CESifo WORKING PAPERS

10248
2023

January 2023

Adolescents' Mental Health and Human Capital: The Role of Socioeconomic Rank

Michaela Pfaffenholz

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10248*

Adolescents' Mental Health and Human Capital: The Role of Socioeconomic Rank

Abstract

I provide evidence on the causal effects of a student's relative socioeconomic status during high school on their mental health and human capital development. Leveraging data from representative US high schools, I utilize between-cohort differences in the distributions of socioeconomic status within schools in a linear fixed effects model to identify a causal rank effect. I find that a higher rank during high school improves a student's depression scores, cognitive ability, self-esteem and popularity. The rank effects are persistent with long-lasting consequences for adult depression and college attainment. Additional analyses emphasize the role of inequality in exacerbating these rank effects.

JEL-Codes: I140, I230, I240.

Keywords: rank, mental health, higher education.

Michaela Paffenholz
ifo Institute – Leibniz Institute for Economic Research
at the University of Munich / Germany
michael.paffenholz@econ.lmu.de

January 18, 2023
I thank Katja Kaufmann and Sebastian Siegloch, as well as participants of the CRC TR 224 Summer School on the Econometrics of Peer Effects and Social Interactions, the Summer School on the Development of Cognitive and Non-Cognitive Skills in Childhood and Adolescence, the Young Researchers Workshops of the CRC TR 224, and seminars at the University of Mannheim, the ZEW, and the University of Munich for helpful comments and feedback. This research uses data from Add Health, a program project funded by grant P01 HD31921 (Harris) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received for this analysis.

1 Introduction

The prevalence of mental health problems and their importance for individuals' lifetime trajectories and the economy as a whole have been increasingly recognized. The estimated total cost of mental health disorders on society was around 3.5% of GDP in 2010 (OECD, 2015). In this context, the mental health of teenagers is of particular interest, as many mental health disorders arise during adolescence, leading to concerns regarding adverse impacts on teenage development. These concerns typically center around potential long-term consequences, emphasizing the importance of an unimpeded development for outcomes such as educational attainment, health, and well-being. This view is supported by a large body of evidence that documents substantial economic and social returns to interventions in adolescence and neurobiological changes in brain regions involved in cognitive and social processes during the second decade of life (Dahl et al., 2018).

A commonly held perception is that teenagers are particularly susceptible to peer influence as they experience a reorientation towards peers and away from parents (Dahl et al., 2018). The notion that social context is an important factor in the human development is widely accepted in the economics of education literature, where peer characteristics are considered important determinants in the production of human capital. Complementary to the traditional peer effects view, which typically emphasizes absolute measures of peer quality, this paper follows the idea that an individual's relative position within their peer group may shape outcomes. The idea that relative characteristics matter for individuals' well-being and development has a long history in sociology and social psychology.[1] However, quantifying the causal effect of such relative attributes is challenging, primarily because social networks are endogenously formed.

This paper provides causal evidence on the effect of the relative socioeconomic status on adolescents' mental health, cognitive ability and educational attainment in the short- and long-run.[2] Motivated by the fact that adolescents spend a significant amount of time in school, I study the role of relative status within high school cohorts, which form a natural reference group for adolescents. My baseline measure of students' socioeconomic status (SES) is the highest level of schooling completed by their head of household, which I use to assign each student the percentile rank in their cohort SES distribution.[3] Studying the role of relative status in the framework of cohort networks allows me to address selection concerns by employing a fixed-effects approach recently

[1] Social Comparison Theory, for example, posits that individuals have the innate drive to evaluate themselves and, in the absence of objective standards, do so in terms of comparisons to others (Festinger, 1954). Social comparison phenomena have been investigated in various settings with the aim to understand the processes by which individuals come to understand themselves through relative comparisons (Suls and Miller, 1977).

[2] According to the theory of relative deprivation, feeling socially and economically deprived relative to a reference group can shape individuals' emotions, cognitions, and behaviors (see e.g. Crosby, 1976; Smith et al., 2012; Stouffer et al., 1949).

[3] My results are robust to variations in the SES definition, as reported in Appendix B.

popularized in the rank-effect literature (Denning et al., 2021; Elsner and Isphording, 2017, 2018; Murphy and Weinhardt, 2020).

Intuitively, my empirical strategy relies on the observation that the ranks of students with the same socioeconomic status can vary substantially across cohorts within the same school. Such variation arises naturally due to fluctuations in the household characteristics of children of school starting age in a school's catchment area over time. As a consequence, I observe "similar" students with the same level of SES, but different relative positions within their cohorts in the same school. Roughly speaking, viewing the between-cohort fluctuations as idiosyncratic allows me to use the within-school differences in SES distributions across cohorts to estimate a causal rank effect. Formally, this view justifies an exogeneity assumption that identifies a causal parameter in a linear fixed effects model.

My empirical analysis is based on data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a nationally representative study in the U.S. that follows students in several waves from their time during high school into adulthood. The Add Health data has four characteristics that make it particularly suitable for my research question. First, it contains detailed information on the school and cohort membership of the surveyed students, providing me with the information necessary to construct cohort networks. Since the primary sampling unit of the survey are schools, the network data is "complete" in the sense that I observe all students within each cohort. Second, it covers multiple cohorts within the same school, a feature that is key for my empirical strategy as outlined above. Third, it contains rich information on students' backgrounds, including parental education, allowing for the construction of different measures of socioeconomic status. Fourth, the data set contains well established outcome measures for depression and cognitive ability. Specifically, depression is measured using the Center of Epidemiological Studies Depression Scale (CES-D; Radloff, 1977) and cognitive ability is measured using the Peabody Picture Vocabulary Test (PPVT; Dunn and Dunn, 2007), an age-specific standardized ability test. Moreover, the data set contains six items similar to or modified from the original Rosenberg self-esteem scale (Rosenberg, 1965) as well as information on friendship networks that allow for the construction of a measure of popularity. The latter two outcomes are closely linked to mental health and social status within a peer group and, taken together with the main outcomes, provide a more comprehensive picture of adolescents' development. Finally, students in the Add Health survey are tracked over a long period of time, allowing me to investigate whether the socioeconomic rank has effects that persist into adulthood, more than 10 years after the initial interviews took place.

My analysis produces three main findings. First, a student's SES rank in their high school cohort has a significant and economically meaningful impact on their development in the short run. Holding the level of socioeconomic status fixed, students with a higher within-cohort rank tend to have better outcomes in terms of depression and cognitive ability. These results are supported by

analogous findings which show that, ceteris paribus, higher ranked students develop higher levels of self-esteem, and are more popular, as measured by different concepts of network centrality. Increasing a student's rank by 25 percentiles (i.e. one standard deviation), decreases depression scores by 13% of a standard deviations and increases cognitive ability and self-esteem scores by 13 and 12% of a standard deviation, respectively. Further, such a rank shift leads to a 10% standard deviation increase in a student's popularity among their peers. To put theses figures into context, I compare the rank effects on depression and cognitive ability to the effects of school quality. Using the school fixed effects as a rough benchmark for school quality, I find that increasing a student's SES rank by one standard deviation is equivalent to increasing school quality by about 50 - 60% of a standard deviation.

Second, these rank effects vary by the degree of cohort inequality, with steeper rank gradients occurring in cohorts with high levels of SES-inequality across all outcome dimensions. These documented patterns may be of independent interest and are consistent with the predictions of a relative deprivation mechanism, suggesting that the salience of inequality is important.

Third, the effect of the socioeconomic rank during high school persists in the long-run. Students with a higher within-cohort rank during high school tend to have better mental health and educational outcomes in adulthood. Increasing a student's cohort rank by 25 percentiles increases the probability of attending and completing college by 5 and 4 percentage points, respectively, and decreases adult depression scores by 0.12 standard deviations. The latter result is consistent with the documented high persistence of depressive symptoms over the life-cycle and emphasizes the importance of adolescent mental health.

Overall, my results suggest that the relative socioeconomic status is an important determinant in shaping adolescents' outcomes, supporting the view that social context should not be treated as a second-order concern when studying human development.[4] My findings can be viewed as a justification for the design and implementation of interventions aimed at mitigating the adverse consequences of relative deprivation.[5] In practice, such efforts could entail interventions aimed at reducing the salience of inequality in schools, such as the provision of school uniforms, subsidized school meals and leisure activities or targeted programs that take into account not only the level of SES, but also the relative position of an individual in a given social environment.

Related Literature My work builds on and contributes to a large body of literature that seeks to understand the determinants of human capital formation and the role of mental health.[6] A

[4] This view is consistent with findings from Butikofer et al. (2020), who show that the school environment causally affects adolescents' mental health and educational attainment.

[5] Importantly, my findings should not be interpreted in support of policies furthering segregation by SES. Such a view would neglect the endogenous consequences of modified peer characteristics.

[6] Influential examples include Cunha and Heckman (2007), Cunha et al. (2006, 2010), Currie and Stabile (2006), and Currie et al. (2010).

consistent finding in this literature is that circumstances and investments early in life have a disproportionate impact in shaping long-term outcomes (e.g. Campbell et al., 2014; Currie, 2009) and that large socioeconomic gaps open up at early ages and persist into adulthood (e.g. Carneiro and Heckman, 2003; Cunha et al., 2006; Currie and Goodman, 2020). These shared patterns are perhaps unsurprising, as concepts of mental health and non-cognitive skills, an important component of human capital, tend to overlap. Moreover, there is evidence that mental health affects processes relevant for the development of cognitive skills (Currie and Stabile, 2006, 2009) and that there are feedback effects of human capital on mental health.

Relating to the large and persistent socioeconomic gaps in (mental) health and human capital outcomes and their implications for lifetime inequality and intergenerational mobility, a growing literature provides estimates of the causal effect of parental background on life outcomes of children and adolescents, providing evidence on the effect of parental education and income on their children's cognitive and non-cognitive ability (Dahl and Lochner, 2012; Lundborg et al., 2014; Milligan and Stabile, 2011), educational attainment (Black et al., 2005; Holmlund et al., 2011; Oreopoulos et al., 2006) as well as health (Lundborg et al., 2014; Milligan and Stabile, 2011).

I contribute to this literature by providing causal evidence on how parental socioeconomic status affects adolescents' mental health and human capital formation. In contrast to the previous literature, which studies the impact of absolute measures of socioeconomic status, I investigate the role of mechanisms that operate through the relative status of a student within their peer group. I draw on a rich theoretical literature[7] in sociology and social psychology that emphasizes the importance of social context, in particular social comparisons and relative deprivation, for individuals' self-evaluations, development and behavior. I investigate the empirical content of these theories in relation to the relative socioeconomic status by applying modern quasi-experimental techniques recently popularized in the rank effects literature discussed below. The importance of relative status for adolescents' mental health is supported by recent evidence from Braghieri et al. (2022), who suggest unfavorable social comparison as the mechanism through which social media negatively affects mental health.

Methodologically, my work is closely related to a growing empirical literature on ordinal rank effects. In particular, a series of recent papers have investigated how relative ability rankings during adolescence impact individuals' educational outcomes and behaviors. This line of work is motivated by the idea that individuals calibrate the perception of their abilities via peer comparisons with consequences for their educational attainment and choices (Delaney and Devereux, 2019; Denning et al., 2021; Elsner and Isphording, 2017; Elsner et al., 2021; Murphy and Weinhardt, 2020), risky behaviors (Elsner and Isphording, 2018), as well as the development of personality traits (Pagani et al., 2019) and depression (Kiessling and Norris, 2022).

[7] Examples include Crosby (1976), Festinger (1954), Stouffer et al. (1949), Wills (1981), and Wood (1989).

One concern in the literature on cognitive ability rankings is that, in the absence of predetermined measures of ability, measures of cognitive ability are endogeneous and existing studies partly rely on strong assumptions to claim causality. For instance, in order to regard their measure of ability as predetermined, Kiessling and Norris (2022) assume that cognitive ability is stable after the age of 10, thus not influenced by e.g. the school environment or peers. In contrast, by studying the importance of the relative socioeconomic status, measured using predetermined parental characteristics such as education, this paper departs from such assumptions and instead treats cognitive ability as an outcome variable. In fact, my findings provide evidence that the relative socioeconomic status not only influences students' mental health, but also the development of their cognitive ability. Thus, while I employ a similar fixed-effects strategy, the challenges I face are different as I discuss in Section 3. More generally, I add to the existing literature on rank effects by focusing on a different dimension along which individual comparisons may matter. This distinction is important as comparisons and relative standings along different dimensions can have vastly different mechanisms and policy implications.

Also closely related to my work are Balsa et al. (2014) and Arduini et al. (2019), who find that differences relative to average peer characteristics in terms of socioeconomic status and body mass index impact risk-taking behavior as measured by alcohol consumption and smoking for young males, as well as eating disorders in female teenagers. The status concerns underlying such comparison mechanisms have also been investigated in adult populations, by studying the impact of relative positions on job satisfaction (Card et al., 2012) and general well-being and satisfaction (Brown et al., 2008). The results reported in these papers are consistent with the findings in Luttmer (2005) and Clark and Oswald (1996), who provide evidence that satisfaction and well-being depends on income relative to an environment specific reference level.

On a conceptual level, my work is also related to a vast literature on peer effects in education (e.g. Bifulco et al., 2011; Carrell et al., 2018; Sacerdote, 2011), in that I recognize the importance of peer groups. In contrast to this literature, I emphasize an individual's relative position within their peer group rather than the effects of absolute measures of peer characteristics, which I treat as nuisance parameters in my model.

Outline of the Paper The rest of this paper is organized as follows. Section 2 describes the Add Health data and the construction of relevant variables. Section 3 presents my empirical strategy and discusses threats to the identification of my model. In Section 4, I present the results of my empirical analysis. Section 5 discusses potential policy implications of my findings and concludes.

2 Data

The data set used for the empirical analysis is the National Longitudinal Study of Adolescent to Adult Health (Add Health; Harris, 2018), explicitly designed to study the link between the social environment and adolescents' health and health-related behavior. During the school year 1994/95, all students in the grades 7-12 of 80 nationally representative high schools and 52 middle schools in the US completed an in-school survey. General student and parental background information, health and health-related behavior as well as information about the school and social network were collected for more than 90,000 adolescents between the age of 12 and 20. Moreover, a sample of around 20,000 students additionally completed a more comprehensive in-home questionnaire with detailed information on behavior, characteristics and health status. Respondents from this in-home-interview (wave I) were followed and re-interviewed in four subsequent waves, administered in 1996 (wave II), 2001-02 (wave III), 2008-09 (wave IV) and 2016-2018 (wave V).[8]

The Add Health survey exhibits four features that are key for the analysis in this paper. First, it contains detailed information on the school and cohort of a student, allowing me to identify the cohort network of a student. Second, it covers multiple cohorts within the same school, allowing me to employ a fixed effects strategy with separate school and cohort fixed effects or school-by-cohort fixed effects. Third, it contains detailed information on the students' background, including parental education as a measure of students' socioeconomic status. Fourth, the data set contains well-established measures of mental health, cognitive ability, and self-esteem as well as information about the friendship networks of students. The scope and detail of the survey questions allow me to obtain an accurate and comprehensive picture of adolescents' development. Finally, students from the in-home sample are tracked over a long period of time, allowing me to study the long-term impacts of relative socioeconomic status during high school.[9]

2.1 Outcome Measures

For the analysis of the short-run effects, the main outcomes I focus on are depression and cognitive ability as adolescence is a critical time for the development of cognitive processes and the onset of mental health problems. In addition, I also consider potential rank effects on a student's self-esteem and popularity during high school. These outcomes are closely linked to mental health and social status within the peer group and, taken together with the two main outcomes, provide a more comprehensive picture of adolescent development. In order to study potential long-run effects of the socioeconomic rank, I look at depression as well as educational attainment in adulthood, more

[8] For further information on the Add Health research design, see Harris et al. (2019).

[9] For the construction of long-term outcomes, I use information from wave IV as this is the most recent data currently available to me.

than 10 years after the initial interview took place.

