hand side, we have externalizing skills. While positive, the difference between the top and bottom deciles is not statistically significant. In general, while the direction of the effect is "increasing" linearly, it is evident that the rank effect is not there. On the right-hand side, we show the same plot for internalizing skills. The trend we observe is indeed linear, although the rank effect on children in the top decile is estimated imprecisely.

5.3 Heterogeneous Rank Effects: Sex

One dimension of heterogeneity that can provide insights into our understanding of the phenomenon is sex at birth. We know that girls normally have better test scores, which can be attributed to better self-control (Duckworth and Seligman, 2005, 2006; Duckworth et al., 2015). We show these expected differences in baseline characteristics in Table 7. Girls perform better in the Age 9 and the Verbal Reasoning Task of the 11-plus test. As we would expect from standardized variables, the magnitude of the difference is almost identical. The difference in non-cognitive skills is even more interesting. Overall, girls have better externalizing and worse internalizing skills than boys (40% and 10% of a unit of standard deviation, respectively). Taking a step back to the psychological literature, boys exhibit worse self-control, while girls showcase a lower sense of self.

We do not have a prior expectation of whether the rank effect should impact children differently based on sex. However, considering that internalizing skills are those impacted by rank, we might expect a stronger rank effect on women. That is what we see in Table 8. The rank effect on girls is 25% larger than on boys. This difference is statistically significant. While there is no effect on girls' externalizing skills, we see a statistically significant effect (at the 5% level) on boys. The size of the effect implies an improvement of almost 4% units of standard deviation as a consequence of a jump in rank of 4 positions (considering an average group size of 40 children). This shows that non-cognitive skills related to self-control might be indeed affected by rank, as studies such as Pagani et al. (2021) and Carneiro et al. (2023) find. On the other hand, girls drive the rank effect on internalizing skills. The coefficient is roughly 50% larger than for boys, and this difference is statistically significant. Overall, rank affects girls only on the internalizing dimension, pointing at improvements in skills such as self-confidence and academic self-concept. Boys' ranking affects both their externalizing and internalizing dimensions.

5.4 Long-Term Outcomes

We exploit the follow-up survey run in 2001, which involves roughly 60% of the original participants, to look at the persistence of the rank effect on long-term outcomes. We focus on educational attainment and mental health.

We initially estimate the rank effect on the probability of participating in the 2001 follow-up survey. Then we look at educational attainments as the probability of achieving an O-level qualification, an A-level qualification, or a degree. O-levels, known as the "Ordinary grade" at the time (in Scotland), were a high-stakes test that students in secondary school would take at 16 years of age. Pupils who passed their O-level exam in a given subject were allowed to take the subject at the A-level grade, after one or more years of courses. The A-levels, named "Higher grade" (again, in Scotland), were high-stakes exams testing the students' proficiency in different subjects at a higher level, and their results granted access to university education.

The first column of Table 9 shows that rank does not impact the probability of responding to the follow-up survey. Because of this, we do not use non-response weights in estimating the rank effect on these long-term outcomes. We then look at the rank effect on the probability of achieving the three levels of education highlighted above. The following columns of Table 9 show a positive and statistically significant rank effect on achieving an O-level and an A-level qualification, but not on the probability of getting a degree. Ranking 4 positions higher in the average group implied a 4% higher probability of attaining the two qualifications. The similarity between the two coefficients (and lack of an effect on the probability of getting a degree) might depend on the fact that those who were getting the O-levels would then also get the A-levels, without necessarily enrolling in higher education after.

It is only natural to ask ourselves if improved non-cognitive skills can affect mental health in the long run. To test that, we use two different proxies for mental health, exploiting the questions asked in the 2001 follow-up survey. First, we look at the probability of enjoying daily activities more or about as usual, rather than less or much less than usual. Secondly, we look at the probability of feeling happy more or about the same as usual, rather than less or much less than usual.

However, we need to adjust our specification slightly. We want to be able to partial-out the effect of educational attainment from the rank effect coefficient on mental health. Therefore, we control for the education of the respondent (where the baseline are those that leave school before getting their O-levels) in the second and third columns of Table 10. The estimated coefficients are positive but imprecise. While Kiessling and Norris (2020) show that rank can impact students' mental health while they are in school, we provide evidence that this effect might potentially last almost 40 years later.

6 Robustness Checks

We present robustness exercises to address three different concerns. First, the possibility of a correlation between rank and heterogeneous effects of the peer group distribution, which could introduce bias into our estimates if not accounted for. Secondly, we want to verify that our main result is not driven by children with extremely poor internalizing skills (especially considering the length of the left tail of the distribution). Finally, despite having relatively small school-cohort groups (on average), we might be worried about the degree of interaction of children in the same group that are assigned to different classes.

6.1 Alternative Specification

As we discussed in section 3, the ideal experiment to study the rank effect would involve comparing groups with similar cognitive skills distributions. That could limit the chances of a correlation between rank and heterogeneous effects of the group distribution. However, the more this comparison is forced, the more variation in rank we would lose, to the point that we would not measure any rank effect. Because our results do not change substantially between baseline and fully controlled specifications, this suggests our choice of school plus cohort fixed effects already addresses passive sorting and potential heterogeneous effects of the group distributions very well. We therefore do not feel the need to introduce further restrictions in order to obtain unbiased estimates.

Still, we follow the blueprint of Denning et al. (2023) and re-estimate the rank effect using an alternative specification. We replace the quadratic polynomial of peer cognitive skills in our baseline equations with features of the cognitive skills distribution of the groups. To capture the rank effect on future test scores, we estimate the following variation of Equation 6:

$$Y_{isc} = \alpha R_{isc} + \beta g(A_{isc}) + \sum_{M=1}^{4} \mathbb{1}(M_{sc} = m)\gamma_{1m} + \sum_{V=1}^{4} \mathbb{1}(V_{sc} = v)\gamma_{2v} + X_i\delta + \lambda_s + \lambda_c + \epsilon_{isc} \quad (10)$$

Where $\sum_{M=1}^{4} \mathbb{1}(M_{sck} = m)$ and $\sum_{V=1}^{4} \mathbb{1}(V_{sck} = v)$ are indicator functions representing the quartiles of the mean and variance in the cognitive skills distributions of the groups. To estimate the rank effect on non-cognitive skills, we change Equation 9 similarly:

$$\begin{aligned} Y_{isc} = \alpha R_{isc} + \beta g(A_{isc}) + \sum_{M=1}^{4} \mathbb{1}(M_{sc} = m)\gamma_{1m} + \sum_{V=1}^{4} \mathbb{1}(V_{sc} = v)\gamma_{2v} + \theta_1 j(H_{isc}) + \\ + \theta_2 k(B_{isc}) + X_i\delta + \lambda_s + \lambda_c + \epsilon_{isc} \end{aligned} \quad (11)$$

We show the results in Table 11. The coefficients of the rank effect on future test scores are nearly identical to those estimated using Equation 6. Similarly, we do not see striking differences in the impact on non-cognitive skills. The rank effect on externalizing skills is again statistically insignificant. The one on internalizing skills decreases slightly in magnitude. Overall, it still implies that jumping 4 positions ahead in a group of 40 children improves internalizing skills by more than 4% of a unit of standard deviation.

