STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS

THE LONG-RUN EFFECTS OF PEERS ON MENTAL HEALTH

BY

LUKAS KIESSLING AND JONATHAN NORRIS

NO 20-06

DEPARTMENT OF ECONOMICS

UNIVERSITY OF STRATHCLYDE GLASGOW

The Long-Run Effects of Peers on Mental Health*

Lukas Kiessling Jonathan Norris

June 1, 2020

Abstract

This paper studies how peers in school affect students' mental health. Guided by a theoretical framework, we find that increasing students' relative ranks in their cohorts by one standard deviation improves their mental health by 6% of a standard deviation conditional on own ability. These effects are more pronounced for low-ability students, persistent for at least 14 years, and carry over to economic long-run outcomes. Moreover, we document a strong asymmetry: Students who receive negative rather than positive shocks react more strongly. Our findings therefore provide evidence on how the school environment can have long-lasting consequences for the well-being of individuals.

Keywords: Peer Effects, Mental Health, Depression, Rank Effects
JEL-Codes: I21, I14, J24

*Lukas Kiessling: Max Planck Institute for Research on Collective Goods, lkiessling@coll.mpg.de; Jonathan Norris, University of Strathclyde, jonathan.norris@strath.ac.uk. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

1 Introduction

Mental health is a growing concern with substantial costs for the economy both in the United States and around the world. In particular, the total costs of mental health disorders are estimated to be as high as 2.5% of the GDP in the U.S. and 3.5% in Europe (OECD, 2015). Many of these mental health issues can be traced back to symptoms during youth as about 20% of all adolescents suffer from diagnosable mental health disorders (Kessler, Angermeyer, and Anthony, 2007), and this number increased by a third between 2005 and 2014 (Mojtabai, Olfson, and Han, 2016). It is therefore important to understand the causes and long-term consequences of mental health disorders during school-age.

In this paper, we study how features of the school environment – in our application information shocks about one's own ability – affect the mental health of students, and how these effects evolve over time. We begin by formulating a theoretical framework in which students are uncertain about their ability and face two decisions: how much time to spend on studying and how much effort to exert while studying. While school performance increases in study time, the return to studying depends on effort choices. Students can either shirk and have a low return to effort or exert effort to have a return equal to their ability. Only the latter allows students to learn about their ability. Due to noise in the educational production function, students learn only imperfectly from observed performance.

Based on this framework, we derive four predictions: First, negative shocks to performance in school decrease a student's belief about their own ability, which we interpret as a particular form of mental health, namely depression, while positive shocks increase this belief (see, e.g., de Quidt and Haushofer, 2016, for a similar assumption). Our interpretation is motivated by leading cognitive (e.g., Beck, 1967) and attributional theories (Seligman, 1972), as well as newer cognitive neuropsychological models of depressions (Clark, Chamberlain, and Sahakian, 2009), which emphasize the crucial role of biased or incorrect beliefs as a source of depression. Also, from the observation that many common depressive symptoms can be seen either as manifestations of these biased beliefs or as direct consequences thereof (de Quidt and Haushofer, 2016). Second, since students refrain from exerting effort once their beliefs about their own ability are sufficiently low, they do not receive any new informative signals implying that negative shocks have stronger effects than positive shocks. Third, the consequences of shocks are more pronounced at the lower end of the ability distribution as students attribute shocks relatively more to their own ability. Fourth, these shocks can have persistent effects over time.

To test these predictions, we exploit idiosyncratic variation in the ability composition of school cohorts as shocks to students' mental health. More specifically, we argue that conditional on a student's school, the ability composition of peers and – holding own and peer ability constant – a student's rank in their cohort is as good as random. We use data from the National Longitudinal Study of Adolescent to Adult Health (AddHealth), a survey of a representative sample of U.S. adolescents in grades 7-12, first interviewed during the 1994/1995 academic year. Importantly, AddHealth repeatedly assesses students' mental health using a well-established measure to diagnose depressions (Center of Epidemiological Studies Depression Scale, CES-D; Radloff, 1977). Based on this measure, we further motivate our analysis by documenting that mental health seems to be malleable before the age of 20, i.e., when respondents are still in school, and show that mental health is persistent over the life-cycle.