Depression Depression is a common mental disorder with potentially long-lasting effects on the individual's quality of life. In this paper, it is measured using the Center for Epidemiologic Studies-Depression Scale (CES-D), a validated international screening test designed to measure depressive symptoms in the general population (Radloff, 1977). The CES-D is one of the most commonly used self-reported measures of depressive symptoms. Psychometric properties in terms of its concurrent validity (i.e. the degree of agreement between the CES-D score and the diagnosis), reliability and internal consistency of the CES-D have been demonstrated to be good in a wide range of clinical and non-clinical populations, including adolescents (see e.g. Lewinsohn et al., 1997; Radloff, 1991; Roberts et al., 1990). The CES-D in the Add Health questionnaire consists of 19 items (e.g. *"You felt sad"*), assessing the frequency with which an individual experiences symptoms associated with depression over the course of the past week.[10] Responses are rated on a scale from 0 (*"never or rarely"*) to 3 (*"most of the time or all of the time"*), resulting in an aggregated measure of the CES-D ranging from 0 to 57, with higher values indicating worse depressive symptoms. Respondents with a score equal to or above 16 are commonly identified to be at risk for clinical depression (Beekman et al., 1995; Radloff, 1977). In the main analysis, I use the aggregate CES-D score as a measure of depression, however, I also use the cut-off of 16 as an indicator for clinical depression in Appendix C.3.

Cognitive Ability As a measure of cognitive ability, I use the Adolescent Health Picture Vocabulary Test (AHPVT), an adapted 87-item version of the Peabody Picture Vocabulary Test (PPVT; Dunn and Dunn, 2007). The Peadody is an assessment of a student's receptive vocabulary and is used to measure verbal intelligence and scholastic aptitude. The test is age-specific and scores are standardized to a mean of 100 and standard deviation of 15 within each age group.

Self-esteem As a measure of a student's self-esteem, I use an adapted 6-item version of the original Rosenberg self-esteem scale (Rosenberg, 1965). The Rosenberg scale asseses an individual's perception of self-worth. In the Add Health data set, students were asked whether they agree or disagree with statements such as *"you have many good qualities"* or *"you have a lot to be proud of"*.[11] Items are scored on a 5-point Likert scale ranging from 1 (*"strongly agree"*) to 5 (*"strongly disagree"*). For the construction of the self-esteem measure, these items are reverse coded to a scale from 0 (*"strongly disagree"*) to 4 (*"strongly agree"*) and aggregated to obtain a score ranging from 0 to 24 such that higher values indicate higher levels of self-esteem.

[10] See Table A.1 in Appendix A for an overview of all items.
[11] See Table A.3 in Appendix A for an overview of all items.

Popularity A student's popularity among their high school peers can be regarded as a reflection of social status and peer acceptance, factors that are essential in the development of adolescents. Having good social relations can have a positive impact on their feelings of self-worth and depressive symptoms. Moreover, adolescents' popularity during high school can be an important predictor of adult success. It has been shown that there is a wage premium associated with a student's popularity, as measured by the number of received friendship nominations (Conti et al., 2013). I use methods from social network analysis to derive a measure of a student's popularity based on their friendship network. In particular, I use detailed information on high school friendship relations collected in the Add Health data set. During the in-school survey in 1994/95, all attending students from each participating school were asked to nominate up to five male and five female friends from a given school roaster. This allows for the construction of different measures of centrality and prestige to describe a student's popularity within their school network.

The simplest measure of centrality is *degree centrality*. In a directed network, one can distinguish between *in-degree* and *out-degree centrality*. In the context of friendship nominations, in-degree (out-degree) centrality counts the number of incoming (outgoing) ties to a node, i.e. the number of friendship nominations each student receives (nominates).

In contrast to degree centrality, where each connection gets equal importance in the construction of a popularity measure, other indicators are based on the idea that some friendship connections should count more than others. One such concept in the Add Health data is *Bonacich centrality*, where individual i's centrality depends not only on the number of friendship connections, but also on the centrality of all individuals she sends ties to. Formally, it is defined as a measure of eigenvector centrality, $B(\alpha,\beta)$:

$$B(\alpha,\beta)_i = \alpha (I - \beta X)^{-1} X 1, \tag{1}$$

where α is a scaling factor; β is the power weight reflecting the extent to which i's centrality depends on the centrality of others (which is set to 0.1); I is the identity matrix; X the adjacency matrix of the entire friendship network; and 1 a column vector of ones.

Alternatively, the measure of *proximity prestige* is based on the number of incoming ties and weighs the fraction of individuals that are in individual $i's$ influence domain (i.e. the fraction of individuals that can reach i) by the average geodesic distance in the influence domain of i. Formally,

$$P_i = \frac{\frac{I_i}{g-1}}{\sum_j \frac{d\left(n_j,n_i\right)}{I_i}}, \tag{2}$$

where I_i is individual i's influence domain, which is the number of individuals it can reach; g is the number of nodes in the friendship network; and $d\left(n_j,n_i\right)$ is the geodesic distance between

individual j and i. All measures are standardized within a school cohort to account for differences in cohort size.

Each of these measures captures different aspects of centrality. In the main part of this paper, I focus on Bonacich centrality to measure students' popularity because it not only relies on the number of friendship ties, but also takes into account the popularity of one's friends. However, since this measure relies on outgoing ties, which could be endogenous to e.g. a students' mental health, I verify the results using different measures of popularity – including a student's indegree, outdegree, and proximity prestige - in Appendix Table C.4.

Long-run Outcomes In order to study long-term effects of a student's socioeconomic rank, I use information from wave IV, when individuals were between 24 and 32 years old, to construct measures of mental health and educational attainment. In particular, I use a short version of the CES-D questionnaire as an indicator for depression. The shorter CES-D score is based on 10 items, thus ranges from 0 to 30.[12] Moreover, I use a student's educational attainment in the form of college attendance and college completion dummies to obtain measures for human capital accumulation in the long-run.

2.2 Socioeconomic Rank

To measure a student's position in comparison to their school peers, I construct their ordinal rank in terms of the socioeconomic background within their school cohort.[13] I follow Balsa et al. (2014) and define an adolescent's socioeconomic status in terms of the highest level of schooling completed by the student's head of the household.[14] In the Add Health data, parental schooling is reported by the students as a categorical variable which I translate into years of schooling by using the midpoints of these categories.[15] One key advantage of using parental schooling as a measure of adolescents' socioeconomic status is that this information is available for all students participating in the in-school questionnaire.

The ordinal rank of a student measures their households' relative position in the distribution of parental socioeconomic status within their school cohort. In a cohort with N students, the student

[12] The 10 items asked in wave IV are indicated with an asterisk in Table A.1

[13] A student's cohort refers to all students attending the same grade at the same school and time.

[14] The head of the household is assumed to be the father unless the respondent reported not living with him or the information is missing. In that case, the mother's highest level of schooling is taken.

[15] Following Balsa et al. (2014), five years of schooling are assigned to parents who completed eight or fewer years of schooling, including those cases in which the child indicated that the parent never went to school or did not know which level the parent completed. Moreover, I assign 10.5 years to parents who completed the eighth grade but did not graduate from high school, 11.5 years to parents who completed a GED, 12 years to parents who graduated from high school, 13.5 years to parents who attended a business, trade, or vocational school after high school, 14 years to parents who received some college education, 16 years to parents who graduated from a college or university, and 20 years to parents who acquired professional training beyond college.

with the lowest status in a cohort is assigned position 1 and the student with the highest status is assigned position N. To account for differences in school cohort size, the student's raw rank is translated into a percentile rank. In particular, the rank of individual i in school s and cohort c is then measured as

$$\text{Rank}_{isc} = \frac{\text{n}_{isc} - 1}{\text{N}_{sc} - 1}, \tag{3}$$

where N_{sc} is the cohort size of school s and cohort c and n_{isc} is student i's ordinal rank position in their school cohort. Rank_{isc} is the percentile rank of student i, ranging from 0 for the lowest ranked student to 1 for the highest ranked student in a given cohort. In the case of ties, students are assigned an average rank.[16]

Given that I observe parental education as a categorical variable with only eight different values, one obvious concern in the construction of the rank is that ties occur frequently and may primarily be responsible for the variation in the rank variable. To address this concern, I use alternative definitions of parental education, including the average educational attainment of both parents. This definition of socioeconomic status helps alleviating such concerns because the variation in the level of parental SES is higher as the SES variable can take on $8^2 = 64$ different values and ties occur less frequently. Robustness checks in Section 4.2 show that the main results are robust to this alternative SES definition.

2.3 Descriptives

Sampling Criteria and Weights For the analytical sample, I only keep individuals with information from both the in-school survey and the in-home survey.[17] Students with missing information on parental socioeconomic background are dropped from the analytical sample. Moreover, I drop students with conflicting school identifiers and students in schools with less than 20 students or cohorts with less than 5 students. Finally, I only keep students with complete information on age, gender and race with at least one non-missing short-run outcome variable. These sampling criteria result in a sample of 13,736 students that were assigned sampling weights.[18]

Descriptive Statistics Table 1 reports summary statistics to describe the main outcome variables as well as sample characteristics of the respondents at the time of the initial interview (wave I) in

[16] As a robustness check in Section 4.2, I use different ways of breaking ties. These analyses yield very similar results and can be found in Appendix B.

[17] This decreases the analytical sample considerably because some schools did not participate in the in-school survey or information on the student's identifier in the school interview is missing. In the main analysis, I exclude these individuals.

[18] The in-home survey of the Add Health data oversamples some groups, thus I use sampling weights in the regression analysis to account for this sampling design. See Chen and Chantala (2014) for details. Results without sampling weights are provided for the main regressions in Appendix C.6.

Panel A-C. I report the mean, standard deviation, and interquartile range. 48% of the students are male, 51% are white, 21% are black, 7% are asian and 16% have a hispanic background. At the time of the initial interview, the average student age is 15.6 years. The students' depression scores range between 0 and 57 with a mean of 11. Around 20% percent of students score equal to or above a score of 16, a commonly used cutoff to indicate individuals at risk of clinical depression.[19] The average self-esteem score in the sample is 17. Popularity and cognitive ability, by construction have a mean close to 100 and 0, respectively. The analytical sample consists of 120 different schools of which 22% have fewer than 401 students, 48% have between 401 and 1,000 students and 30% have more than 1,000 students. The sample consists of 421 different school cohorts[20] with an average cohort size of 182.

Attrition Of the 13,736 individuals from the main analysis, 10,912 remain in the sample of wave IV for the long-run analysis. The summary statistics of the long-run outcome variables are presented in panel D of Table 1. The average CES-D score is 7.4[21], 69% of the sample has been enrolled in college, 35% have already completed a college degree. A more detailed description of the long-run sample can be found in Table C.1. The average respondent in the long-run sample is 28 years old. The sample characteristics of the 10,912 individuals that remain in the sample for the long-run analysis are fairly similar to the initial sample of 13,736 individuals, indicating that attrition is not a major concern for my analysis of the long-term effects. I further address this concern in Appendix B by showing that attrition status is not related to ranks as shown in Appendix Table B.7. Moreover, I re-estimate the results for the main analysis based on students that remain in the sample in wave IV in the Appendix Table B.8 and find very similar results. I therefore conclude that attrition is unlikely to affect my long-term results.

[19] See Figure D.3 in Appendix D.2 for the distribution of the depression score.
[20] Some schools do not have all grades 7-12.
[21] The CES-D score in wave IV only consists of 10 items instead of the 19 items in wave I.

Table 1: Sample Descriptives

A. Contemporaneous Outcomes (wave I)

	mean	sd	25th	median	75th	n
Depression	11.01	7.46	6.00	10.00	15.00	13,683
Cognitive ability	100.95	14.50	92.00	101.00	112.00	13,115
Self-esteem	17.13	4.49	15.00	18.00	20.00	11,917
Popularity (indegree)	0.05	1.02	-0.70	-0.17	0.58	12,883
Popularity (bonacich)	0.04	1.01	-0.81	-0.08	0.76	12,883

B. Individual Characteristics (wave I)

	mean	sd	25th	median	75th	n
SES	13.38	3.61	12.00	12.00	16.00	13,736
Family income	46.99	54.17	22.00	40.00	60.00	10,331
Age	15.63	1.69	14.00	16.00	17.00	13,736
Male	0.48	0.50				13,736
Ever repeated a grade	0.19	0.39				13,736
White	0.51	0.50				13,736
Black	0.21	0.41				13,736
Asian	0.07	0.25				13,736
Hispanic	0.16	0.37				13,646

C. School and Cohort Characteristics (wave I)

	mean	sd	25th	median	75th	n
School characteristics:						
Small (<401 students)	0.22	0.41				120
Medium (401-1000 students)	0.48	0.50				120
Large (> 1000 students)	0.30	0.46				120
Cohort characteristics:						
Cohort size	181.62	127.73	87.00	159.00	243.00	421
Mean SES	13.45	1.37	12.60	13.30	14.17	421
SD SES	3.05	0.60	2.69	3.03	3.36	421

D. Long-run Outcomes (wave IV)

	mean	sd	25th	median	75th	n
Depression	7.40	3.82	5.00	7.00	9.00	10,901
College attendance	0.69	0.46	0.00	1.00	1.00	10,911
College completion	0.35	0.48	0.00	0.00	1.00	10,911

Note: This table describes the analytical sample for the main analysis. Panels A-C describe the main outcome variables as well as individual, school and cohort level characteristics of the respondents in the sample for the short-run analysis, i.e. at the time of the initial interview (wave I). Panel D describes the outcome variables of the respondents that remain in the long-run sample (wave IV). The table displays the mean, standard deviation, and interquartile range of the variables as well as the number of observations. SES is measured in years of education (as outlined in section 2.2), annual family income is measured in thousands U.S. $.

3 Empirical Strategy

I seek to estimate the causal effect of a student's socioeconomic rank in their high-school cohort on a set of short and long-run outcomes related to adolescent development. To that end, I exploit variation in the socioeconomic composition of different cohorts within the same school. Such variation arises naturally due to fluctuations in the household characteristics of children of school-starting age in a school's catchment area over time. Utilizing only within-school variation allows me to address concerns regarding the non-random selection of students into schools, which confounds estimates based on global comparisons.

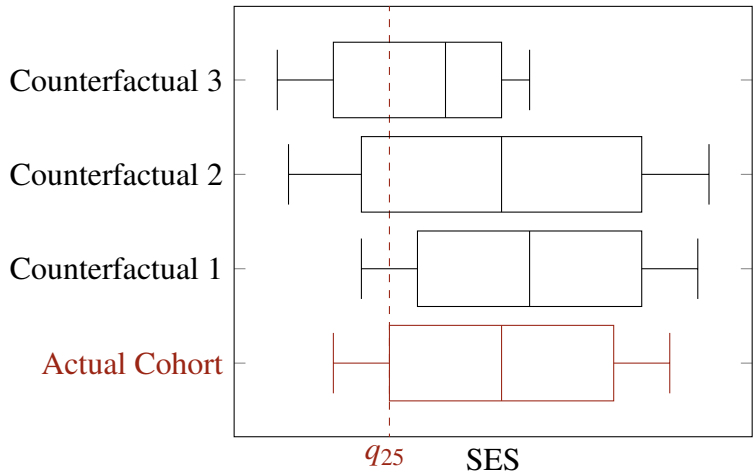
Viewing the observed within-school variation in ranks conditional on SES as quasi-random motivates a conditional mean independence assumption that identifies the causal effect of the rank variable in a linear fixed effects model. In the following, I begin by describing the mechanisms that generate the identifying variation in the composition of cohorts before discussing the functional form of my model and potential threats to its identification.

3.1 Intuition

Intuitively, the idea underlying my empirical strategy is the following counterfactual thought experiment: A student of a given socioeconomic background would potentially have had a different rank, had they been a member of a different cohort in their school. With this in mind, I estimate counterfactuals by comparing the outcomes of students of the same socioeconomic background in the same school that differ with respect to the socioeconomic rank assigned to them in their respective cohort. Consequently, my strategy requires variation in ranks within school-SES strata across cohorts. This identifying variation is generated by within-school differences in the shape of the SES distributions across cohorts. For example, consider a student of a given socioeconomic background in a given school and cohort such that the student is located at the 25th percentile in their actual cohort SES distribution. Figure 1 illustrates how this student's rank would have differed in counterfactual cohorts that differ from the factual cohort distribution with respect to the mean (Counterfactual 1), the variance (Counterfactual 2), or in general shape (Counterfactual 3). In each counterfactual cohort, the student factually positioned at the 25th percentile would have been assigned a different rank. My empirical strategy seeks to recover the causal effect of a student's relative socioeconomic cohort rank using such within-school across-cohort comparisons.