6.2 Excluding the Children with Severe Internalizing Issues

As we can see from Figure 5, the distribution of internalizing skills is characterized by long left tails, which capture severe behavioral issues. Since our main finding revolves around how the rank effect impacts internalizing skills and how that explains why we measure a rank effect in the first place, we want to assess the role of these extreme values in driving the coefficient.

We progressively exclude a higher percentage of the children with severe internalizing issues (starting from 2% and up to 5%) and re-estimate the rank effect. Results are shown in Table 12. As expected, we observe a progressive drop in the magnitude of the coefficient, while the significance remains strong. We argue these changes are inconsistent with what we would see if extreme internalizing issues drove the results. This is also consistent with the linearity of the rank effect that we observed previously.

6.3 Classroom Rank Effect

The natural group within which we define and consider the effects of rank may be classes within a school. In our analysis, we have constructed rank within school by cohort groups. We have ignored students' assignment to different classrooms for two reasons. First, we are concerned that within schools and cohorts, better performing students may actively be sorted into different classrooms. Second, and related, if such streaming is applied, it may be carried out either before or after we observe classroom assignment[19].

However, we can assess within-classroom rank effects without being concerned about sorting or streaming. We restrict our sample to school-cohort groups with only one class per cohort. This restriction reduces our sample size by roughly 4 times, leaving us with around 120 school-cohort groups (belonging to 21 different schools), and an average group size of 26 students.

We show the rank effect on test scores and non-cognitive skills in Table 13. The coefficients are similar to those previously estimated, with a slight decrease in those related to the 11-plus test, and an increase in those for non-cognitive skills. We need to consider that, as the average group size is reduced, the effect expressed in units of the standard deviation of the outcome follows a hypothetical jump of between 2 to 3 positions, rather than 3 to 4. This similarity supports the results estimated using the entire sample. Moreover, our estimates might represent a lower bound of the classroom rank effect, at least in the absence of streaming.

[19]By examining the average cognitive skills across classes, we can see students were sorted according to their test scores. However, our knowledge of the institutional setting cannot draw further conclusions. For example, we know that streaming happened before grade 3, as the youngest cohort in our sample was already divided into classes with different average test scores when we observed them the first time (halfway through third grade, in December 1962). Also, we are unsure if we can take the classroom indicator at face value. We do not know if students were reshuffled depending on the subject, and the indicator only refers to one of these subjects. Moreover, we observe many children characterized by classroom indicators that involve only one, or sometimes a handful of students. Focusing on school-cohort groups rather than school-cohort-class groups does not harm our identification strategy, since we are not assuming randomization beyond school-cohort groups.

7 Conclusion

We study the rank effect in a context where a rule based on the month of birth creates idiosyncratic variation in the peer groups. We use a standardized test administered at 9 years of age to construct the ranking of the students within their school-cohort group. Students with identical baseline test scores rank differently in their peer group due to these random changes in group composition across birth cohorts and between schools. We then estimate the rank effect test score achieved in a high-stakes test at 11 years old, non-cognitive skills, educational attainment, and mental health 40 years later.

This study breaks new ground by delving deeper into the mechanisms driving the rank effect. We introduce a novel approach that leverages the distinction between externalizing and internalizing skills. While acknowledging the limitation of not examining individual skills, we avoid the challenge of disentangling highly interconnected components. It is well-established that internalizing skills encompass various aspects like self-confidence, academic self-concept, and ability expectations. However, focusing on the broader externalizing/internalizing divide offers a more comprehensive view of the rank effect and its underlying mechanisms, particularly how these two facets interact. We show that the effect on future performance is matched by a similar-sized one on internalizing skills. This goes in the direction of the "self-concept" hypothesis. That mechanism may be dominant, but it can also coexist with "self-control", particularly for boys' academic performance. Interestingly, the rank effect impacts both boys' externalizing and internalizing skills. Girls see only their internalizing skills improve as a consequence of their rank, but by a factor that is 50% larger. Overall, this highlights that both boys and girls improve their worse non-cognitive skill thanks to the rank effect. But also that their production function might require different inputs to produce the same output.

We also offer a two-pronged contribution to understanding the long-term influence of non-cognitive skills. Firstly, we demonstrate the persistence of these skills throughout life (Attanasio et al., 2020b,a). Our findings reveal that the rank effect, which compares an individual's standing within their peer group, still impacts self-reported mental health nearly four decades later. Notably, this effect operates primarily through increased educational attainment, which itself can be attributed to the rank effect. This aligns with research highlighting the long-term benefits of childhood interventions aimed at enhancing non-cognitive skills Heckman and Kautz (2012); Heckman et al. (2013).

References

Achenbach, T. M. and Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin*, 85(6):1275–1301.

Attanasio, O., Blundell, R., Conti, G., and Mason, G. (2020a). Inequality in socio-emotional skills: A cross-cohort comparison. *Journal of Public Economics*, 191:104171.

Attanasio, O., Ãureo de Paula, and Toppeta, A. (2020b). The Persistence of Socio-Emotional Skills: Life Cycle and Intergenerational Evidence. Working Papers 2020-066, Human Capital and Economic Opportunity Working Group.

Behar, L. and Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology*, 10(5):601–609.

Berk, L. E. (2023). *Development Through the Lifespan*. Pearson Education, 10th edition.

Boyle, M. H. and Jones, S. C. (1985). Selecting measures of emotional and behavioral disorders of childhood for use in general populations. *Journal of Child Psychology and Psychiatry*, 26(1):137–159.

Carneiro, P., Cruz-Aguayo, Y., Salvati, F., and Schady, N. (2023). The effect of classroom rank on learning throughout elementary school: Experimental evidence from ecuador. *Journal of Labor Economics*, 0(ja):null.

Clark, D. and Del Bono, E. (2016). The long-run effects of attending an elite school: Evidence from the united kingdom. *American Economic Journal: Applied Economics*, 8(1):150–76.

Creemers, D. H., Scholte, R. H., Engels, R. C., Prinstein, M. J., and Wiers, R. W. (2013). Damaged self-esteem is associated with internalizing problems. *Frontiers in Psychology*, 152(4).

Delaney, J. M. and Devereux, P. J. (2020). *Rank Effects in Education: What Do We Know So Far?*, pages 1–24. Springer International Publishing.

Delaney, J. M. and Devereux, P. J. (2021). High school rank in math and english and the gender gap in stem. *Labour Economics*, 69:101969.

Denning, J. T., Murphy, R., and Weinhardt, F. (2023). Class Rank and Long-Run Outcomes. *The Review of Economics and Statistics*, 105(6):1426–1441.

Duckworth, A. L., Quirk, A., Gallop, R., Hoyle, R. H., Kelly, D. R., and Matthews, M. D. (2019). Cognitive and noncognitive predictors of success. *Proceedings of the National Academy of Sciences*, 116(47):23499–23504.

Duckworth, A. L. and Seligman, M. E. (2005). Self-discipline outdoes iq in predicting academic performance of adolescents. *Psychological Science*, 16(12):939–944. PMID: 16313657.

Duckworth, A. L. and Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 1(98):198–208.

Duckworth, A. L., Shulman, E. P., Mastronarde, A. J., Patrick, S. D., Zhang, J., and Druckman, J. (2015). Will not want: Self-control rather than motivation explains the female advantage in report card grades. *Learn Individual Differences*, 39(1):13–23.

Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., Murphy, B. C., Losoya, S. H., and Guthrie, I. K. (2001). The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development*, pages 1112–1134.

Elsner, B. and Isphording, I. E. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, 35(3):787–828.

Elsner, B. and Isphording, I. E. (2018). Rank, sex, drugs, and crime. *Journal of Human Resources*, 53(2):356–381.

Elsner, B., Isphording, I. E., and Zölitz, U. (2021). Achievement Rank Affects Performance and Major Choices in College. *The Economic Journal*, 131(640):3182–3206.

Goulas, S., Griselda, S., and Megalokonomou, R. (2022). Comparative advantage and gender gap in stem. *Journal of Human Resources*, pages 0320–10781R2.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.

Heckman, J. J. and Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4):451–464. European Association of Labour Economists 23rd annual conference, Paphos, Cyprus, 22-24th September 2011.

Hershberger, S. L. (2005). *Factor Score Estimation*. John Wiley & Sons, Ltd.

Iloeje, S. O. and Meme, J. (1992). Rutter's Behaviour Scale (B2) for Children (Teacher's Scale): Validation and Standardization for Use on Nigerian Children. *Journal of Tropical Pediatrics*, 38(5):235–239.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1):141–151.

Kiessling, L. and Norris, J. (2020). The long-run effects of peers on mental health. Discussion paper series of the max planck institute for research on collective goods, Max Planck Institute for Research on Collective Goods.

Klein, J. M., Goña Salves, A., and Silva, C. F. (2009). The rutter children behaviour questionnaire for teachers: from psychometrics to norms, estimating caseness. *PsicoUSF*, 14:157 – 165.

Lawlor, D. A., Smith, G. D., and Ebrahim, S. (2006). Season of birth and childhood intelligence: findings from the aberdeen children of the 1950s cohort study. *British Journal of Educational Psychology*, 76(3):481–499.

Marsh, H. W. and Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47(1):213–231.

McGee, R., Williams, S., Bradshaw, J., Chapel, J. L., Robins, A., and Silva, P. A. (1985). The rutter scale for completion by teachers: Factor structure and relationships with cognitive abilities and family adversity for a sample of new zealand children. *Journal of Child Psychology and Psychiatry*, 26(5):727–739.

Megalokonomou, R. and Zhang, Y. (2022). How good am i? effects and mechanisms behind salient ranks. Technical report.

Murphy, R. and Weinhardt, F. (2020). Top of the Class: The Importance of Ordinal Rank. *The Review of Economic Studies*, 87(6):2777–2826.

Narusyte, J., Ropponen, A., Alexanderson, K., and Svedberg, P. (2017). Internalizing and externalizing problems in childhood and adolescence as predictors of work incapacity in young adulthood. *Social Psychiatry and Psychiatric Epidemiology*, 52(9):1159–1168.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443–460.

Pagani, L., Comi, S., and Origo, F. (2021). The Effect of School Rank on Personality Traits. *Journal of Human Resources*, 56(4):1187–1225.

Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry*, 8(1):1–11.

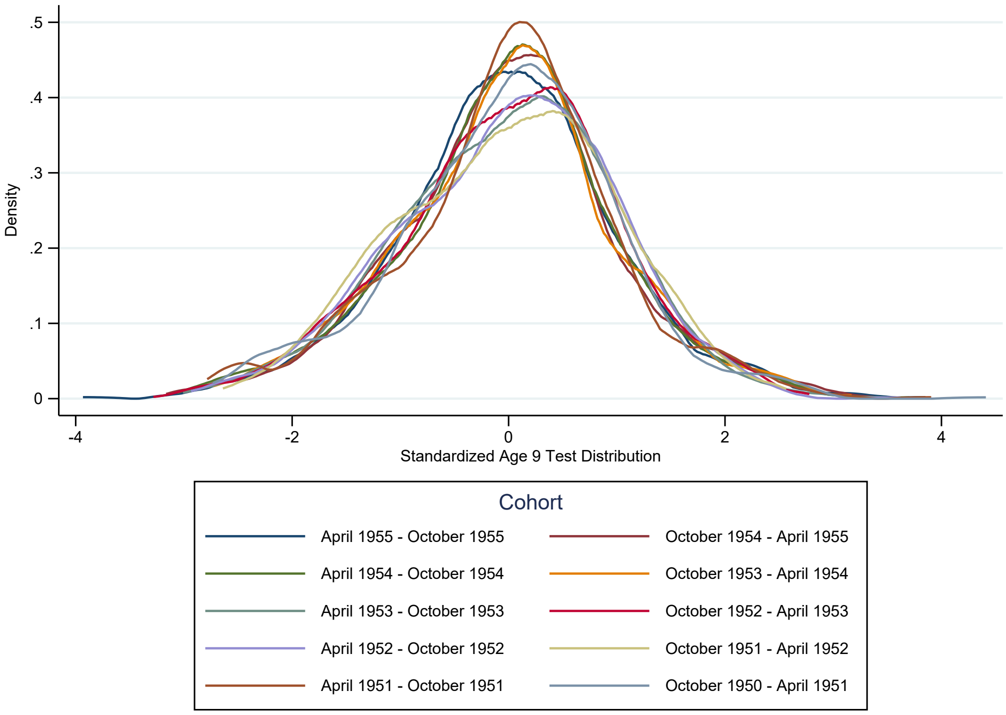
Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they, and how much do we know thus far? *Handbook of the Economics of Education*, 3:681–703.

Santrock, J. W. and Feldman, R. S. (2020). *Child Development*. McGraw-Hill Education, 19th edition.

Schonell, F. J. and Adams, R. H. (1940). *The essential intelligence test*. Edinburgh, London: Oliver & Boyd.

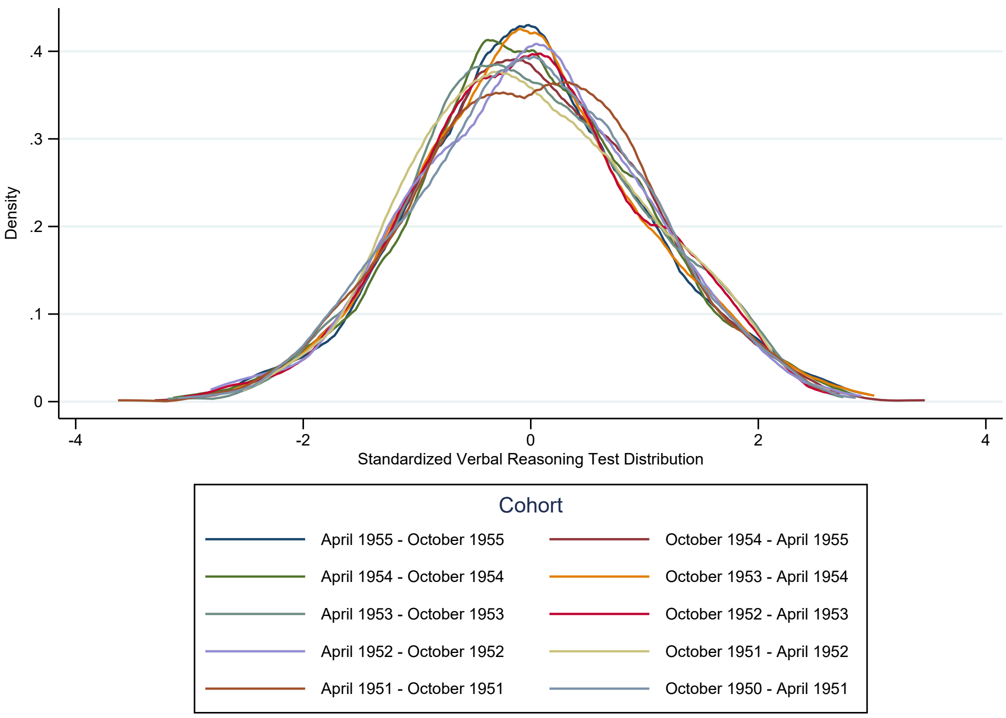
Sorrenti, G., Zölitz, U., Ribeaud, D., and Eisner, M. (2024). The Causal Impact of Socio-Emotional Skills Training on Educational Success. *The Review of Economic Studies*, page rdae018.

Figure 1: Distribution of Age 9 Test Results, by Cohort



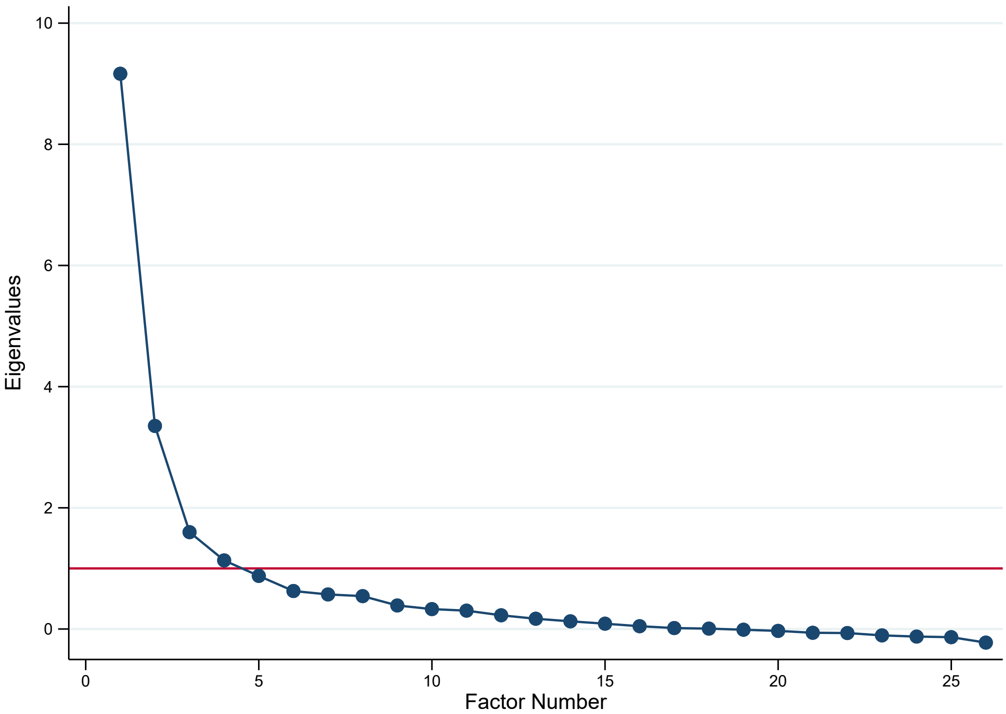
Notes: We show the distribution of the outcomes of the Age 9 Test, by cohort. The test scores are standardized at the cohort level. We include only children born between October 1950 and September 1955. The number of observations is 9,368.

Figure 2: Distribution of Age 9 Test Results, by Cohort



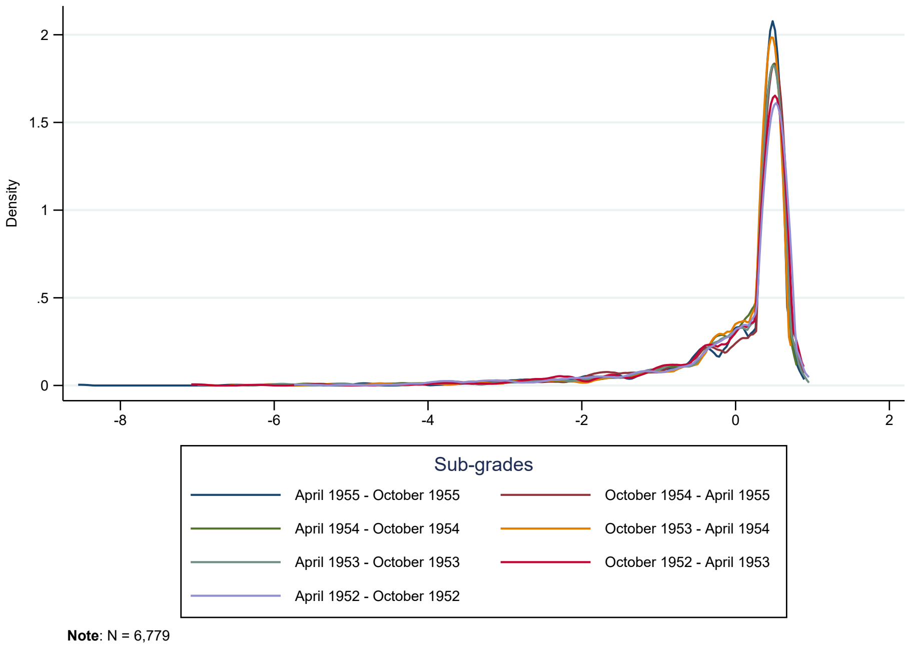
Notes: We show the distribution of the outcomes of the Verbal Reasoning Test, by cohort. The test scores are standardized at the cohort level. We include only children born between October 1950 and September 1955. The number of observations is 9,698.

Figure 3: Scree Plot of the Eigenvalues: Iteration 1



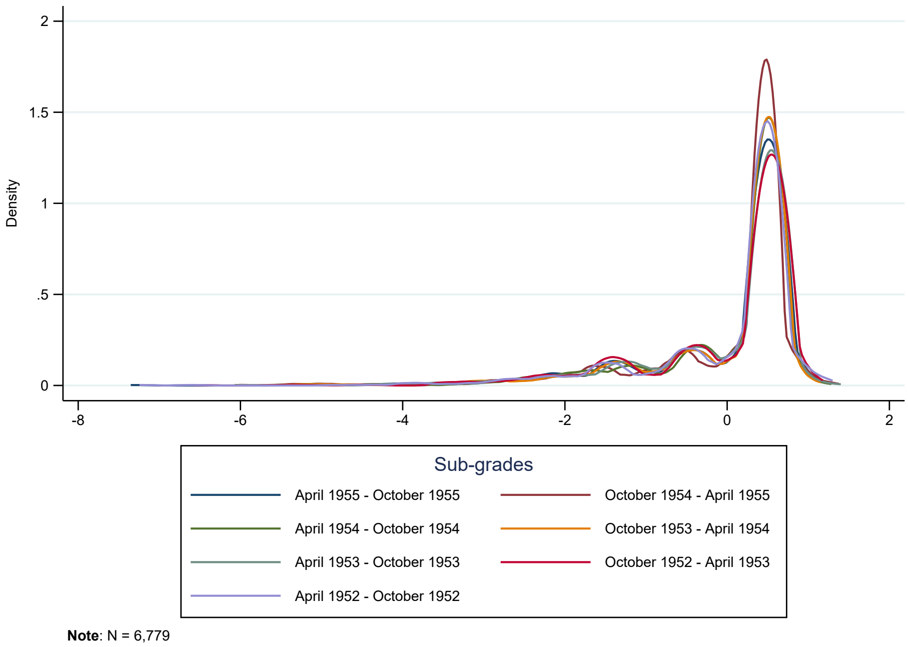
Notes: The graph plots the eigenvalue of each factor estimated through the first iteration of factor analysis. The number of observations is 6,779.