In theory there are variety of mechanisms that can generate within-school between cohort differences in SES distributions. For example, variation in the timings of birth around school year specific enrolment dates can generate differences in the share of highly educated parents between cohorts. Similarly, variations in cohort sizes, as explained by Hoxby (2000), are likely to induce

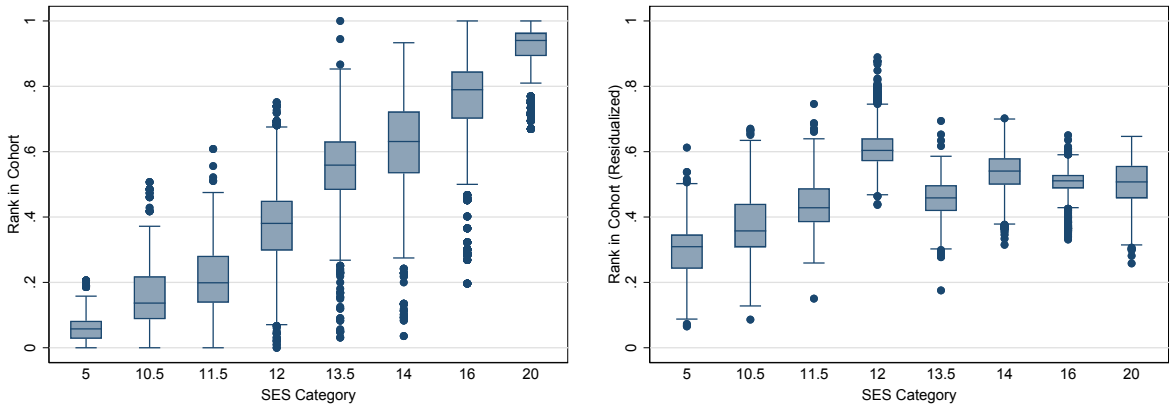
Figure 1: Illustration of Identifying Variation



Note: This figure illustrates how differences in the SES distribution across cohorts lead to variation in the rank variable for a fixed level of SES. The figure depicts a hypothetical cohort (red) and fixes the SES level of a student ranked at the 25th percentile in this cohort. In the three counterfactuals, I show how a student with the fixed level of SES would be ranked in cohorts with a different mean (counterfactual 1), a different variance (counterfactual 2), or a generally different shape in the SES distribution (counterfactual 3).

differences in the shape of cohort SES distributions. While such differences are typically negligible on aggregate levels such as school districts, they can produce pronounced differences on the school level, provided there is some heterogeneity in the types of households attracted by each school. The extent to which such variation exists in a given data set is an empirical question. Figure 2 shows the variation in cohort ranks within each SES category for the schools and cohorts sampled in the Add Health survey.

Figure 2: Unconditional and Conditional Variation in Ranks



(A) Unconditional Variation. (B) Conditional Variation.

Note: This figure plots the variation in SES ranks for each education category (5 "8th grade or less", 10.5 "Completed 8th grade, but no high school degree", 11.5 "GED", 12 "High school degree", 13.5 "Business or vocational school after high school", 14 "Some college", 16 " College degree", 20 "Professional degree"). For each category, I display the median, the 25th and 75th percentiles, and the minimum and maximum of the rank distribution. In panel A, I plot the unconditional variation in ranks. In panel B, I present the variation in ranks conditional on separate school and cohort fixed effects as well as individual and school cohort specific controls.

While Panel A shows the unconditional variation in ranks within each SES category, Panel B displays the variation conditional on separate school and cohort fixed effects as well as individual and cohort level observables used in my preferred model specification. The figure illustrates three important points: First, globally there is substantial variation in ranks within each SES category. Second, unsurprisingly, most variation is observed around the center of the SES distribution, where almost all ranks are observed in certain cohort environments. Finally, the conditional variation in ranks is substantially smaller, which has important implications for the interpretation of my estimates, as it illustrates what type of counterfactuals my estimates are based upon. This is important to keep in mind when interpreting the rank coefficients in my model and extrapolating towards "extreme" counterfactuals. Specifically, it is unlikely that, for a given level of SES, a student is ranked top in one cohort and bottom in a different cohort in the same school. The last observation also illustrates the main practical challenge reflected in my modelling choice: I seek to solve a trade-off between flexibility and precision. While more flexible functional forms mitigate misspecification concerns, they come at the cost of less precise estimates. This is because in order to pin down the rank effect, I require sufficient variation in ordinal ranks within the strata defined by my model.

3.2 Empirical Model

I impose the following general additively separable fixed effects model that relates the outcome y_{isc} of student i in school s and cohort c to their cohort rank according to

$$y_{isc} = \beta \text{Rank}_{isc} + f(\text{SES}_{isc}) + \gamma \mathbf{X}_{isc} + g(s,c) + u_{isc}. \tag{4}$$

As discussed in Section 2.2, the rank variable is approximately uniformly distributed on $[0,1]$ by construction. The vector X_{isc} contains predetermined individual-level characteristics such as age in days, gender and ethnicity. The functions f and g denote flexible functional forms of a student's level of SES as well as different school and cohort fixed effects specifications.

The model parameter of interest is β, which captures the causal effect of the ordinal rank on the respective outcome. Note that, while my counterfactual thought experiment compared students within schools, the constant effects assumption underlying β justifies across-school comparisons in residualized outcomes and ranks. Following textbook arguments, β is identified under the following strict exogeneity assumption:

$$\mathbb{E}[u_{isc} | \text{Rank}_{isc}, \text{SES}_{isc}, \mathbf{X}_{isc}, g(s,c)] = 0. \tag{5}$$

The strict exogeneity assumption (5) is conditional on the functional form assumption in equation (4) in the sense that its interpretation and plausibility depend on the choices for the functions f and g. Consequently, the key challenge is to parameterize these functions such that assump-

tion (5) is plausible, keeping in mind the flexibility-precision trade-off mentioned above. For f, I consider different dummy-specifications that capture SES-bin specific averages ($f(SES_{isc}) = \sum_{j=1}^{K} \delta_j D_j(SES_{isc})$).[22]

For g, I consider three different choices: (i) separate school (120) and cohort (6) fixed effects ($g(s,c) = \lambda_s + \lambda_c$), (ii) separate school and cohort fixed effects augmented by school cohort specific control variables ($g(s,c) = \lambda_s + \lambda_c + \alpha \mathbf{W}_{sc}$), as well as (iii) school-by-cohort (421) fixed effects ($g(s,c) = \lambda_{sc}$).

My initial model contains separate school and cohort fixed effects. This model uses variation in the socioeconomic rank within schools and rules out systematic self-selection of students into schools as a confounding factor. In this model, the strict exogeneity assumption requires that all cohort level unobservables are uncorrelated with the rank variable. This model is best viewed as a rough approximation, as school cohort specific characteristics such as the average SES are mechanically correlated with the rank and likely to have an effect on outcomes via traditional peer effect mechanisms as pointed out in Elsner and Isphording (2017). Arguments along these lines motivate my second model, where I include observable school cohort characteristics to mitigate omitted variable concerns at the cohort-level. Specifically, $\mathbf{W}_{sc}$ includes the mean and standard deviation of cohort-SES, the fraction of repeaters, the gender composition, and share of white students in each school cohort.

While including a set of school cohort specific charactristics makes the strict exogeneity assumption appear more plausible, I cannot rule out the existence of relevant unobserved cohort characteristics that impact outcomes via less obvious peer-effect mechanisms. In particular, Elsner and Isphording (2017) discuss dynamic selection along unobservable cohort characteristics as a potential threat to the strict exogeneity assumption. Such concerns motivate my third model which includes school-by-cohort fixed effects, effectively ruling out that school cohort specific confounders drive my estimation results. This approach compares students across all school cohorts after removing all school cohort specific mean differences. Note, that in this model β is still estimable from differences in the shape of the SES distribution.

The last specification of my model guards my empirical results from potential confounding caused by school and school cohort specific unobservable characteristics. However, strict exogeneity also posits the absence of individual level unobservables that correlate with the residualized rank. While my research design does not allow me to rule out the existence of such individual level confounders, the institutional setting I study provides some arguments mitigating such concerns. The arguably most important behavioral assumption that I rely on is that parents cannot

[22] In my preferred specification, I assign the SES levels to four different categories: "high school or less", "some college or vocational training", "college", and "postgraduate". The grouping of SES categories into bins is varied in Appendix B. Alternatively, I also consider a linear and quadratic function of SES.

exactly anticipate the relative socioeconomic rank of their child in a specific cohort when making their school choice. While my design allows and accounts for choices based on (unobservable) school and cohort characteristics, school choices based on ranks would violate strict exogeneity. Abstracting from the fact that it appears unlikely that parents have the necessary information to make such a choice, rank based school choices would likely lead to strategic delays in enrolment, since there are limited school options available in each school district. Appendix B contains evidence showing that the data does not support the notion of strategic enrolment delays, suggesting that rank based school choices are not a major concern for my analysis.

Balancing Tests To support the approach of using variation in students' ranks, stemming from the differences in SES-compositions across cohorts, to identify causal effects, I conduct a set of balancing tests. The purpose of these balancing tests is to examine whether the SES rank, conditional on controls as well as school and cohort fixed effects, can explain predetermined student characteristics. Specifically, I use polygenic scores (PGSs; Braudt and Harris, 2020) for different traits, behaviors, and diseases (phenotypes) as outcome variables in these balancing tests because PGSs are based only on individuals' genetic code, thus measures that are fixed at conception. In wave IV of the Add Health survey, saliva samples were collected and, following quality control procedures, genotyped data for 9,974 individuals in four genetic ancestry groups (European ancestry, African ancestry, Hispanic ancestry, and East Asian ancestry) is available. Based on this sample, Add Health constructed different polygenic scores in accordance with summary statistics stemming from genome-wide association studies (GWASs). These PGSs represent an association between allele frequencies at individual single-nucleotide polymorphisms (SNPs) and different phenotypes. In the current context, I focus on PGSs for educational attainment (Lee et al., 2018; Okbay et al., 2016) and different types of mental disorders, including Attention-Deficit/Hyperactivity Disorder (Demontis et al., 2017; Neale et al., 2010), Bipolar Disorder (Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011), Major Depressive Disorder (Psychiatric GWAS Consortium et al., 2013; Wray et al., 2018), Schizophrenia (Psychiatric Genomics Consortium, 2014), and Mental Health Cross Disorder (Psychiatric Genomics Consortium et al., 2013). All polygenic scores are standardized to have a mean of zero and standard deviation of one within ancestry groups to account for between group population stratification. Since GWASs are predominantly conducted on groups of European ancestry, I restrict the sample to this ancestry group, which leaves me with a sample of 3,975 (3,961) individuals in the specification with separate school and cohort (school-by-cohort) fixed effects.[23]

The balancing tests for the PGSs are presented in Appendix Table C.2. I report the coeffi-

[23] While this sample is very restrictive, regression analyses based on this sample yield similar results in terms of the rank effects on the main outcomes.

cients (and standard errors) for separate regressions that include all controls from equation (4) and either account for separate school and cohort fixed effects (columns 1) or school-by-cohort fixed effects (column 2). Appendix Table C.2 reveals no evidence that students socioeconomic rank is systematically correlated with students' polygenic scores on education and mental health disorders. Regression coefficients on the socioeconomic rank are statistically insignificant, thus lending credibility to the empirical strategy used to identify the causal effect of a student's rank.

4 Results

This section presents the results of my empirical analysis. I first show the average effect of a student's socioeconomic rank on contemporaneous outcomes of adolescent development, specifically depression, cognitive ability, self-esteem, and a student's popularity. In Section 4.2, I show that these result hold for a series of robustness checks. I then study potential heterogeneities in the rank effect (section 4.3), emphasizing the role of inequality in exacerbating the impact of the socioeconomic rank. In section 4.4, I proceed to look into the long-term effects of socioeconomic rank on depression and educational attainment and how much of the long-run effects are mediated through the observed contemporaneous effects (section 4.5).

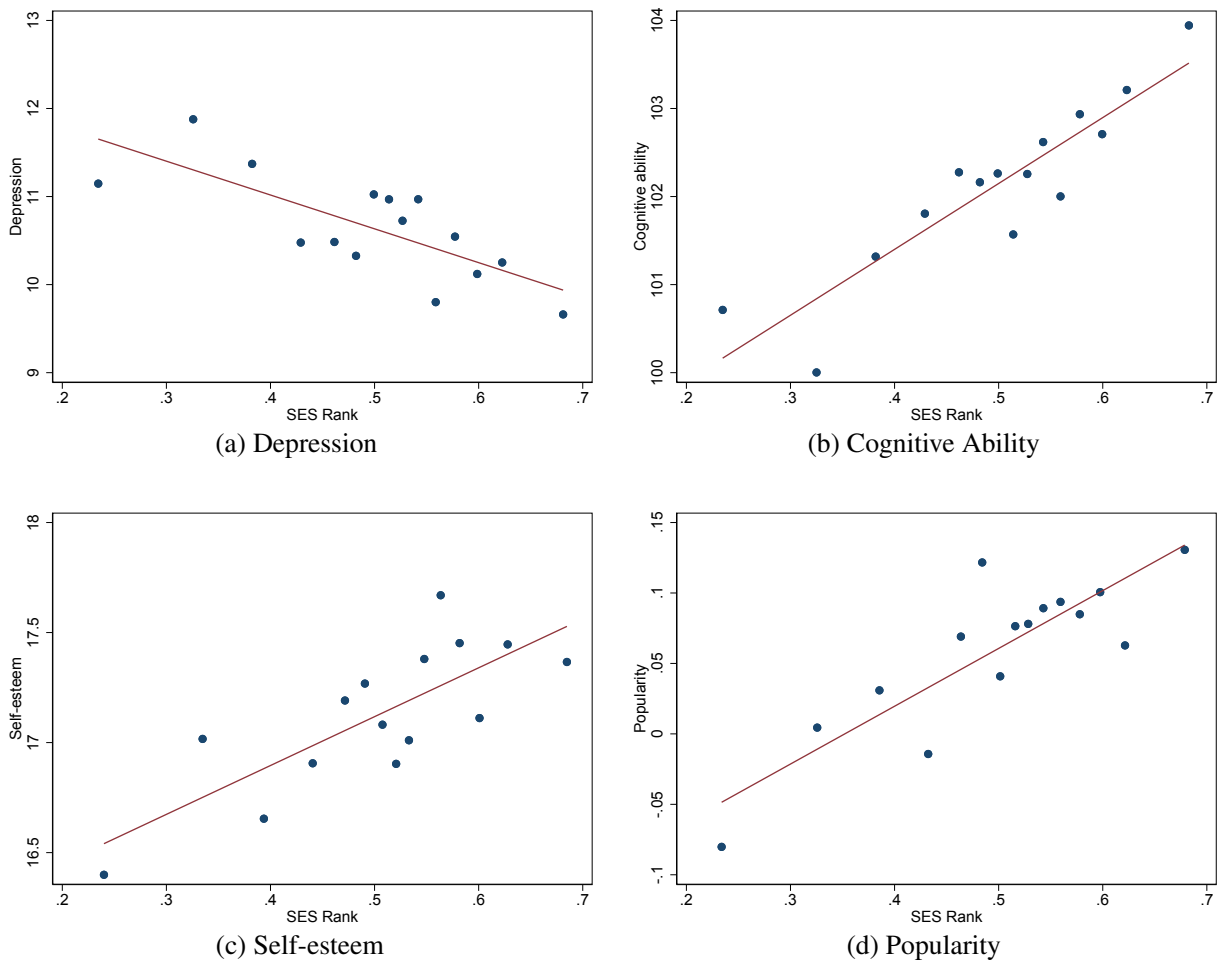
4.1 Average Effect of the Socioeconomic Rank

In this section, I analyze the effect of a student's socioeconomic rank within a school cohort on the contemporary outcomes depression, cognitive ability, self-esteem and popularity. Figure 3 visualizes OLS regressions of equation (4) for each of these outcomes with separate school and cohort fixed effects as well as all individual and school cohort-level controls. I find a negative relationship between the socioeconomic rank and depression: for a given level of socioeconomic status, a higher rank reduces the student's depression score, that is a higher rank is associated with lower depressive symptoms. Conversely, cognitive skills, self-esteem and popularity are positively related to the socioeconomic rank. Students with a higher rank have better cognitive skills, higher levels of self-esteem and are more popular in comparison to their cohort peers.

These findings are substantiated in Table 2 which reports the β-coefficients for different specifications of equation (4) for each of the four outcome variables: depression[24], cognitive ability, self-esteem, and popularity. When interpreting the rank coefficients, it is important to keep in mind that, while the rank variable as defined in Section 2.2 has the support [0,1], extreme counterfactuals are unlikely to occur within a given school. In fact, Figure 2 demonstrates that students of the

[24] In Appendix Table C.3, I additionally estimate the rank effect on the probability to be classified as being at risk for clinical depression, measured as an indicator variable for CES-D $\geq$ 16.