Figure 4: Externalizing Skills Distribution, by Cohort



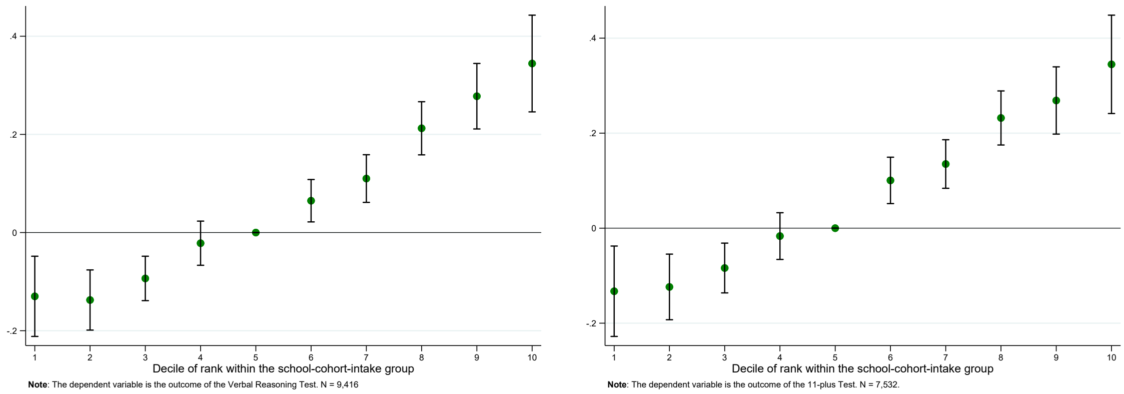
Notes: The graph shows the distribution of externalizing skills, by cohort. The variables are standardized at the cohort level. The number of observations is 6,779.

Figure 5: Internalizing Skills Distribution, by Cohort



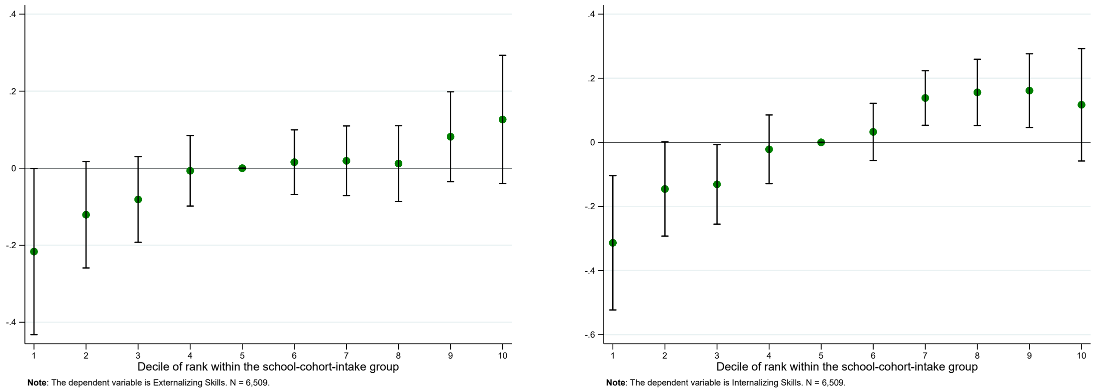
Notes: The graph shows the distribution of internalizing skills, by cohort. The variables are standardized at the cohort level. The number of observations is 6,779.

Figure 6: Rank effect on the (standardized) outcome of the Verbal Reasoning and 11-plus tests, by rank decile



Notes: The graph shows the rank effect on the Verbal Reasoning and the 11-plus test scores by rank decile.

Figure 7: Rank effect on (standardized) externalizing and internalizing skills, by rank decile



Notes: The graph shows the rank effect on externalizing and internalizing skills by rank decile.

Table 1: Grade Progression Mechanism and Survey Timings

Intake		Sub-grade										
	Survey Timings	Oct '50 - Mar '51	Apr '51 - Sep '51	Oct '51 - Mar '52	Apr '52 - Sep '52	Oct '52 - Mar '53	Apr '53 - Sep '53	Oct '53 - Mar '54	Apr '54 - Sep '54	Oct '54 - Mar '55	Apr '55 - Sep '55	Survey Timings
Jan '56		1st Grade	-	-	-	-	-	-	-	-	-	
Aug '56		1st Grade	1st Grade	-	-	-	-	-	-	-	-	
Jan '57		2nd Grade	1st Grade	1st Grade	-	-	-	-	-	-	-	
Aug '57	Age 7 Test ↘	2nd Grade	2nd Grade	1st Grade	1st Grade	-	-	-	-	-	-	
Jan '58		3rd Grade	2nd Grade	2nd Grade	1st Grade	1st Grade	-	-	-	-	-	
Aug '58		3rd Grade	3rd Grade	2nd Grade	2nd Grade	1st Grade	1st Grade	-	-	-	-	
Jan '59		4th Grade	3rd Grade	3rd Grade	2nd Grade	2nd Grade	1st Grade	1st Grade	-	-	-	
Aug '59	Age 9 Test ↘	4th Grade	4th Grade	3rd Grade	3rd Grade	2nd Grade	2nd Grade	1st Grade	1st Grade	-	-	
Jan '60		5th Grade	4th Grade	4th Grade	3rd Grade	3rd Grade	2nd Grade	2nd Grade	1st Grade	1st Grade	-	
Aug '60		5th Grade	5th Grade	4th Grade	4th Grade	3rd Grade	3rd Grade	2nd Grade	2nd Grade	1st Grade	1st Grade	
Jan '61		6th Grade	5th Grade	5th Grade	4th Grade	4th Grade	3rd Grade	3rd Grade	2nd Grade	2nd Grade	1st Grade	
Aug '61		6th Grade	6th Grade	5th Grade	5th Grade	4th Grade	4th Grade	3rd Grade	3rd Grade	2nd Grade	2nd Grade	
Jan '62		7th Grade	6th Grade	6th Grade	5th Grade	5th Grade	4th Grade	4th Grade	3rd Grade	3rd Grade	2nd Grade	↙ Age 7 Test
Aug '62	Reading Test ↘	7th Grade	7th Grade	6th Grade	6th Grade	5th Grade	5th Grade	4th Grade	4th Grade	3rd Grade	3rd Grade	↙ Reading Test
Jan '63		Junior Secondary School	7th Grade	7th Grade	6th Grade	6th Grade	5th Grade	5th Grade	4th Grade	4th Grade	3rd Grade	
Aug '63		November 1962: 11 + Test		7th Grade	7th Grade	6th Grade	6th Grade	5th Grade	5th Grade	4th Grade	4th Grade	
Jan '64	Rutter (scale b) ↘	-	-	Junior Secondary School	7th Grade	7th Grade	6th Grade	6th Grade	5th Grade	5th Grade	4th Grade	↙ Age 9 Test
Aug '64		-	-	November 1963: 11 + Test		7th Grade	7th Grade	6th Grade	6th Grade	5th Grade	5th Grade	↖ Rutter (scale b)
Jan '65		-	-	-	-	Junior Secondary School	7th Grade	7th Grade	6th Grade	6th Grade	5th Grade	
Aug '65		-	-	-	-	November 1964: 11 + Test		7th Grade	7th Grade	6th Grade	6th Grade	
Jan '66		-	-	-	-	-	-	Junior Secondary School	7th Grade	7th Grade	6th Grade	
Aug '66		-	-	-	-	-	-	November 1965: 11 + Test		7th Grade	7th Grade	
Jan '67		-	-	-	-	-	-	-	-	Junior Secondary School	7th Grade	
Aug '67		-	-	-	-	-	-	-	-	November 1966: 11 + Test		