Figure 3: Average Effect of the Socioeconomic Rank



(a) Depression (b) Cognitive Ability

(c) Self-esteem (d) Popularity

Note: Each panel visualizes the effect of the socioeconomic rank based on the linear fixed effects specification in equation (4), accounting for the level of SES, individual (age in days, gender, race) and school cohort specific (mean and standard deviation of SES, fraction of repeaters, male share, and share of white students in the cohort) controls as well as for separate school and cohort fixed effects. Both the x- and y-variables are residualized and the sample mean of each variable is added back to the residuals. The panels display the average values of (a) depression, (b) cognitive ability, (c) self-esteem, and (d) popularity for 15 equally large rank bins.

same socioeconomic background are not ranked top in one cohort and ranked bottom in a different cohort of the same school. Within a given school, the variation in the rank variable for a given level of SES is much smaller. In order to facilitate the interpretation of my results, I re-scale the coefficient estimates to represent a more realistic comparison. In particular, the reported coefficient estimates represent the effect of a 25 percentage point increase in the ordinal rank.[25] That is, the reported coefficients always compare a student that is ranked at, for example, the median to a student that is ranked at the 75th percentile of their cohort SES-distribution.

[25] This is a more realistic counterfactual, but by no means a small change given the conditional variation observed in Figure 2. A 25 percentage point increase approximately corresponds to a one-standard deviation increase in the rank variable.

In column (1) of Table 2, I estimate the rank coefficient, controlling for the level of socioeconomic status as well as separate school and cohort fixed effects. Holding constant the level of socioeconomic status, moving from the median to the 75th percentile rank within a cohort is associated with an improvement of -1.09 points in the depression score, 2.30 points in the cognitive ability test score, 0.49 points on the self-esteem scale, and an increase in popularity by 0.13 standard deviations within a student's cohort. In column (2), when accounting for student characteristics, most of the rank coefficients are moderately smaller in absolute size, but qualitatively robust.

As discussed in section 3.2, this specification is unlikely to fulfill the strict exogeneity assumption because school cohort specific characteristics such as the average SES are mechanically correlated with the rank and likely to have an effect on the outcomes via traditional peer effects. In column (3), I therefore additionally control for school cohort specific characteristics to disentangle the socioeconomic rank effect from potential confounders at the school cohort level. The rank coefficients change only slightly. For a given level of socioeconomic status, increasing a student's rank by 25 percentiles, decreases the depression score by -0.96 points or 13% of a standard deviation[26], and increases the cognitive ability test score by 1.87 points (0.13 standard deviations) and the self-esteem score by 0.56 points (0.12 standard deviations). Further, such a rank shift leads to a 0.10 standard deviations increase in a student's popularity among their peers. These findings hold when estimating equation (4) using school-by-cohort fixed effects in column (4) to absorb all school cohort-specific characteristics as discussed in section 3.2. Overall, the estimated rank coefficients are relatively stable across specifications, lending credibility to the observed rank effects.

In order to get a better idea of the estimated effect sizes, I proceed by comparing these rank effects to the effects associated with a change in school quality. Schools differ along multiple dimensions, such as teacher quality, school facilities, or peer quality, and attending a better school is generally associated with significant gains in students' outcomes. Comparable to Murphy and Weinhardt (2020), I use the size of the school fixed effects from equation (4) as a benchmark for overall school quality. This allows me to compare the estimated rank effects to the effects associated with a change in school quality. For depression and cognitive ability, a one-standard deviation increase in school quality is associated with a 1.6 point decrease in the depression score and a 3.5 point increase in the cognitive ability test scores. This implies that increasing the socioeconomic rank by 25 percentiles, holding constant school quality, is equivalent to increasing school quality by approximately 50-60% of a standard deviation, net of the rank of a student. In Appendix C.7,

[26] This result is confirmed by the finding that the SES rank has a negative effect on the probability of being classified as at risk for clinical depression (CES-D $\geq$ 16). Table C.3 in Appendix C shows that a 25 percentile increase in rank leads to a 4 percentage point lower likelihood of being classified as at risk for clinical depression.

Table 2: Average Effect of the Socioeconomic Rank

	(1)	(2)	(3)	(4)
Panel A: Depression				
CES-D	-1.09	-0.96	-0.96	-0.99
	(0.21)	(0.21)	(0.22)	(0.21)
Effect size	[-0.15]	[-0.13]	[-0.13]	[-0.13]
Number of observations	13,683	13,683	13,683	13,683
Panel B: Cognitive Ability				
Peabody	2.30	1.69	1.87	1.71
	(0.35)	(0.34)	(0.34)	(0.35)
Effect size	[0.16]	[0.12]	[0.13]	[0.12]
Number of observations	13,115	13,115	13,115	13,115
Panel C: Self-esteem				
6-item Rosenberg	0.49	0.52	0.56	0.57
	(0.13)	(0.13)	(0.13)	(0.13)
Effect size	[0.11]	[0.12]	[0.12]	[0.13]
Number of observations	11,917	11,917	11,917	11,917
Panel D: Popularity				
Bonacich	0.13	0.11	0.10	0.11
	(0.02)	(0.02)	(0.02)	(0.02)
Number of observations	12,883	12,883	12,883	12,883
Level of SES	yes	yes	yes	yes
Individual controls	no	yes	yes	yes
Cohort controls	no	no	yes	no
School and cohort FE	yes	yes	yes	no
School x cohort FE	no	no	no	yes

Note: Standard errors clustered at school level in parantheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) with the outcome variables: depression (panel A) cognitive ability (panel B), self-esteem (panel C), and popularity (panel D). The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Results with the original rank scale are presented in Appendix C.5. The effect size is calculated in terms of the standard deviation of the outcome variable. Column (1) includes separate school and cohort fixed effects and controls for the absolute level of SES. In column (2), individual controls (age in days, gender, and race) are added. In column (3), school cohort specific controls (mean and standard deviation of SES, fraction of repeaters, male share, and share of white students in the cohort) are additionally included. Column (4) controls for individual characteristics and school-by-cohort fixed effects. Wave I cross-sectional weights are used.

I characterize the aspects of school quality captured by the fixed effects by providing evidence on the correlation between the estimates and standard indicators of school quality.

4.2 Robustness

Strategic Delay of School Entry The central identifying assumption for a causal interpretation of the rank coefficient is the strict exogeneity condition. One potential concern regarding this condition is that parents may strategically delay their child's school entry, thereby imposing a potential violation of this assumption. In order to address this concern, I restrict my sample to students whose age is sufficiently close - within one standard deviation - to the average age in their school cohort. The argument here is that, for these students, strategic delays can be plausibly ruled out as a confounding factor. Based on the results presented in Appendix Table B.1, I conclude that strategic delay is not a threat to identification. Compared to the baseline, the estimates for depression and self-esteem are moderately larger while the results for cognitive ability and popularity are comparable to the baseline estimate.

Functional Form and SES-bins One concern for the identification of a causal rank effect could be misspecification in the regression model. Importantly, the plausibility of the strict exogeneity assumption always depends on the functional form assumption of my regression model in equation (4). This includes $f(SES)$, which defines the way in which I control for the level of SES in the model. Note again, that the choice of f is subject to a flexibility-precision trade-off in the sense that a more flexible choice restricts the rank variation that remains in order to estimate the rank coefficient. In the baseline estimation, I use four different SES-bins to capture SES-bin specific averages. Alternatively, one could think of different combinations to bin the SES categories or use a linear or quadratic function of SES. The results of these estimations can be found in Appendix Table B.2. For the main outcomes, depression and cognitive ability, as well as for popularity, the results remain robust to all alternative specifications of f. For self-esteem, the results remain robust when using alternative SES-bins, however the rank effect vanishes when using a linear or quadratic specification. I do not necessarily take this as evidence against a rank effect on self-esteem because misspecification could be a bigger issue in these alternative specifications. In fact, the baseline model with SES-bin specific averages allows for more flexibility and is thus less likely to suffer from misspecification than a linear or quadratic function of SES.

Breaking Ties When computing a student's rank within a cohort, one decision one has to make is how to break ties. In the main analysis, students are assigned the average rank in case of ties. Alternative ways to break ties include assigning students the lower rank, i.e. only counting students with a strictly lower socioeconomic status when ordering students, or to assign students the higher

rank, i.e. only counting students with a strictly higher socioeconomic status. In order to verify that the results are not driven by the way to break ties, I re-estimate the rank coefficient, constructing the rank variable according to each of the two alternatives. The results are presented in Appendix Table B.3. While the way to break ties has an impact on the size of the estimated regression coefficients, the results remain qualitatively robust to the alternative definitions.

Definition of Socioeconomic Status So far in this paper, the socioeconomic status of a student is measured as the educational attainment of the student's father.[27] Alternatively, I could define the socioeconomic status based on mothers' educational attainment, the highest level of educational attainment or the average educational attainment of both parents. In Appendix Table B.4, I compare these different definitions of socioeconomic status and find that the estimates are robust to the precise definition. Moreover, as discussed in Section 2.2, using the average educational attainment of both parents allows me to address and alleviate concerns regarding the high incidence of ties in the construction of the rank variable.

More generally, the socioeconomic status is a complex construct, determined by a combination of social and economic factors. Among the most common measures of SES is education, but also income. The reason I do not rely on income in the baseline construction of a student's socioeconomic status is twofold. First, income is only asked for in the in-home questionnaire, therefore only a fraction of each cohort received this question. Second, almost a quarter of the in-home sample has missing income information. Taken together, information on income is only available for slightly more than 15,000 students while parental education is available for nearly 80,000 students. While students participating in the in-home questionnaire were selected randomly, non-responses in the income variable are likely to be selective. Nonetheless, income might arguably be a more "visible" indicator for socioeconomic status, therefore I perform a robustness exercise in which I construct the rank measure based on family income (adjusted by family size). The results are reported in Appendix Table B.5 with qualitatively similar results as in the baseline for all outcomes but self-esteem. While the regression coefficients on the income rank are considerably smaller in size compared to the education rank, column (3) of Table B.5 demonstrates that a considerable part of this reduction in the coefficient size is likely due to the smaller and more selected sample. Overall the results remain qualitatively robust to the different definitions of a student's socioeconomic status.

Four-factor Model of Depression When originally developed, a factor analysis by Radloff (1977) showed that the CES-D can be divided into four subscales that represent different fac-

[27] Exceptions are made if the student reports not living with father or the father's information is missing. In this case, the mother's educational attainment is used.

tors, but are all symptoms related to depression. The four factors identified by Radloff (1977) have been confirmed in various studies, however, alternative factor structures have been proposed as well. Using principal components analysis with varimax rotation in the Add Health data, I find 4 factors with eigenvalues greater than one that account for 51% of the variance.[28] Compared to Radloff (1977), some items loaded differently on the four factors. Table A.2 shows the rotated factor loadings of all items. Including items with factor loadings above 0.40 identifies 4 factors with similar interpretation to Radloff (1977):

I. Depressed affect: *bothered, appetite, blues, depressed, failure, fearful, lonely, sad, worth living*

II. Positive affect: *good, hopeful, happy, enjoyed life*

III. Somatic symptoms: *mind, tired, get started*

IV. Interpersonal problems: *unfriendly, disliked*

Each of the 4 subscales' score is computed as the sum of the items and divided by the number of items to facilitate the comparison between the subscales. Regression results of equation (4) with the four subscales of depression as outcome variables are presented in Table B.6. Strikingly, the socioeconomic rank of a student in their high school cohort has an impact on all four factors of the CES-D. Moreover, the effect size seems to be comparable across factors, though slightly larger for positive affect.[29] This confirms the main results and demonstrates that the rank effect on depression is not driven by a single factor or item in the depression score.

4.3 Heterogeneous Effects

In this section, I study potential heterogeneities in the effect of the socioeconomic rank along multiple dimensions. First, I explore whether the degree of inequality within a school cohort impacts the magnitude to which the socioeconomic rank affects depression, cognitive ability, self-esteem, and popularity during high school. I then proceed to study heterogeneities along the individual level, including gender and race.

Exploring the Role of Inequality By construction, the measure of socioeconomic rank estimates a student's relative position, but ignores any notion of distance between peers. However, the distance between two rank positions may matter for the degree to which the socioeconomic

[28] The corresponding scree plot can be found in Appendix Figure A.1.

[29] As indicated in Table A.1, items for positive effects were reverse coded such that - similar to the other items - higher values indicate worse conditions.

rank affects adolescent development. In line with relative deprivation theory, higher degrees of inequality likely lead to larger differences between the desired situation and one's own, thus elicit higher degrees of envy, shame and humiliation and could intensify competition among peers. In this section, I therefore study the extent to which inequality within a student's comparison group, i.e. the school cohort, affects the socioeconomic rank gradient.

Inequality is measured using the standard deviation of the SES-distribution within a school cohort. All school cohorts are then ordered according to the magnitude of this standard deviation and divided into quintiles. Cohorts with the lowest degree of inequality are assigned to the first quintile and cohorts with the highest degree of inequality are assigned to the 5th quintile.

To test for the role of inequality in the relationship between the socioeconomic rank and students' contemporaneous outcomes, I estimate equation (4), interacting the rank with indicators for each inequality quintile. Table 3 depicts a clear pattern in the estimated coefficients on the socioeconomic rank for the different quintiles: Irrespective of the level of inequality within a cohort, the relationship between a student's socioeconomic rank and all four contemporaneous outcomes holds. However, the estimated rank coefficients increase in absolute size with the degree of inequality within a cohort. Holding fixed the level of socioeconomic status, higher ranked students gain more compared to lower ranked students when inequality is high in their cohort. This pattern is quite striking in its consistency across outcomes.

The observed pattern is reassuring as it confirms my main results and is consistent with theories of relative deprivation. From an equity perspective, these results can be viewed as a motivation for policy interventions aimed at reducing the salience of inequality as this could mitigate the adverse effects of relative deprivation.

Heterogeneity Along the Rank Distribution One natural question that arises is whether the observed rank effects exist along the complete rank distribution or whether they only materialize for lower-ranked students. To address this question, I construct an indicator variable, $\mathbb{1}(Rank_{isc} > 0.5)$, that takes on the value 1 if the student is ranked above the median in their school cohort and 0 otherwise. I estimate equation (4), interacting the rank variable with this indicator variable and report the resulting coefficient estimates and 95% confidence intervals for students below and above the median rank in their cohorts in Figure D.1. The rank effects exists for both students below and above the median rank in their cohort. However, the pattern suggests a heterogeneous effect along the rank distribution since the rank effects are more pronounced for students with ranks below the median.

Other Heterogeneities In a next step, I study potential effect heterogeneities along individual characteristics. Specifically, I look at differences in the rank effect by gender and race. Figure D.2

Table 3: Heterogeneous Effect of the SES Rank by the Degree of Inequality

Inequality quintile	1st	2nd	3rd	4th	5th
Depression	-0.80	-0.91	-1.07	-1.03	-1.19
	(0.22)	(0.22)	(0.26)	(0.27)	(0.31)
Cognitive ability	1.31	1.73	1.93	2.02	3.30
	(0.38)	(0.34)	(0.36)	(0.36)	(0.49)
Self-esteem	0.38	0.54	0.68	0.69	0.65
	(0.14)	(0.15)	(0.14)	(0.14)	(0.16)
Popularity	0.09	0.09	0.11	0.12	0.12
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Note: Standard errors clustered at school level in parantheses; Inequality quintiles group school cohorts into quintiles based on the standard deviation of the school cohort-level SES distribution. The table reports the estimated rank coefficients when interacting the socioeconomic rank with indicators of these inequality quintiles in equation (4) for each outcome: depression, cognitive ability, self-esteem, and popularity. The model specification includes separate school and cohort fixed effects and controls for school cohort specific controls (mean and standard deviation of SES, fraction of repeaters, male share, and share of white students in the cohort), the level of SES, and individual controls (age in days, gender, and race). The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Wave I cross-sectional weights are used.

in Appendix D depicts the estimated rank effects that result from interacting the rank in equation (4) with gender or race dummies. For gender, the results show that, along all outcomes, both boys and girls are affected by their socioeconomic rank position. If anything, girls tend to react slightly stronger to their rank position, however, the depicted differences are not necessarily statistically significant. For race, I first distinguish white students from students with any other racial background. The results show that differences between the two groups are not necessarily statistically significant, but the estimated rank coefficients are systematically stronger for white students. A more detailed split by race shows, however, that the coarse classification into white and non-white students masks substantial heterogeneities across the races and outcomes.

4.4 Persistence of Effects

A natural question that arises in the context of the observed socioeconomic rank effects is whether these effects are persistent. The importance of mental health as well as cognitive, non-cognitive and social skills for human capital development would suggest that the relative socioeconomic status has long-term consequences for economic success and well-being. I therefore investigate the long-term effects of socioeconomic rank on depression and educational attainment during adulthood, that is when respondents are between 24 and 32 years old. To this end, I estimate equation (4), using wave IV outcome measures for the 10-items CES-D score and dummies for college attendance and college completion as dependent variables.

The results of the different model specifications are presented in Table 4. Similarly to before, the reported coefficients represent the effect of an increase in rank by 25 percentiles (one standard

deviation). Overall, the estimated coefficients are very stable across the different model specifications, signaling that a higher within-cohort rank during high school is associated with significantly lower depression scores and better odds at attending and completing college. The coefficient estimates in column (3) and (4) imply that a 25 percentile increase in the socioeconomic rank during high school reduces depression scores by 0.45 points. This is equivalent to a reduction by 0.12 standard deviations, an effect size similar to the one reported on short-run depression. This finding is consistent with evidence documenting the persistence of mental health problems. Further, a 25 percentile increase in the socioeconomic rank is associated with a higher likelihood of attending and completing college by 5 and 4 percentage points, respectively.

Similar as before, I use the estimated school fixed effects from equation (4) to compare the rank effect to the effect of school quality on the outcomes. A one-standard deviation increase in school quality is associated with a decrease of 0.8 points in long-term depression and a 11 and 16 percentage points higher likelihood of attending and completing college, respectively. With respect to college attendance and completion, this implies that increasing the socioeconomic rank by 25 percentiles, holding constant school quality, is equivalent to increasing the school quality by 0.5 and 0.2 standard deviations, holding constant the rank of a student. In regards to mental health, such a rank increase is equivalent to an increase in school quality by 0.6 standard deviations.

Table 4: Persistence of the Rank Effect

	(1)	(2)	(3)	(4)
Panel A: Long-run Depression				
CES-D (10 items)	-0.43	-0.42	-0.45	-0.45
	(0.15)	(0.15)	(0.15)	(0.15)
Effect size	[-0.11]	[-0.11]	[-0.12]	[-0.12]
Number of observations	10,901	10,901	10,901	10,901
Panel B: College				
Attending college	0.06	0.04	0.05	0.05
	(0.01)	(0.01)	(0.01)	(0.01)
Completing college	0.05	0.04	0.04	0.04
	(0.01)	(0.01)	(0.01)	(0.01)
Number of observations	10,911	10,911	10,911	10,911
Level of SES	yes	yes	yes	yes
Individual controls	no	yes	yes	yes
Cohort controls	no	no	yes	no
School and cohort FE	yes	yes	yes	no
School x cohort FE	no	no	no	yes

Note: Standard errors clustered at school level in parantheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) for the long-run outcomes: the 10-item CES-D (panel A) and dummies for college completion and college attendance (panel B). The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Results with the original rank scale are presented in Appendix C.5. The effect size is calculated in terms of the standard deviation of the outcome variable. Column (1) includes separate school and cohort fixed effects and controls for the absolute level of SES. In column (2), individual controls (age in days, gender, and race) are added. In column (3), school cohort specific controls (mean and standard deviation of SES in the cohort, fraction of repeaters, male share, share of white students in the cohort) are additionally included. Column (4) controls for individual characteristics and school-by-cohort fixed effects. Wave IV cross-sectional weights are used.

4.5 Mediation Analysis

In light of the persistent effects of a student's socioeconomic rank within their high school cohort on mental health and educational attainment, an interesting question to ask is to what extent these long-run effects are mediated by the observed short-run effects. Specifically, I am interested in the importance of adolescent mental health as a mediator. Since the available data does not allow for an appropriate causal mediation analysis, which at the minimum would require some type of sequential ignorability assumption (see e.g. Imai et al., 2011), which is almost certainly violated in the present context, it is beyond the scope of this paper to provide a full answer to this question.

However, my results allow me to conduct back-of-the-envelope calculations that are suggestive of the relative importance of depression for long-run educational attainment. The goal is to split the "total" effect of the rank on long-run outcomes into a "direct" effect and an "indirect" effect. The "indirect" effect refers to the effect of rank on long-run depression and educational

attainment that operates through mediators. The mediators of interest, m_{isc}, are adolescent depression, dep_{isc}, cognitive ability, cog_{isc}, self-esteem, $self_{isc}$, and popularity, pop_{isc}. To this end, I estimate a set of equations, regressing each of the long-run outcomes, y_{isc}, and each of the mediators on the socioeconomic rank using equation (4). Moreover, I estimate an auxiliary regression in which the potential mediators are added as regressors when estimating equation (4) for the long-run outcomes. For example, in the case of depression as mediator, I estimate the following set of equations:

$$y_{isc} = \alpha_1 + \beta_1 \text{Rank}_{isc} + f(\text{SES}_{isc}) + \mathbf{X}_{isc}\gamma_1 + g(s,c) + u_{1,isc}$$
$$m_{isc} = \alpha_2 + \beta_2 \text{Rank}_{isc} + f(\text{SES}_{isc}) + \mathbf{X}_{isc}\gamma_2 + g(s,c) + u_{2,isc}$$

with $m_{isc} = dep_{isc}$.

Finally, I estimate the auxiliary regression in which all potential mediators are added as regressors:

$$y_{isc} = \alpha_3 + \beta_3 \text{Rank}_{isc} + \beta_d \text{dep}_{isc} + \beta_c \text{cogn}_{isc} + \beta_s \text{self}_{isc} + \beta_p \text{pop}_{isc} + f(\text{SES}_{isc}) + \mathbf{X}_{isc}\gamma_3 + g(s,c) + u_{3,isc}.$$

For each outcome, the effect mediated through depression is then defined as the product of $\beta_2\beta_d$. Dividing this product by the total rank effect β_1 yields the share of the socioeconomic rank effect mediated by depression. Analogously, the shares mediated through cognitive ability, self-esteem, and popularity are estimated.

Table 5 presents the results from this exercise. It reports the total socioeconomic rank effect from regressing the long-run outcomes on the socioeconomic rank according to equation (4) in column (1).[30] In column (2), all mediators are added as regressors. This reduces the rank coefficient considerably for all long-term outcomes. The right-hand side of Table 5 displays the computed shares of the total effect that are mediated by adolescent depression, cognitive ability, self-esteem, and popularity measured in high school for each of the three long-term outcomes. Unsurprisingly, adolescent depression is the most important mediator for the relationship between socioeconomic rank and the depression score in adulthood, accounting for more than 25% of the total rank effect. In combination with self-esteem, a mediator closely connected to mental health, almost 35% of the total rank effect are mediated by these two factors. Cognitive ability and popularity are only weak mediators. In comparison, cognitive ability in high school is the most important factor that mediates the effect of socioeconomic rank on college attainment, both in terms of college attendance

[30] The estimated regression coefficients deviate slightly from the reported coefficients in Table 4 because the sample was reduced to individuals with complete information on all mediators. Otherwise, the specifications in column (1) of Table 5 and column (3) of Table 4 are identical.

Table 5: Mediation Analysis

	(1)	(2)	% of total effect mediated			
			Depression	Cognitive ability	Self-esteem	Popularity
Panel A: Long-run Depression						
CES-D (10 items)	-0.48	-0.31	25.90	0.48	8.53	2.25
	(0.18)	(0.16)				
Number of observations	8,519	8,519				
Panel B: College						
Attending college	0.04	0.03	8.37	24.08	4.72	11.10
	(0.02)	(0.02)				
Completing college	0.04	0.02	10.47	27.31	6.18	15.51
	(0.01)	(0.01)				
Number of observations	8,526	8,526				
Level of SES	yes	yes				
Individual controls	yes	yes				
Cohort controls	yes	yes				
School and cohort FE	yes	yes				
Mediators	no	yes				

Note: Standard errors clustered at school level in parantheses; The table reports rank coefficients from a regression of long-run outcomes on the socioeconomic rank according to equation (4) with separate school and cohort fixed effects as well as controls for the level of SES, individual (age in days, gender, and race) as well as school cohort specific (mean and standard deviation of SES in the cohort, fraction of repeaters, male share, share of white students) controls. The sample is reduced to individuals with complete information on all mediators. Column (1) replicates the results from column (3) of Table 4 with the reduced sample size. Column (2) reports the rank coefficients from auxilliary regressions that add all potential mediators (depression, cognitive ability, self-esteem and popularity during high school) as regressors. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Columns (3)-(6) display the share of the rank effect that is mediated by: depression, cognitive ability self-esteem, and popularity during high school. Wave IV cross-sectional weights are used.

and college completion. Roughly 25-30% of the rank effect on educational attainment is mediated through this factor. However, depression, self-esteem, and popularity also seem to be important channels through which the rank effect impacts educational attainment. Taken together, these three factors are equally important for educational attainment, compared to cognitive ability.

5 Conclusion

Motivated by the importance of mental health for adolescents' unimpeded development, this paper provides new causal evidence on the effect of relative parental socioeconomic status on adolescents' mental health, cognitive ability and educational attainment. I show that the relative socioeconomic status has a significant and economically meaningful impact on adolescents' personal development that persists into adulthood.

The short-run effects documented in my analysis demonstrate that socioeconomic ranks impact teenagers' development along several important and interrelated dimensions. In particular, I find that higher ranks lead to reductions in depression scores, improved cognitive ability and self-esteem as well as higher levels of social integration as measured by popularity.

While my data does not allow me to pin down the specific mechanisms underlying the causal rank effects, the patterns I document are consistent with theories of social comparisons and relative deprivation widely accepted in the sociology and social psychology literature. I document that the estimated rank effects are more pronounced in cohorts with higher levels of SES-inequality across all considered outcome dimensions, suggesting that social comparisons have non-negligible impacts on adolescents' mental health and behavior.

Strikingly, the rank effects on depression persist into adulthood with effect sizes almost identical to those documented in the short-run. My findings are consistent with evidence documenting high levels of persistence of mental health disorders, highlighting the importance of mental health and interventions designed to reduce risks to mental health during adolescence. I also find substantial long-run effects on educational attainment as measured by college attendance and completion.

An important question that arises in this context is to what extent mental health problems impede the accumulation of human capital. While the data available does not allow me to conduct an appropriate mediation analysis, I provide suggestive evidence that depression does in fact impede human capital development as measured by educational attainment. As a complete assessment of the economic costs of mental health disorders of teenagers requires quantifying this link, future research on this question is needed.

The results documented in this paper can be viewed as motivation and justification for policies aimed at reducing the salience of inequality in schools. From an equity perspective, such policies could be an effective tool to mitigate the adverse mental health and human capital consequences of relative deprivation and thus enhance educational outcomes and intergenerational mobility. Importantly, the potential gains of such policies for lower-ranked students outweigh the potential losses of higher-ranked students because the rank effects are stronger for students ranked below the median. Concrete efforts of this type could entail the provision of paid-for school meals, school uniforms and subsidized leisure activities. Alternatively, policies targeting individuals based on parental background (e.g. mentoring programs) should not only do so on the basis of absolute SES, but also consider relative SES in order to counteract the negative rank effects documented in this paper.

References

Arduini, T., D. Iorio, and E. Patacchini (2019). "Weight, Reference Points, and the Onset of Eating Disorders". *Journal of Health Economics 65*, 170–188.

Balsa, A. I., M. T. French, and T. L. Regan (2014). "Relative Deprivation and Risky Behaviors". *The Journal of Human Resources 49* (2), 446–471.

Beekman, A. T. F., D. J. H. Deeg, T. van Tilburg, J. H. Smit, C. Hooijer, and W. van Tilburg (1995). "Major and Minor Depression in Later Life: a Study of Prevalence and Risk Factors". *Journal of Affective Disorders 36* (1-2), 65–75.

Bifulco, R., J. M. Fletcher, and S. L. Ross (2011). "The Effect of Classmate Characteristics on Post-secondary Outcomes: Evidence from the Add Health". *American Economic Journal: Economic Policy 3* (1), 25–53.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital". *American Economic Review 95* (1), 437–449.

Braghieri, L., R. Levy, and A. Makarin (2022). "Social Media and Mental Health". *American Economic Review 112* (11), 3660–93.

Braudt, D. B. and K. M. Harris (2020). "Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health) – Release 2". *Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill*.

Brown, G. D. A., J Gardiner, A. J. Oswald, and J Qian (2008). "Does Wage Rank Affect Employees' Well-being?" *Industrial Relations: A Journal of Economy and Society 47* (3), 355–389.

Butikofer, A., R. Ginja, F. Landaud, and K. V. Løken (2020). "School Selectivity, Peers, and Mental Health". *NHH Dept. of Economics Discussion Paper* (21).

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan (2014). "Early Childhood Investments Substantially Boost Adult Health". *Science 343* (6178), 1478–1485.

Card, D., A. Mas, E. Moretti, and E. Saez (2012). "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction". *American Economic Review 102* (6), 2981–3003.

Carneiro, P. M. and J. J. Heckman (2003). "Human Capital Policy". *Inequality in America: What Role for Human Capital Policies,* ed. by J. J. Heckman and A. Krueger. MIT Press.

Carrell, S. E., M. Hoekstra, and E. Kuka (2018). "The Long-Run Effects of Disruptive Peers". *American Economic Review 108* (11), 3377–3415.

Chen, P. and K. Chantala (2014). "Guidelines for Analyzing Add Health Data". *Carolina Population Center, University of North Carolina at Chapel Hill 710*.

Clark, A. E. and A. J. Oswald (1996). “Satisfaction and Comparison Income”. *Journal of Public Economics 61* (3), 359–381.

Conti, G., A. Galeotti, G. Müller, and S. Pudney (2013). “Popularity”. *Journal of Human Resources 48* (4), 1072–1094.

Crosby, F. (1976). “A Model of Egoistical Relative Deprivation.” *Psychological review 83* (2), 85.

Cunha, F. and J. Heckman (2007). “The Technology of Skill Formation”. *American Economic Review 97* (2), 31–47.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). “Interpreting the Evidence on Life Cycle Skill Formation”. *Handbook of the Economics of Education 1*, 697–812.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). “Estimating the Technology of Cognitive and Noncognitive Skill Formation”. *Econometrica 78* (3), 883–931.

Currie, J. (2009). “Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development”. *Journal of Economic Literature 47* (1), 87–122.

Currie, J. and J. Goodman (2020). “Parental Socioeconomic Status, Child Health, and Human Capital”. *The Economics of Education*. Elsevier, 239–248.

Currie, J. and M. Stabile (2006). “Child Mental Health and Human Capital Accumulation: the Case of ADHD”. *Journal of Health Economics 25* (6), 1094–1118.

——— (2009). “4. Mental Health in Childhood and Human Capital”. *The Problems of Disadvantaged Youth: An Economic Perspective*. Ed. by J. Gruber. University of Chicago Press, 115–148.

Currie, J., M. Stabile, P. Manivong, and L. L. Roos (2010). “Child Health and Young Adult Outcomes”. *Journal of Human Resources 45* (3), 517–548.

Dahl, G. B. and L. Lochner (2012). “The Impact of Family Income on Child Achievement: Evidence From the Earned Income Tax Ccredit”. *American Economic Review 102* (5), 1927–1956.

Dahl, R. E., N. B. Allen, L. Wilbrecht, and A. B. Suleiman (2018). “Importance of Investing in Adolescence From a Developmental Science Perspective”. *Nature 554*, 441–450.

Delaney, J. and P. J. Devereux (2019). “The Effect of High School Rank in English and Math on College Major Choice”. *CEPR Discussion Paper No. DP14205*.

Demontis, D. et al. (2017). “Discovery of the First Genome-wide Significant Risk Loci for ADHD”. *bioRxiv*.

Denning, J. T., R. Murphy, and F. Weinhardt (2021). “Class Rank and Long-Run Outcomes”. *The review of Economics and Statistics*, 1–45.

Dunn, L. M. and D. M. Dunn (2007). *The Peabody Picture Vocabulary Test*. 4th. Bloomington: MN: NCS Pearson.

Elsner, B. and I. E. Isphording (2017). “A Big Fish in a Small Pond: Ability Rank and Human Capital Investment”. *Journal of Labor Economics 35* (3), 787–828.

——— (2018). “Rank, Sex, Drugs, and Crime”. *Journal of Human Resources 53* (2), 356–381.

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

Festinger, L (1954). “A Theory of Social Comparison Processes”. *Human Relations 7* (2), 117–140.

Harris, K. M. (2018). “The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994-1996; Wave III, 2001-2002; Wave IV, 2007-2009; Wave V, 2016-2018 [machine-readable data file and documentation]”. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

Harris, K. M., C. T. Halpern, E. A. Whitsel, J. M. Hussey, L. A. Killeya-Jones, J. Tabor, and S. C. Dean (2019). “Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health)”. *International Journal of Epidemiology 48* (5), 1415–1415k.