Notes: The table summarizes the (adjusted) grade progression mechanism through which students advance their grades in primary school. We show the path for all the cohorts in our sample, separated into sub-grades. Children born between the beginning of October and the end of March will enter during the January admission window, while those born between the beginning of April and the end of September will start during the August admission window. We highlight those that are the key moments relative to our study. In green, we locate the time the Age 7 Test took place, for each sub-grade cohort. In blue, we do the same for the Age 9 Test. Both these tests were administered in November for the October-January cohorts, and in May for the April-September cohorts. The goal was to have kids taking these tests at the same relative age. In magenta, we highlight December 1962, when the Reading Test occurred. Finally, in red, we locate March 1964, when teachers filled out the Rutter Questionnaire (scale b) from which we extracted our social and emotional skills measures. Students who started primary school in January will enter junior secondary school for one term, before taking the 11+ Test in November and starting secondary school at the same time as the children in the following sub-grade.

Table 2: Descriptive Statistics of the Sample

	Mean	SD	Median	Min	Max	N
School and Cohort Characteristics						
School Size	362	157	307	168	750	28
Group Size	36	19	30	12	123	275
Individual Characteristics						
Females (%)	48	50	-	-	-	9,959
Socioeconomic Status (%)	91	28	-	-	-	9,959
Height at School Entry (cm)	107	5.13	107	73	137	9,310
Birth Weight (kg)	4.989	1.025	4,853	0.454	7.711	9,959
Number of Siblings	1.54	0.72	1	1	5	9,959
Cognitive and Non-Cognitive Skills (standardized)						
Age 9 Test	0	1	0.045	4.410	-3.933	9,698
Verbal Reasoning Test	0	1	-0.033	3.463	-3.627	9,600
11-plus Test	0	1	0.004	3.007	-3.544	7,739
Externalizing Skills	0	1	0.474	0.957	-8.549	6,779
Internalizing Skills	0	1	0.479	1.401	-7.352	6,779
Future Outcomes						
O-level (%)	60	49	-	-	-	5,862
A-level (%)	33	47	-	-	-	5,862
Degree (%)	15	36	-	-	-	5,862
Enjoys Daily Life (%)	89	31	-	-	-	5,862
Happy (%)	88	33	-	-	-	5,862

Notes: We report the characteristics of schools and the school-cohort groups, the individual characteristics of the children such as share of female, share of high socioeconomic status children, height at the first medical exam (in school), birth weight, and number of siblings. We also include measures of individual cognitive skills (calculated using the outcome of the Age 9 Test), the 11-plus Test (separating between Verbal Reasoning Task and overall test score), and measures of Externalizing and Internalizing skills (all standardized at the cohort level). Finally, we show the percentage of the children who responded to the 2001 follow-up Survey that achieved an O-level, A-level, and Degree-level of education, and whether they declared enjoying daily activities or being happy. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Balancing Exercise: Individual Characteristics, Rank, and Peer Cognitive Skills

Variables	**Woman**	**High SES**	**Height**	**Weight**	**Birth Weight**	**Siblings**
Equation 7: Conditional relation between individual characteristics and rank						
Percentile Rank	0.018	0.017	0.167	0.165*	0.052	-0.089
	(0.045)	(0.027)	(0.105)	(0.096)	(0.097)	(0.063)
Equation 8: Conditional relation between individual characteristics and peer quality						
Mean of Peer Cognitive Skills	-0.0001	-0.003	-0.015	-0.024	0.005	0.008
	(0.007)	(0.004)	(0.023)	(0.018)	(0.017)	(0.011)
Standard Deviation of Peer Cognitive Skills	-0.005	0.005	-0.019	-0.011	0.001	0.0001
	(0.005)	(0.003)	(0.016)	(0.011)	(0.011)	(0.008)
25th Percentile of Peer Cognitive Skills	0.004	-0.004	-0.011	-0.014	-0.001	-0.006
	(0.007)	(0.004)	(0.022)	(0.016)	(0.015)	(0.010)
50th Percentile of Peer Cognitive Skills	-0.004	-0.004	0.001	-0.001	0.020	0.010
	(0.006)	(0.003)	(0.020)	(0.015)	(0.015)	(0.010)
75th Percentile of Peer Cognitive Skills	0.001	-0.001	-0.010	-0.017	0.006	0.010
	(0.006)	(0.003)	(0.016)	(0.013)	(0.015)	(0.010)
Observations	9,698	9,698	9,465	9,458	9,698	9,698

Notes: We estimate the relationship between ranking (Equation 7)/peer cognitive skills (Equation 8) at the school-cohort group level on different characteristics of the students: their probability of being females, their probability of coming from an advantaged socioeconomic background, their height and weight at the time of their first medical exam, and their number of siblings. Both rank and peer cognitive skills are based on the results of the Age 9 Test. We restrict our sample to children born between October 1950 and September 1955. Standard errors are clustered at the school-cohort-group level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Rotated Factor Loadings from the Exploratory Factor Analysis based on the 26 items of the Rutter Questionnaire for Teachers

	Iteration 1		Iteration 2	
Variable	Externalizing	Internalizing	Externalizing	Internalizing
Restless	**0.78**	0.00	**0.79**	-0.02
Truant	**0.68**	0.18	**0.67**	0.16
Fidgety	**0.77**	0.01	**0.76**	-0.01
Destroys Belongings	**0.89**	-0.07	**0.89**	-0.07
Fights Others	**0.87**	-0.04	**0.88**	-0.004
Disliked	**0.67**	0.33	**0.68**	0.33
Anxious	-0.16	**0.85**	-0.15	**0.86**
Solitary	0.11	**0.62**	0.12	**0.60**
Irritable	**0.75**	0.04	**0.76**	0.06
Often Unhappy and Miserable	0.22	**0.75**	0.24	**0.76**
Tics	0.39	0.32	-	-
Sucks Finger	0.26	0.25	-	-
Nail Biting	0.24	0.13	-	-
Trivial Absences	0.38	0.34	-	-
Disobedient	**0.87**	-0.12	**0.87**	-0.11
Poor Concentration	**0.57**	0.24	**0.56**	0.20
Afraid	-0.14	**0.85**	-0.12	**0.84**
Fussy over particular child	-0.18	**0.55**	-0.16	**0.58**
Often Lies	**0.86**	0.004	**0.86**	0.01
Stealing	**0.71**	-0.02	**0.70**	0.003
Wet/Soiled Themselves	0.26	0.29	-	-
Often Aching	0.16	**0.53**	0.17	**0.49**
Tearful	0.20	**0.63**	0.21	**0.65**
Stutters	0.20	0.29	-	-
Speech Difficulties	0.19	0.21	-	-
Bullies Others	**0.85**	-0.11	**0.85**	-0.09

Notes: We iterate exploratory factor analysis to decide which items to retain out of the 26 in the Rutter Questionnaire for Teachers. We report the factor loadings and the communities for the oblique rotated total variance matrix. We restrict our sample to children born between April 1951 and September 1955, since we want to include only children who were in primary school when the Rutter Questionnaire was completed (March 1964). In total, we have 6,779 children.