Holmlund, H., M. Lindahl, and E. Plug (2011). “The Causal Effect of Parents’ Schooling on Children’s Schooling: A Comparison of Estimation Methods”. *Journal of economic literature 49* (3), 615–51.

Hoxby, C. M. (2000). “The Effects of Class Size on Student Achievement: New Evidence from Population Variation”. *The Quarterly Journal of Economics 115* (4), 1239–1285.

Imai, K., L. Keele, D. Tingley, and T. Yamamoto (2011). “Unpacking the Black Box of Causality: Learning About Causal Mechanisms From Experimental and Observational Studies”. *American Political Science Review 105* (4), 765–789.

Kiessling, L. and J. Norris (2022). “The Long-Run Effects of Peers on Mental Health”. *The Economic Journal*, 1–42.

Lee, J. J. et al. (2018). “Gene Discovery and Polygenic Prediction from a 1.1-Million-Person GWAS of Educational Attainment”. *Nature Genetics 50* (8), 1112.

Lewinsohn, P. M., J. R. Seeley, R. E. Roberts, and N. B. Allen (1997). “Center for Epidemiologic Studies Depression Scale (CES-D) as a Screening Instrument for Depression Among Community-Residing Older Adults”. *Psychology and Aging 12* (2), 277–287.

Lundborg, P., A. Nilsson, and D.-O. Rooth (2014). “Parental Education and Offspring Outcomes: Evidence From the Swedish Compulsory School Reform”. *American Economic Journal: Applied Economics 6* (1), 253–278.

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

Milligan, K. and M. Stabile (2011). "Do Child Tax Benefits Affect the Well-Being of Children? Evidence From Canadian Child Benefit Expansions". *American Economic Journal: Economic Policy 3* (3), 175–205.

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

Neale, B. M. et al. (2010). "Meta-analysis of Genome-wide Association Studies of Attention-deficit/Hyperactivity Disorder". *Journal of the American Academy of Child & Adolescent Psychiatry 49* (9), 884–897.

OECD (2015). *Fit Mind, Fit Job: From Evidence to Practice in Mental Health and Work*. Paris: Mental Health and Work, OECD Publishing.

Okbay, A. et al. (2016). "Genome-wide Association Study Identifies 74 Loci Associated with Educational Attainment". *Nature 533* (7604), 539–542.

Oreopoulos, P., M. E. Page, and A. H. Stevens (2006). "The Intergenerational Effects of Compulsory Schooling". *Journal of Labor Economics 24* (4), 729–760.

Pagani, L., S. Comi, and F. Origo (2019). "The Effect of School Rank on Personality Traits". *Journal of Human Resources*.

Psychiatric Genomics Consortium, C.-D. G. of the et al. (2013). "Identification of Risk Loci With Shared Effects on Five Major Psychiatric Disorders: a Genome-wide Analysis". *The Lancet 381* (9875), 1371–1379.

Psychiatric Genomics Consortium, S. W. G. of the (2014). "Biological Insights From 108 Schizophrenia-associated Genetic Loci". *Nature 511* (7510), 421–427.

Psychiatric GWAS Consortium, M. D. D. W. G. of the et al. (2013). "A Mega-analysis of Genome-wide Association Studies for Major Depressive Disorder". *Molecular Psychiatry 18* (4).

Psychiatric GWAS Consortium Bipolar Disorder Working Group (2011). "Large-scale Genome-wide Association Analysis of Bipolar Disorder Identifies a New Susceptibility Locus Near ODZ4". *Nature Genetics 43* (10), 977–983.

Radloff, L. S. (1977). "The CES-D Scale: A Self Report Depression Scale for Research in the General Population". *Applied Psychological Measurements 1* (3), 385–401.

——— (1991). "The Use of the Center for Epidemiologic Studies Depression Scale in Adolescents and Young Adults". *Journal of Youth and Adolescence 20* (2), 149–166.

Roberts, R. E., J. A. Andrews, P. M. Lewinsohn, and H. Hops (1990). "Assessment of depression in adolescents using the Center for Epidemiologic Studies Depression Scale." *Psychological Assessment: A Journal of Consulting and Clinical Psychology 2* (2), 122.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

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*. Ed. by E. A. Hanushek, S Machin, and L Woessmann. Vol. 3. Elsevier, 249–277.

Smith, H. J., T. F. Pettigrew, G. M. Pippin, and S. Bialosiewicz (2012). “Relative Deprivation: A Theoretical and Meta-Analytic Review”. *Personality and Social Psychology Review 16* (3), 203–232.

Stouffer, S. A., E. A. Suchman, L. C. DeVinney, S. A. Star, and R. M. Williams Jr (1949). “The American Soldier: Adjustment During Army Life. (Studies in Social Psychology in World War II), Vol. 1”.

Suls, J. M. and R. L. Miller (1977). *Social Comparison Processes: Theoretical and Empirical Perspectives*. Hemisphere.

Wills, T. A. (1981). “Downward Comparison Principles in Social Psychology”. *Psychological Bulletin 90* (2), 245–271.

Wood, J. V. (1989). “Theory and Research Concerning Social Comparisons of Personal Attributes”. *Psychological Bulletin 106* (2), 231–248.

Wray, N. R. et al. (2018). “Genome-wide Association Analyses Identify 44 Risk Variants and Refine the Genetic Architecture of Major Depression”. *Nature Genetics 50* (5), 668–681.

Online Appendix

Adolescents' Mental Health and Human Capital: The Role of Socioeconomic Rank

Michaela Paffenholz

A Appendix: Outcome Measures

A.1 The CES-D

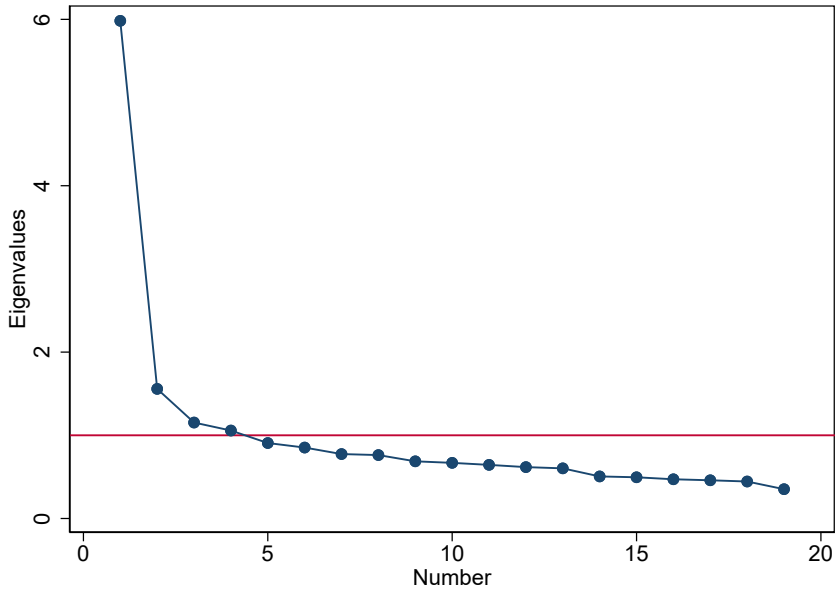
The Center of Epidemiologic Studies-Depression (CES-D) asks about the frequency with which an individual experienced symptoms associated with depression in the last week. The response options range from 0 to 3 for each item (0 = Never or rarely, 1 = Sometimes, 2 = A lot of the time, 3 = Most of the time or all of the time). Positively worded items were reverse coded. The CES-D is constructed as the sum of all items and ranges from 0 - 57 with higher scores indicating a higher degree of depressive symptoms. A score equal to or above 16 is commonly referred to as a cutoff for being at risk for clinical depression.

Table A.1: The CES-D

Measure	Item	Scale
CES-D	You were bothered by things that don't usually bother you.*	Never 0 – 3 most/all of the time
	You didn't feel like eating, your appetite was poor.	
	You felt that you could not shake off the blues, even with help from your family and your friends.*	
	You felt you were just as good as other people. (reverse coded)*	
	You had trouble keeping your mind on what you were doing.*	
	You felt depressed.*	
	You felt that you were too tired to do things.*	
	You felt hopeful about the future. (reverse coded)	
	You thought your life had been a failure.	
	You felt fearful.	
	You were happy. (reverse coded)*	
	You talked less than usual.	
	You felt lonely.	
	People were unfriendly to you.	
	You enjoyed life. (reverse coded)*	
	You felt sad.*	
	You felt that people disliked you.*	
	It was hard to get started doing things.	
	You felt life was not worth living.	

Note: This table displays the items in wave I of the Add Health data set that were used to construct the outcome variable depression (CES-D). Positively worded questions were reverse coded. The final CES-D score was computed as the sum of all items. Items marked with an asterisk (*) indicate questions that were also asked during the wave IV interview and were used to construct the CES-D 10-item measure of depression in the long-run.

Figure A.1: Screeplot of a Principal Component Analysis of the CES-D



Note: This figure presents a screeplot of principal components, using all items of the CES-D in wave I. It identifies four factors with eigenvalues larger than 1.

Table A.2: Factor Loadings of CES-D Items

	Depressed Affect	Positive Affect	Somatic Symptoms	Interpersonal Problems
You were bothered by things that usually don't bother you.	**0.54**	0.09	0.33	0.07
You didn't feel like eating, your appetite was poor.	**0.46**	0.09	0.35	-0.09
You could not shake the blues, even with help from your friends and family.	**0.73**	0.14	0.21	0.05
You felt that you were just as good as other people.	0.09	**0.68**	0.04	0.12
You had trouble keeping your mind on what you were doing.	0.30	0.10	**0.60**	0.11
You felt depressed.	**0.76**	0.18	0.20	0.13
You felt that you were too tired to do things.	0.20	0.11	**0.69**	0.14
You felt hopeful about the future.	0.00	**0.76**	0.09	0.01
You thought your life had been a failure.	**0.57**	0.22	-0.01	0.33
You felt fearful.	**0.46**	0.02	0.16	0.30
You were happy.	0.32	**0.68**	0.09	0.05
You talked less than usual.	0.29	0.12	0.32	0.10
You felt lonely.	**0.65**	0.12	0.16	0.22
People were unfriendly to you.	0.09	0.02	0.15	**0.81**
You enjoyed life.	0.31	**0.68**	0.07	0.11
You felt sad.	**0.70**	0.14	0.16	0.21
You felt that people disliked you.	0.24	0.13	0.11	**0.78**
It was hard to get started doing things.	0.12	0.08	**0.70**	0.22
You felt life was not worth living.	**0.54**	0.20	-0.06	0.32

Note: This table reports factor loadings of each item in the CES-D for the four principal component factors. Bold items with loadings larger than 0.4 are assigned to the fours factors: depressed affect, positive affect, somatic symptoms, and interpersonal problems.

A.2 Self-Esteem Scale

To measure self-esteem, six items similar to or adapted from the Rosenberg self-esteem scale (Rosenberg, 1965) were used. Students were asked how much they agreed or disagreed on a 5-point Likert scale with the statements presented in the table below. The final score is computed as the sum of all items and ranges from 0 - 24 with higher values indicating higher self-esteem.

Table A.3: The Adapted Rosenberg Self-Esteem Scale

Measure	Do you agree or disagree that you...	Scale
Rosenberg Self-Esteem	have many good qualities	Strongly disagree 0 - 4 Strongly agree
	have a lot to be proud of	
	like yourself just the way you are	
	feel you are doing things just about right	
	feel socially accepted	
	feel loved and wanted	

Note: This table displays the items in wave I of the Add Health data set that were used to construct the self-esteem score. Items are originally rated on a scale from 1 (Strongly agree) to 5 (Strongly disagree) and were reverse coded and scaled to range from 0-4. The overall score of self-esteem is computed as the sum of all items.

B Appendix: Robustness Checks

Table B.1: Test for Strategic Delay

	Depression	Cognitive ability	Self-esteem	Popularity
Rank Coefficient	-1.20	1.86	0.68	0.10
	(0.27)	(0.40)	(0.17)	(0.03)
Number of observations	9,733	9,358	8,518	9,183

Note: Standard errors clustered at school level in parantheses; The table reports the estimated rank coefficients from estimating equation (4) with the level of SES, all individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) controls as well as separate school and cohort fixed effects. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Each column refers to a different outcome. The sample is restricted to individuals within 1 standard deviation of the average age level in the school cohort. Wave I cross-sectional weights are used.

Table B.2: Alternative SES-Bins and Functional Form

	4 SES bins (Baseline)	3 SES bins	linear SES	quadratic SES
Panel A: Depression				
CES-D	-0.96	-0.75	-0.82	-0.97
	(0.22)	(0.18)	(0.26)	(0.27)
Number of observations	13,683	13,683	13,683	13,683
Panel B: Cognitive Ability				
Peabody	1.87	1.92	1.07	1.29
	(0.34)	(0.31)	(0.41)	(0.42)
Number of observations	13,115	13,115	13,115	13,115
Panel C: Self-esteem				
6-item Rosenberg	0.56	0.52	0.06	0.11
	(0.13)	(0.12)	(0.15)	(0.15)
Number of observations	11,917	11,917	11,917	11,917
Panel D: Popularity				
Bonacich	0.10	0.09	0.06	0.08
	(0.02)	(0.02)	(0.03)	(0.03)
Number of observations	12,883	12,883	12,883	12,883

Note: Standard errors clustered at school level in parantheses; The table reports the estimated rank coefficients from estimating equation (4) with the level of SES, all individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) controls as well as separate school and cohort fixed effects. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Each column refers to a different specification of $f()$ in equation (4). In column (1), SES is controlled for through 4 SES-bins ("high school or less", "some college" "college", and "postgraduate"). In column (2), SES is controlled for through 3 SES-bins ("high school or less", "some college" "at least college"). In column (3) and (4), linear and quadratic functions of the SES variable are used, respectively. Wave I cross-sectional weights are used.

Table B.3: Alternative Ways to Break Ties

	Average (Baseline)	Lower	Higher
Panel A: Depression			
CES-D	-0.96	-1.20	-0.62
	(0.22)	(0.26)	(0.16)
Number of observations	13,683	13,683	13,683
Panel B: Cognitive Ability			
Peabody	1.87	2.01	1.32
	(0.34)	(0.44)	(0.25)
Number of observations	13,115	13,115	13,115
Panel C: Self-esteem			
6-item Rosenberg	0.56	0.47	0.43
	(0.13)	(0.16)	(0.10)
Number of observations	11,917	11,917	11,917
Panel D: Popularity			
Bonacich	0.10	0.10	0.08
	(0.02)	(0.03)	(0.02)
Number of observations	12,883	12,883	12,883

Note: Standard errors clustered at school level in parantheses; The table reports the estimated rank coefficients from estimating equation (4) with the level of SES, all individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) controls as well as separate school and cohort fixed effects. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Different methods to calculate the rank, in particular different rules to break ties for students with the same socioeconomic status, are used. The 'Average' rank coincides with the baseline estimate; ties are assigned the average rank of the tied positions. The 'Lower' rank is computed counting the number of individuals with a strictly lower socioeconomic status. In contrast, the 'Higher' rank assigns the rank based on the number of individuals with a strictly higher socioeconomic status. Wave I cross-sectional weights are used.

Table B.4: Alternative Definitions of SES

	Father (Baseline)	Mother	Highest Education	Average Eduation
Panel A: Depression				
CES-D	-0.96	-0.83	-0.93	-0.80
	(0.22)	(0.22)	(0.26)	(0.20)
Number of observations	13,683	13,683	13,683	13,683
Panel B: Cognitive Ability				
Peabody	1.87	2.16	2.31	1.69
	(0.34)	(0.34)	(0.42)	(0.31)
Number of observations	13,115	13,115	13,115	13,115
Panel C: Self-esteem				
6-item Rosenberg	0.56	0.40	0.35	0.53
	(0.13)	(0.14)	(0.16)	(0.13)
Number of observations	11,917	11,917	11,917	11,917
Panel D: Popularity				
Bonacich	0.10	0.11	0.08	0.12
	(0.02)	(0.03)	(0.03)	(0.03)
Number of observations	12,883	12,883	12,883	12,883
Number of observations	13,580	13,580	13,580	13,580

Note: Standard errors clustered at school level in parantheses; The table reports the rank coefficients from estimating equation (4) with the level of SES, all individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) controls as well as separate school and cohort fixed effects. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Each column refers to a different definition on how to define a students' socioeconomic status. SES is defined as the father's educational attainment in column (1), the mother's educational attainment in column (2), the highest educational attainment of both parents in column (3), and the average parental education in column (4). Wave I cross-sectional weights are used.

Table B.5: Alternative Definition of SES - Income

	Father's Education (Baseline)	Family Income	Father's Education (Adjusted Sample)
Panel A: Depression			
CES-D	-0.96	-0.34	-0.54
	(0.22)	(0.14)	(0.23)
Number of observations	13,683	10,010	10,010
Panel B: Cognitive Ability			
Peabody	1.87	1.59	0.68
	(0.34)	(0.24)	(0.30)
Number of observations	13,115	9,640	9,640
Panel C: Self-esteem			
6-item Rosenberg	0.56	0.02	0.39
	(0.13)	(0.09)	(0.13)
Number of observations	11,917	8,869	8,869
Panel D: Popularity			
Bonacich	0.10	0.06	0.12
	(0.02)	(0.02)	(0.02)
Number of observations	12,883	9,390	9,390

Note: Standard errors clustered at school level in parentheses; The table reports the estimated rank coefficients from estimating equation (4) with the level of SES, all individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) controls as well as separate school and cohort fixed effects. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Different measures of SES are used to compute students' SES ranks. Column (19 represents the baseline results; in column (2), SES is defined in terms of family income (adjusted by household size); in column (3), the baseline specification is estimated for the reduced sample from column (2).

Table B.6: Rank Effect on 4 Factors of Depression

	Positive affect	Depressed affect	Somatic symptoms	Interpersonal probblems
Rank coefficient	-0.08	-0.04	-0.05	-0.05
	(0.02)	(0.01)	(0.02)	(0.01)
Number of observations	13,683	13,683	13,683	13,683

Note: Standard errors clustered at school level in parantheses; The table reports the coefficient on the socioeconomic rank for the 4 facors of depression that have been identified via principal component analysis in Appendix A.1: (i) depressed affect, (ii) positive affect, (iii) somatic symptoms, and (iv) interpersonal problems. Controls include the absolute level of SES, individual controls (age in days, gender, race) and school cohort controls (mean and standard deviation of SES in the cohort, fraction of repeaters, male share, and share of white students in the cohort) as well as separate school and cohort fixed effects. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Wave I cross-sectional weights are used.

Table B.7: Test for Attrition Bias

	(1)	(2)
Rank	0.01	0.01
	(0.01)	(0.01)
Number of observations	13,736	13,736
Level of SES	yes	yes
Individual conrols	yes	yes
Cohort controls	yes	no
School and cohort FE	yes	no
School x cohort FE	no	yes

Note: Standard errors clustered at school level in parantheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) with an indicator for attrition as outcome. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Column (1) uses the specification with separate school and cohort fixed effects and controls for the level of SES, and individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) controls. Column (2) uses school-by-cohort fixed effects and controls for the level of SES and individual characteristics. Wave I cross-sectional weights are used.

Table B.8: Short-Run Effects Based on Long-Run Sample (Wave IV)

	Depression	Cognitive ability	Self-esteem	Popularity
Rank	-0.82	1.63	0.45	0.10
	(0.25)	(0.37)	(0.14)	(0.02)
Number of observations	10,875	10,430	9,523	10,238

Note: Standard errors clustered at school level in parantheses; The table reports the estimated rank coefficients from estimating equation (4) with the level of SES, individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) controls as well as separate school and cohort fixed effects. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. The sample is restricted to individuals that remained in the long-run sample in wave IV. Wave I cross-sectional weights are used.

C Appendix: Additional Tables

C.1 Descriptives - Long Run

Table C.1: Descriptives of the Long-Run Sample

Long-run Outcomes						
	mean	**sd**	**25th**	**median**	**75th**	**n**
Depression	7.40	3.82	5.00	7.00	9.00	10,901
College attendance	0.69	0.46	0.00	1.00	1.00	10,911
College completion	0.35	0.48	0.00	0.00	1.00	10,911
B. Individual Characteristics						
	mean	**sd**	**25th**	**median**	**75th**	**n**
SES	13.39	3.59	12.00	12.00	16.00	10,912
Family income	47.22	52.29	23.00	40.00	60.00	8,386
Age	28.47	1.73	27.00	29.00	30.00	10,912
Male	0.46	0.50				10,912
White	0.53	0.50				10,912
Black	0.21	0.41				10,912
Asian	0.06	0.23				10,912
Hispanic	0.15	0.36				10,912

Note: This table describes the sample characteristics of the individuals that remain in the sample in the long-run analysis (wave IV). Panel A describes the main long-run outcome variables. Panel B describes individual sample characteristics, measured in wave I, of this sample. The table displays the mean, standard deviation, and interquartile range of the variables as well as the number of observations. SES is measured in years of education (as outlined in section 2.2), annual family income is measured in thousands U.S. $.

C.2 Balancing Table

Table C.2: Balancing Tests

	(1)	(2)
Polygenic scores for education		
Educational attainment (Okbay et al., 2016)	0.01	0.01
	(0.05)	(0.05)
Educational attainment (Lee et al., 2018)	0.04	0.04
	(0.04)	(0.05)
Polygenic scores for mental health disorders		
Attention-Deficit/Hyperactivity Disorder (Neale et al., 2010)	-0.02	-0.02
	(0.03)	(0.03)
Attention-Deficit/Hyperactivity Disorder (Demontis et al., 2017)	0.03	0.02
	(0.05)	(0.05)
Bipolar Disorder (Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011)	0.01	0.02
	(0.04)	(0.04)
Major Depressive Disorder (Psychiatric GWAS Consortium et al., 2013)	0.02	0.01
	(0.05)	(0.05)
Major Depressive Disorder (Wray et al., 2018)	-0.05	-0.05
	(0.04)	(0.04)
Schizophrenia (Psychiatric Genomics Consortium, 2014)	0.01	0.01
	(0.03)	(0.03)
Mental Health Cross Disorder (Psychiatric Genomics Consortium et al., 2013)	-0.01	-0.02
	(0.03)	(0.03)
Number of observations	3,975	3,961
Level of SES	yes	yes
Individual controls	yes	yes
Cohort controls	yes	no
School and cohort FE	yes	no
School x cohort FE	no	yes

Note: Standard errors clustered at school level in parantheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) with polygenic scores for education and mental health disorders as outcome variables. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Column (1) includes all controls as well as separate school and cohort fixed effects. Column (4) controls for individual characteristics and school-by-cohort fixed effects. Wave I cross-sectional weights are used.

C.3 Risk of Clinical Depression

Table C.3: Average Effect of the Socioeconomic Rank on Risk of Clinical Depression

	(1)	(2)	(3)	(4)
Panel A: Depression				
CES-D $\geq$ 16	-0.05	-0.04	-0.04	-0.04
	(0.01)	(0.01)	(0.01)	(0.01)
Number of observations	13,683	13,683	13,683	13,683
Level of SES	yes	yes	yes	yes
Individual controls	no	yes	yes	yes
Cohort controls	no	no	yes	no
School and cohort FE	yes	yes	yes	no
School x cohort FE	no	no	no	yes

Note: Standard errors clustered at school level in parentheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) with an indicator for CES-D $\geq$ 16 as outcome. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Column (1) only controls for separate school and cohort fixed effects and the absolute level of SES. In column (2), individual controls (age in days, gender, and race) are added. In column (3), school cohort specific controls (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) are additionally included. Column (4) controls for individual characteristics and school-by-cohort fixed effects. Wave I cross-sectional weights are used.

C.4 Alternative Measures for Popularity

Table C.4: Average Effect of the Socioeconomic Rank on Alternative Measures of Popularity

	(1)	(2)	(3)	(4)
Panel A: Depression				
Indegree	0.09	0.08	0.08	0.08
	(0.02)	(0.02)	(0.02)	(0.02)
Number of observations	12,883	12,883	12,883	12,883
Outdegree	0.11	0.08	0.07	0.08
	(0.02)	(0.02)	(0.02)	(0.02)
Number of observations	12,883	12,883	12,883	12,883
Prestige	0.15	0.14	0.14	0.14
	(0.03)	(0.03)	(0.03)	(0.03)
Number of observations	11,731	11,731	11,731	11,730
Level of SES	yes	yes	yes	yes
Individual controls	no	yes	yes	yes
Cohort controls	no	no	yes	no
School and cohort FE	yes	yes	yes	no
School x cohort FE	no	no	no	yes

Note: Standard errors clustered at school level in parentheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) with different measures of popularity as outcome. The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Column (1) only controls for separate school and cohort fixed effects and the absolute level of SES. In column (2), individual controls (age in days, gender, and race) are added. In column (3), school cohort specific controls (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) are additionally included. Column (4) controls for individual characteristics and school-by-cohort fixed effects. Wave I cross-sectional weights are used.

C.5 Main Results With Oringinal Rank Scale

Table C.5: Short-Run Rank Effects - Original Rank Scale

	(1)	(2)	(3)	(4)
Panel A: Depression				
CES-D	-4.37	-3.85	-3.84	-3.97
	(0.85)	(0.83)	(0.86)	(0.84)
Number of observations	13,683	13,683	13,683	13,683
Panel B: Cognitive Ability				
Peabody	9.20	6.78	7.48	6.84
	(1.41)	(1.36)	(1.37)	(1.41)
Number of observations	13,115	13,115	13,115	13,115
Panel C: Self-esteem				
6-item Rosenberg	1.96	2.08	2.22	2.27
	(0.51)	(0.52)	(0.53)	(0.53)
Number of observations	11,917	11,917	11,917	11,917
Panel D: Popularity				
Bonacich	0.54	0.43	0.41	0.43
	(0.09)	(0.08)	(0.09)	(0.09)
Number of observations	12,883	12,883	12,883	12,883
Level of SES	yes	yes	yes	yes
Individual controls	no	yes	yes	yes
Cohort controls	no	no	yes	no
School and cohort FE	yes	yes	yes	no
School x cohort FE	no	no	no	yes

Note: Standard errors clustered at school level in parentheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) with the outcome variables: depression (panel A) cognitive ability (panel B), self-esteem (panel C), and popularity (panel D). The rank variable is **not** re-scaled and the reported coefficients represent the effect of a change from the bottom rank to the top rank, i.e. from rank 0 to rank 1. Column (1) includes separate school and cohort fixed effects and controls for the absolute level of SES. In column (2), individual controls (age in days, gender, and race) are added. In column (3), school cohort specific controls (mean and standard deviation of SES in the cohort, fraction of repeaters, male share, and share of white students in the cohort) are additionally included. Column (4) controls for individual characteristics and school-by-cohort fixed effects. Wave I cross-sectional weights are used.

Table C.6: Long-Run Rank Effects - Original Rank Scale

	(1)	(2)	(3)	(4)
Panel A: Long-run Depression				
CES-D (10 items)	-1.74	-1.69	-1.80	-1.80
	(0.59)	(0.60)	(0.62)	(0.61)
Number of observations	10,901	10,901	10,901	10,901
Panel B: College				
Attending college	0.24	0.17	0.19	0.18
	(0.05)	(0.05)	(0.05)	(0.05)
Completing college	0.21	0.16	0.15	0.17
	(0.05)	(0.04)	(0.04)	(0.05)
Number of observations	10,911	10,911	10,911	10,911
Level of SES	yes	yes	yes	yes
Individual controls	no	yes	yes	yes
Cohort controls	no	no	yes	no
School and cohort FE	yes	yes	yes	no
School x cohort FE	no	no	no	yes

Note: Standard errors clustered at school level in parantheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) for the long-run outcomes: the 10-item CES-D (panel A) and dummies for college completion and college attendance (panel B). The rank variable is **not** re-scaled and the reported coefficients represent the effect of a change from the bottom rank to the top rank, i.e. from rank 0 to rank 1. Column (1) includes separate school and cohort fixed effects and controls for the absolute level of SES. In column (2), individual controls (age in days, gender, and race) are added. In column (3), school cohort specific controls (mean and standard deviation of SES in the cohort, fraction of repeaters, male share, share of white students in the cohort) are additionally included. Column (4) controls for individual characteristics and school-by-cohort fixed effects. Wave IV cross-sectional weights are used.

C.6 Main Results Without Sampling Weights

Table C.7: Short-Run Rank Effects - No Sampling Weights

	(1)	(2)	(3)	(4)
Panel A: Depression				
CES-D	-0.87	-0.70	-0.73	-0.73
	(0.15)	(0.15)	(0.15)	(0.14)
Number of observations	13,683	13,683	13,683	13,683
Panel B: Cognitive Ability				
Peabody	1.72	1.16	1.37	1.36
	(0.30)	(0.28)	(0.26)	(0.26)
Number of observations	13,115	13,115	13,115	13,115
Panel C: Self-esteem				
6-item Rosenberg	0.46	0.46	0.52	0.52
	(0.10)	(0.10)	(0.10)	(0.10)
Number of observations	11,917	11,917	11,917	11,917
Panel D: Popularity				
Bonacich	0.10	0.08	0.08	0.08
	(0.03)	(0.02)	(0.02)	(0.02)
Number of observations	12,883	12,883	12,883	12,883
Level of SES	yes	yes	yes	yes
Individual controls	no	yes	yes	yes
Cohort controls	no	no	yes	no
School and cohort FE	yes	yes	yes	no
School x cohort FE	no	no	no	yes

Note: Standard errors clustered at school level in parantheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) with the outcome variables: depression (panel A) cognitive ability (panel B), self-esteem (panel C), and popularity (panel D). The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Column (1) includes separate school and cohort fixed effects and controls for the absolute level of SES. In column (2), individual controls (age in days, gender, and race) are added. In column (3), school cohort specific controls (mean and standard deviation of SES in the cohort, fraction of repeaters, male share, and share of white students in the cohort) are additionally included. Column (4) controls for individual characteristics and school-by-cohort fixed effects. No weights are used.

Table C.8: Long-Run Rank Effects - No Sampling Weights

	(1)	(2)	(3)	(4)
Panel A: Long-run Depression				
CES-D (10 items)	-0.39	-0.38	-0.41	-0.39
	(0.08)	(0.08)	(0.08)	(0.09)
Number of observations	10,901	10,901	10,901	10,901
Panel B: College				
Attending college	0.05	0.03	0.04	0.04
	(0.01)	(0.01)	(0.01)	(0.01)
Completing college	0.05	0.03	0.04	0.04
	(0.01)	(0.01)	(0.01)	(0.01)
Number of observations	10,911	10,911	10,911	10,911
Level of SES	yes	yes	yes	yes
Individual controls	no	yes	yes	yes
Cohort controls	no	no	yes	no
School and cohort FE	yes	yes	yes	no
School x cohort FE	no	no	no	yes

Note: Standard errors clustered at school level in parantheses; The table reports the estimated coefficients on the socioeconomic rank from different specifications of equation (4) for the long-run outcomes: the 10-item CES-D (panel A) and dummies for college completion and college attendance (panel B). The rank variable is re-scaled such that the reported coefficients represent the effect of a 25 percentile increase in rank. Column (1) includes separate school and cohort fixed effects and controls for the absolute level of SES. In column (2), individual controls (age in days, gender, and race) are added. In column (3), school cohort specific controls (mean and standard deviation of SES in the cohort, fraction of repeaters, male share, share of white students in the cohort) are additionally included. Column (4) controls for individual characteristics and school-by-cohort fixed effects. No weights are used.