Table 5: Rank Effect on the 11-plus Test

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Verbal Reasoning Test			11-plus Test		
Percentile Rank	0.625***	0.631***	0.625***	0.607***	0.614***	0.644***
	(0.055)	(0.055)	(0.062)	(0.064)	(0.063)	(0.071)
Cognitive Skills	0.685***	0.679***	0.681***	0.698***	0.693***	0.684***
	(0.018)	(0.018)	(0.019)	(0.020)	(0.020)	(0.021)
Cognitive Skills Squared	0.002	0.001	0.002	-0.007	-0.008	-0.008
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Mean of Group Cognitive Skills			-0.002			0.010
			(0.013)			(0.016)
SD of Group Cognitive Skills			-0.005			0.005
			(0.008)			(0.010)
Observations	9,416	9,416	9,416	7,532	7,532	7,532
School Fixed Effects	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X
Sex	-	X	X	-	X	X
Socioeconomic Status	-	X	X	-	X	X
Month of Birth	-	X	X	-	X	X
Mean of Group Cognitive Skills	-	-	X	-	-	X
SD of Group Cognitive Skills	-	-	X	-	-	X

Notes: We estimate the relationship between the standardized Verbal Reasoning Test and the standardized 11-plus Test and percentile rank (based on the Age 9 Test score). We restrict our sample to children born between October 1950 and September 1955 when estimating the rank effect on the Verbal Reasoning Test, to children born between October 1951 and September 1955 when estimating the rank effect on the 11-plus Test. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Rank Effect on Non-Cognitive Skills

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Externalizing Skills				Internalizing Skills		
Percentile Rank	0.096	0.102	0.167	0.172	0.399***	0.402***	0.600***	0.583***
	(0.126)	(0.125)	(0.140)	(0.140)	(0.150)	(0.150)	(0.155)	(0.153)
Cognitive Skills	0.217***	0.207***	0.188***	0.185***	0.058***	0.059***	-0.0002	0.006
	(0.039)	(0.039)	(0.041)	(0.041)	(0.047)	(0.046)	(0.045)	(0.045)
Cognitive Skills Squared	-0.057***	-0.050***	-0.051***	-0.050***	-0.014	-0.016*	-0.019**	-0.020**
	(0.008)	(0.008)	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)	(0.010)
Mean of Group Cognitive Skills			0.020	0.018			0.060	0.056
			(0.028)	(0.028)			(0.041)	(0.041)
SD of Group Cognitive Skills			0.014	0.016			0.042*	0.041
			(0.017)	(0.017)			(0.025)	(0.025)
Height				0.005				-0.029**
				(0.013)				(0.014)
Height Squared				0.002				-0.001
				(0.006)				(0.007)
Birth Weight				-0.004				0.045***
				(0.012)				(0.013)
Birth Weight Squared				0.002				0.012
				(0.007)				(0.009)
Observations	6,615	6,615	6,615	6,509	6,615	6,615	6,615	6,509
School Fixed Effects	X	X	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X	X	X
Sex	-	X	X	X	-	X	X	X
Socioeconomic Status	-	X	X	X	-	X	X	X
Month of Birth	-	X	X	X	-	X	X	X
Mean of Group Cognitive Skills	-	-	X	X	-	-	X	X
SD of Group Cognitive Skills	-	-	X	X	-	-	X	X
Physical Development (Height)	-	-	-	X	-	-	-	X
Physical Development Squared (Height)	-	-	-	X	-	-	-	X
Physical Development (Birth Weight)	-	-	-	X	-	-	-	X
Physical Development Squared (Birth Weight)	-	-	-	X	-	-	-	X

Notes: We estimate the relationship between our standardized indicators of externalizing and internalizing skills and percentile rank (based on the Age 9 Test score). We restrict our sample to children born between April 1952 and September 1955. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Baseline Characteristics of Cognitive and Non-Cognitive Skills: Sex Heterogeneity

Variables	Male	Female	Combined	T-Test
Externalizing Skills	-0.175	0.186	0.000	-0.361***
	(0.020)	(0.013)	(0.012)	(0.024)
Observations	3,494	3,285	6,779	
Internalizing Skills	0.050	-0.053	0.000	0.103***
	(0.016)	(0.018)	(0.012)	(0.024)
Observations	3,494	3,285	6,779	
Age 9 Test	-0.044	0.048	0.000	-0.092***
	(0.015)	(0.014)	(0.010)	(0.024)
Observations	5,042	4,656	9,698	
Verbal Reasoning Test	-0.042	0.046	0.000	-0.088***
	(0.014)	(0.014)	(0.010)	(0.024)
Observations	4,990	4,610	9,600	

Notes: We report the mean values of externalizing and internalizing skills, as well as the mean values of the Age 9 Test score and the Verbal Reasoning Test, separately for boys and girls. All the variables are standardized to have a mean of 0 and a standard deviation of 1. We run a two-sample t-test of the difference between the means and report the results. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Rank Effect by Sex at Birth

	(1)	(2)	(3)	(4)
	Test Type		Non-Cognitive Skill	
Variables	Verbal Reasoning	11-plus	Externalizing	Internalizing
Percentile Rank # Males	0.544***	0.560***	0.370**	0.437***
	(0.065)	(0.073)	(0.159)	(0.162)
Percentile Rank # Females	0.681***	0.705***	0.044	0.679***
	(0.063)	(0.073)	(0.135)	(0.157)
P-Value for T-Test of Difference	0.137***	0.145***	-0.326***	0.241***
	(0.033)	(0.037)	(0.080)	(0.089)
Observations	9,416	7,532	6,509	6,509
School Fixed Effects	X	X	X	X
Cohort Fixed Effects	X	X	X	X
Cognitive Skills	X	X	X	X
Cognitive Skills Squared	X	X	X	X
Sex	X	X	X	X
Socioeconomic Status	X	X	X	X
Month of Birth	X	X	X	X
Mean of Group Cognitive Skills	X	X	X	X
SD of Group Cognitive Skills	X	X	X	X
Physical Development (Height)	-	-	X	X
Physical Development Squared (Height)	-	-	X	X
Physical Development (Birth Weight)	-	-	X	X
Physical Development Squared (Birth Weight)	-	-	X	X

Notes: We estimate the relationship between the standardized Verbal Reasoning Test, the standardized 11-plus Test, externalizing and internalizing skills, and percentile rank (based on the Age 9 Test score). We interact with percentile rank with a dummy that takes value one when the child is a female. We report the marginal effect of rank on female and male students. We restrict our sample to children born between October 1950 and September 1955. Furthermore, we restrict our sample to include only responders to the 2001 follow-up survey. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9: Rank Effect on Responding to the Survey and Educational Attainment

	(1)	(2)	(3)	(4)
Outcome Variable	Respondent in 2001	O-level	A-level	Degree
Percentile Rank	0.038	0.387***	0.423***	0.018
	(0.058)	(0.068)	(0.060)	(0.051)
Observations	9,698	5,729	5,729	5,729
School Fixed Effects	X	X	X	X
Cohort Fixed Effects	X	X	X	X
Cognitive Skills	X	X	X	X
Cognitive Skills Squared	X	X	X	X
Sex	X	X	X	X
Socioeconomic Status	X	X	X	X
Month of Birth	X	X	X	X
Mean of Group Cognitive Skills	X	X	X	X
SD of Group Cognitive Skills	X	X	X	X

Notes: We estimate the relationship between the probability of responding to the 2001 follow-up survey, achieving an O-level, A-level, or degree qualification, and percentile rank (based on the Age 9 Test score). We restrict our sample to children born between October 1950 and September 1955. When we estimate the rank effect on educational attainments we restrict our sample to include only responders to the 2001 follow-up survey. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10: Rank Effect on Mental Health

	(1)	(2)	(3)	(4)
Outcome Variable	Probability of Enjoying Daily Activities		Probability of Being Happy	
Percentile Rank	0.077*	0.071	0.107**	0.092*
	(0.043)	(0.043)	(0.051)	(0.052)
O-Level	-	0.034***	-	0.049***
	-	(0.010)	-	(0.010)
A-level	-	-0.014	-	-0.009
	-	(0.012)	-	(0.011)
Degree	-	0.005	-	0.008
	-	(0.013)	-	(0.013)
Observations	5,729	5,729	5,729	5,729
School Fixed Effects	X	X	X	X
Cohort Fixed Effects	X	X	X	X
Cognitive Skills	X	X	X	X
Cognitive Skills Squared	X	X	X	X
Sex	X	X	X	X
Socioeconomic Status	X	X	X	X
Month of Birth	X	X	X	X
Mean of Group Cognitive Skills	X	X	X	X
SD of Group Cognitive Skills	X	X	X	X

Notes: We estimate the relationship between the probability of "enjoying daily activities" and "being happy", and percentile rank (based on the Age 9 Test score). We restrict our sample to children born between October 1950 and September 1955. Furthermore, we restrict our sample to include only responders to the 2001 follow-up survey. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11: Alternative Specification: Rank Effect on Performance and Non-Cognitive Skills

	(1)	(2)	(3)	(4)
	Test Type		Non-Cognitive Skill	
Variables	Verbal Reasoning	11-plus	Externalizing	Internalizing
Percentile Rank	0.642***	0.651***	0.155	0.427***
	(0.061)	(0.071)	(0.139)	(0.154)
Observations	9,416	7,532	6,509	6,509
School Fixed Effects	X	X	X	X
Cohort Fixed Effects	X	X	X	X
Cognitive Skills	X	X	X	X
Cognitive Skills Squared	X	X	X	X
Sex	X	X	X	X
Socioeconomic Status	X	X	X	X
Month of Birth	X	X	X	X
Quartile of Mean of Group Cognitive Skills	X	X	X	X
Quartile of Variance of Group Cognitive Skills	X	X	X	X
Physical Development (Height)	-	-	X	X
Physical Development Squared (Height)	-	-	X	X
Physical Development (Birth Weight)	-	-	X	X
Physical Development Squared (Birth Weight)	-	-	X	X

Notes: We estimate the relationship between the standardized Verbal Reasoning Test, the standardized 11-plus Test, externalizing and internalizing skills, and percentile rank (based on the Age 9 Test score). We restrict our sample to children born between October 1950 and September 1955 when estimating the rank effect on the Verbal Reasoning Test, to children born between October 1951 and September 1955 when estimating the rank effect on the 11-plus Test, and to children born between April 1952 and September 1955 when estimating the rank effect on non-cognitive skills. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 12: Rank Effect on Internalizing Skills: Excluding Extreme Values

	(1)	(2)	(3)	(4)
Variables		Internalizing Skills		
Exclude from Sample Lowest:	2%	3%	4%	5%
Percentile Rank	0.463***	0.455***	0.395***	0.367***
	(0.126)	(0.120)	(0.110)	(0.100)
Observations	6,375	6,309	6,242	6,171
School Fixed Effects	X	X	X	X
Cohort Fixed Effects	X	X	X	X
Cognitive Skills	X	X	X	X
Cognitive Skills Squared	X	X	X	X
Sex	X	X	X	X
Socioeconomic Status	X	X	X	X
Month of Birth	X	X	X	X
Mean of Group Cognitive Skills	X	X	X	X
SD of Group Cognitive Skills	X	X	X	X
Physical Development (Height)	X	X	X	X
Physical Development Squared (Height)	X	X	X	X
Physical Development (Birth Weight)	X	X	X	X
Physical Development Squared (Birth Weight)	X	X	X	X

We estimate the relationship between standardized internalizing skills and percentile rank (based on the Age 9 Test score). We restrict our sample to children born between April 1952 and September 1955. Furthermore, we progressively restrict our sample to exclude the bottom 2%, 3%, 4%, and 5% of the children according to their internalizing skills. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 13: Classroom Rank Effect

	(1)	(2)	(3)	(4)
	Test Type		Non-Cognitive Skill	
Variables	Verbal Reasoning	11-plus	Externalizing	Internalizing
Percentile Rank	0.548***	0.521***	0.381	0.680**
	(0.108)	(0.111)	(0.281)	(0.293)
Observations	2,183	1,857	1,598	1,598
School Fixed Effects	X	X	X	X
Cohort Fixed Effects	X	X	X	X
Cognitive Skills	X	X	X	X
Cognitive Skills Squared	X	X	X	X
Sex	X	X	X	X
Socioeconomic Status	X	X	X	X
Month of Birth	X	X	X	X
Mean of Group Cognitive Skills	X	X	X	X
SD of Group Cognitive Skills	X	X	X	X
Physical Development (Height)	-	-	X	X
Physical Development Squared (Height)	-	-	X	X
Physical Development (Birth Weight)	-	-	X	X
Physical Development Squared (Birth Weight)	-	-	X	X

Notes: We estimate the relationship between the standardized Verbal Reasoning Test, the standardized 11-plus Test, externalizing and internalizing skills, and percentile rank (based on the Age 9 Test score). We restrict our sample to children born between October 1950 and September 1955 when estimating the rank effect on the Verbal Reasoning Test, to children born between October 1951 and September 1955 when estimating the rank effect on the 11-plus Test, and to children born between April 1952 and September 1955 when estimating the rank effect on non-cognitive skills. Furthermore, we restrict our sample to those school-cohort groups that were not broken down into multiple classrooms. Standard errors are clustered at the school-cohort level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.