C.7 School Quality

Table C.9: Correlation of School Fixed Effects and Indicators of School Quality

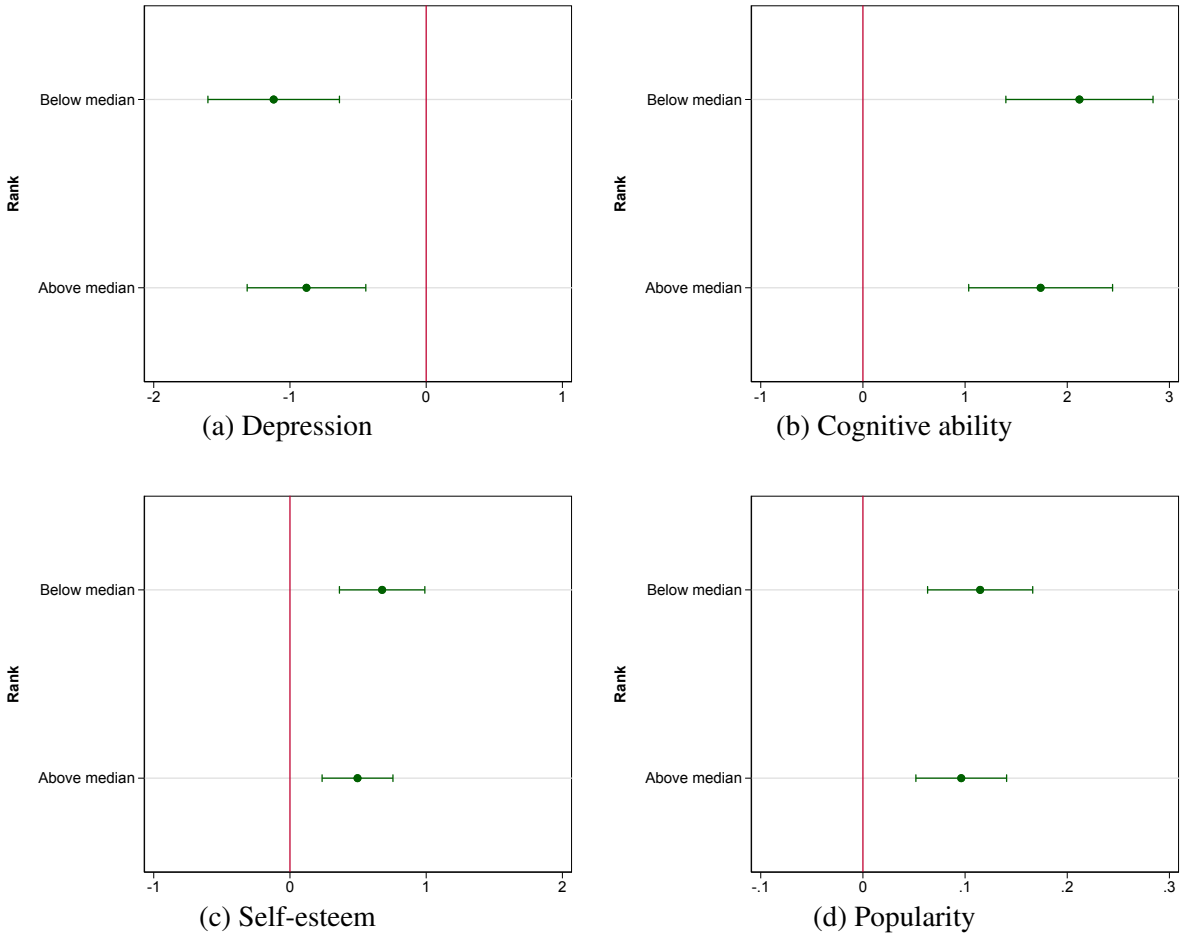
	Short-run outcomes		Long-run outcomes		
	Depression	Cognitive ability	Depression	College attendance	College completion
Average SES	-0.71	0.88	0.07	0.05	-0.05
	(0.16)	(0.37)	(0.09)	(0.01)	(0.01)
Fraction: college students	0.79	0.82	-1.63	0.50	0.84
	(1.32)	(3.06)	(0.78)	(0.10)	(0.06)
Average class size	-0.00	-0.17	-0.01	0.00	-0.00
	(0.03)	(0.07)	(0.02)	(0.00)	(0.00)
% School drop outs	-0.01	0.02	0.01	-0.00	-0.00
	(0.04)	(0.09)	(0.02)	(0.00)	(0.00)
% Student retention	0.02	-0.20	0.02	0.00	-0.00
	(0.03)	(0.07)	(0.02)	(0.00)	(0.00)
Teacher-student ratio	-4.72	0.15	-2.19	0.46	0.02
	(2.46)	(5.69)	(1.44)	(0.19)	(0.12)
% Teachers with MA or higher	-0.01	-0.00	0.00	0.00	-0.00
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
% Teacher with $>$ 5 year tenure	0.01	-0.00	-0.00	0.00	-0.00
	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)
% Teacher with $<$ 1 year tenure	-0.03	0.02	0.00	-0.00	0.00
	(0.02)	(0.04)	(0.01)	(0.00)	(0.00)
Additional controls:					
School size	yes	yes	yes	yes	yes
Region	yes	yes	yes	yes	yes
Urbanicity	yes	yes	yes	yes	yes
School type	yes	yes	yes	yes	yes

Note: Standard errors reported in parantheses; The table reports the estimated coefficients of a regression of the school fixed effects on different indicators of school quality. The school fixed effects are estimated from equation (4) with the level of SES, individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) controls as well as separate school and cohort fixed effects for each of the main short- and long-run outcomes: depression and cognitive ability in the short-run and depression, college attendance, and college completion in the long-run. Measures of school quality include the school-average SES, the fraction of students attending college in the long-run, the average class size, the average percent of dropouts across all grades, the ratio of full-time teachers to students, and the percent of teachers with at least an MA degree.

D Appendix: Additional Figures

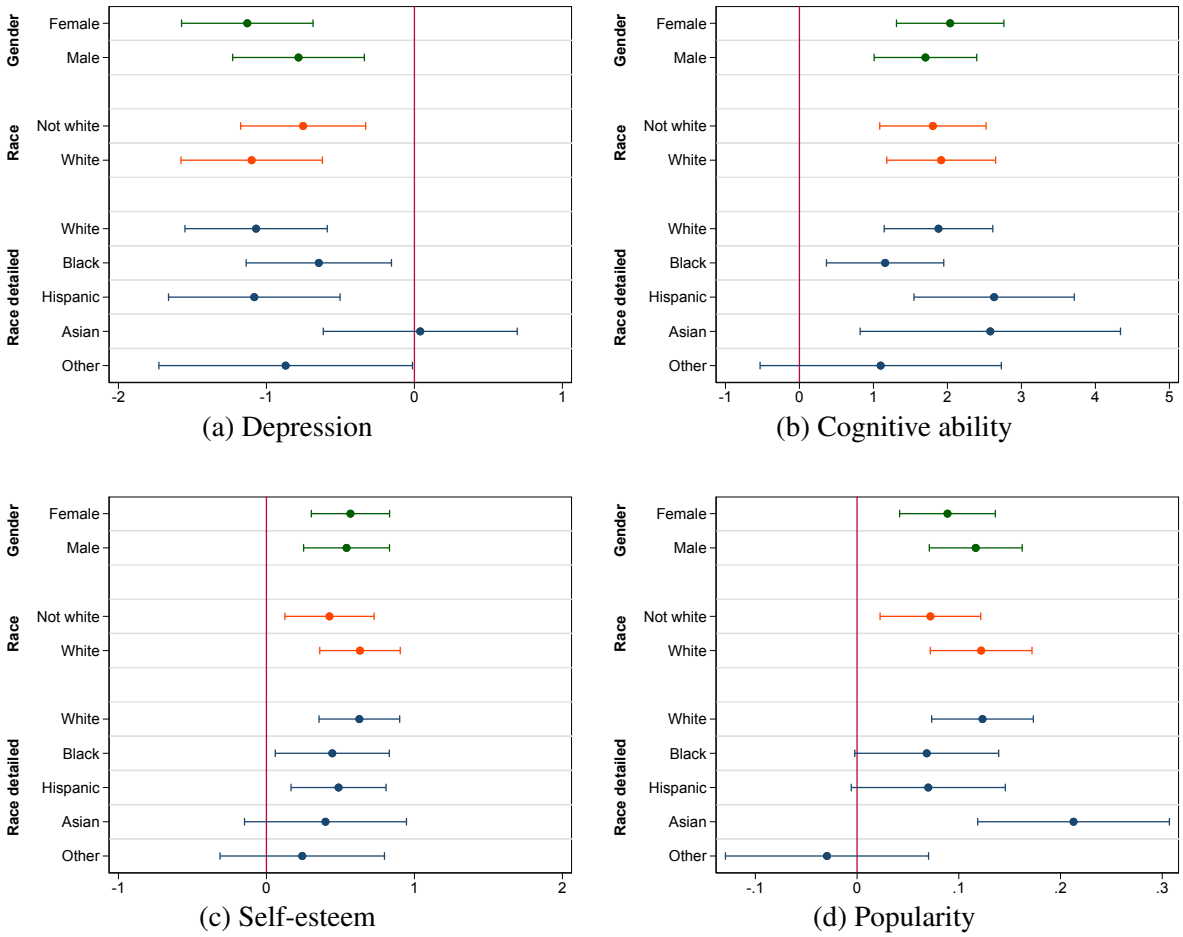
D.1 Heterogeneous Effects

Figure D.1: Heterogeneity by Rank



Note: The figure shows the rank effect for each of the short-run outcomes for the two subgroups: (i) students with a rank at or below the median, and (ii) students with a rank above the median. It displays point estimates with 95% confidence intervals. To get the depicted coefficients, the rank is interacted with an indicator variable $\mathbb{1}(Rank_{isc} > 0.5)$ in equation (4) with separate school and cohort fixed effects and controls for the level of SES, individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) characteristics. The depicted rank coefficients are re-scaled to represent the effect of a 25 percentile increase in the socioeconomic rank. Wave I cross-sectional weights are used.

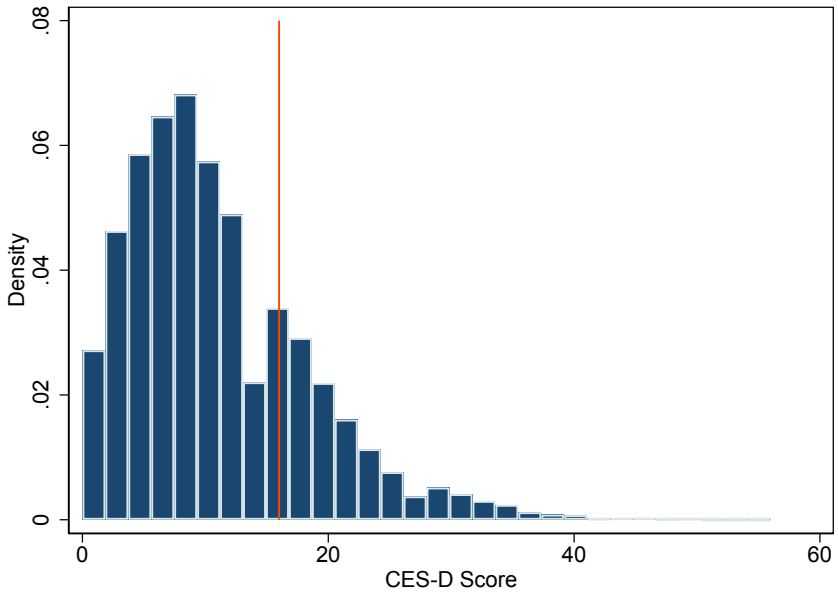
Figure D.2: Heterogeneity by Individual Characteristics



Note: The figure shows heterogeneities in the rank effect for each of the four outcomes by gender and race. It displays point estimates with 95% confidence intervals. To get the depicted coefficients, the rank is interacted with dummies for either gender or race in equation (4) with separate school and cohort fixed effects and controls for the level of SES, individual (age in days, gender, and race) and school cohort (mean and standard deviation of SES in the cohort, fraction of repeaters, the male share, and the share of white students in a cohort) characteristics. The rank variable is re-scaled such that the depicted coefficients represent the effect of a 25 percentile increase in rank. Wave I cross-sectional weights are used.

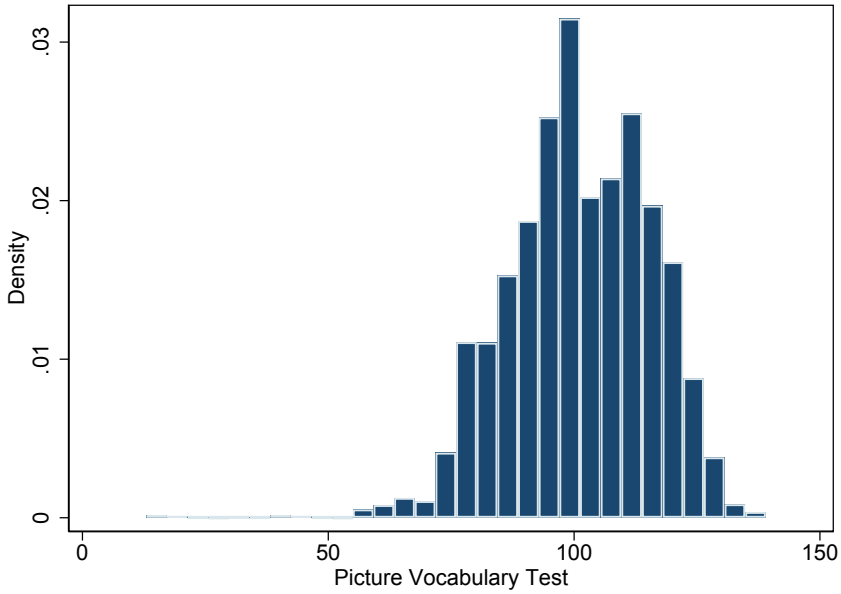
D.2 Outcome Measures

Figure D.3: Distribution of the Depression Score



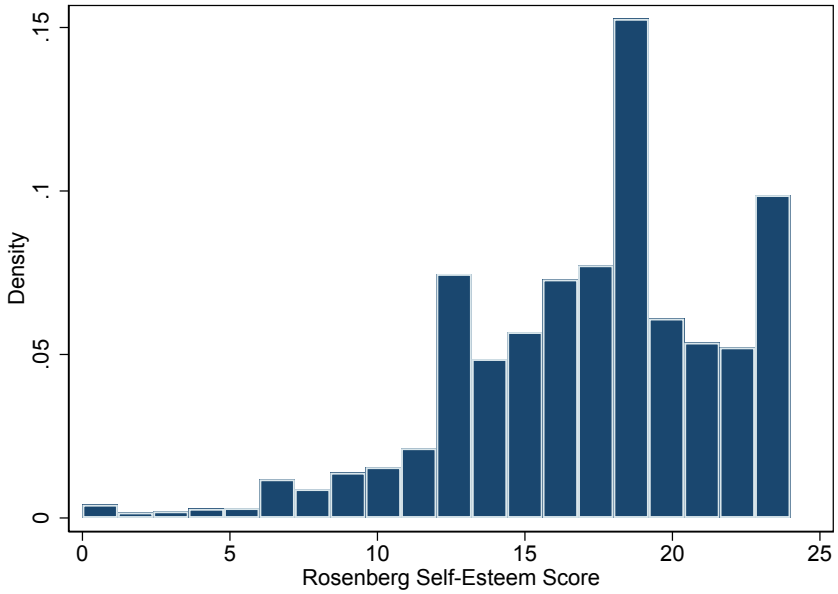
Note: The figure displays the distribution of depression scores (CES-D) in the short-run sample with 13,683 observations. The red line represents a score of 16, which is a common cut-off for being at risk for clinical depression.

Figure D.4: Distribution of the Cognitive Ability Score



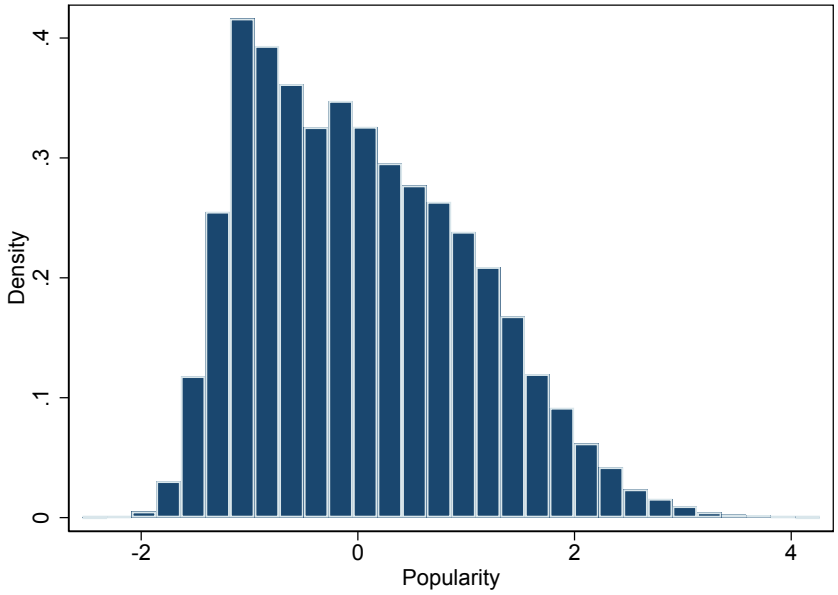
Note: The figure displays the distribution of the cognitive (Peabody) test scores in the short-run sample with 13,115 observations.

Figure D.5: Distribution of the Self-esteem Score



Note: The figure displays the distribution of the self-esteem scores in the short-run sample with 11,917 observations.

Figure D.6: Distribution of the Popularity Score



Note: The figure displays the distribution of the popularity scores in the short-run sample with 12,883 observations.