We find that shocks measured by students' ranks in their cohorts affect mental health. Increasing a student's rank by one standard deviation improves their mental health by 6% of a standard deviation (SD) on average. This effect is of similar size as the difference in the mental health of children from college and non-college educated households, or two thirds of the difference between students raised by both parents rather than a single parent. Furthermore, our effect size amounts to about a sixth to a third of average effect sizes reported for positive psychology interventions targeted at psychological or subjective well-being and depression (e.g., trainings, exercises and therapies aimed at increasing positive feelings; Bolier et al., 2013). Our effects stem from natural variation in the ability composition of school cohorts rather than from targeted intervention. Thus, we interpret the estimated effects as large, providing strong support for the central prediction of our model that such shocks (e.g., variation in ranks) affect mental well-being.

We then explore the robustness of this result. First, we find that our estimates do not change once we include additional controls. These include (non-linear) peer effects in several dimensions, allowing for school-specific time trends, or identifying effects within school cohorts rather than within schools and across cohorts.[1] Second, we show that our effects do not depend on functional form assumptions by (i) adopting a post-double selection (PDS) Lasso proposed by Belloni, Chernozhukov, and Hansen (2014), (ii) using several alternative parametric specifications, and (iii) using different definitions of ranks. Third, a sensitivity analysis based on Oster (2019) indicates that unobservables are unlikely to drive our estimates and we further show evidence that sorting into schools based on ranks is unlikely to occur. Finally, we explore several sources of non-classical measurement error using a series of Monte Carlo simulations. These simulations suggest that measurement error is unlikely to explain our results. In fact, they consistently lead us to underestimate the magnitude of the actual effects.

Subsequently, we explore whether our effects are driven by positive or negative shocks. In line with the second prediction of our theoretical framework, we provide evidence that our results are driven by negative shocks, i.e., having a lower rank in one's school cohort than across students from all schools and cohorts. Moreover, we explore specific facets of mental health underlying the CES-D measure. We find that the ranks mainly induce a lack of positive attitudes and a lack of motivation rather than affecting loneliness and attributions to other external or social factors.

[1]Rather than exploiting variation in the ability distribution across cohorts and within schools, the last strategy listed follows Murphy and Weinhardt (forthcoming) and exploits variation of students' ranks within cohorts and compares these students to others from different cohorts and schools after all observed and unobserved characteristics on a school-cohort level are removed.

Moving to our third prediction, we find that this effect is indeed larger at the lower end of the ability distribution: Effects for students from the lowest decile of the ability distribution are about three times larger than average effects and amount to approximately 0.70 SD. This effect slowly fades out for higher deciles. When studying further heterogeneities, we find that students with different sociodemographic characteristics are affected rather equally. Furthermore, the rank effect does not vary strongly with school characteristics, with one exception: lower teacher-student ratios are associated with smaller rank effects. This suggests that less teacher turnover in a given classroom fosters student-teacher relationships alleviating the consequences of negative shocks.

Next, we explore how the heterogeneous effects in own ability evolve over time. We find that these effects are persistent and last from adolescence to adulthood in line with our fourth prediction. In fact, we observe that rank effects for low-ability students are at least as large in wave IV, 14 years after baseline when individuals in our dataset are 26-32 years old, as they are in wave I, when they were adolescents, and fade out for higher deciles.

Finally, we investigate whether these effects carry over to other economic long-run outcomes and find that better mental health in adolescence is highly predictive of better economic long-run outcomes in adulthood. We find significant and sizable associations of CES-D scores and college graduation, employment status, income, as well as paying bills on time, ever being married, and ever being arrested conditional on a rich set of background characteristics. While these effects are correlational in nature, we also estimate the causal effect of ranks among peers on these long-run outcomes and calculate how much of these effects are mediated by mental health in adolescence. Notably, we find the same heterogeneous pattern in ability that we observed for mental health, and we estimate that across outcomes at least 5 to 7% of these effects are mediated by mental health in adolescence.

We focus on mental health for several reasons. First, mental health and psychological well-being are important outcomes in their own right by directly entering an individual's utility function (Kahneman and Deaton, 2010). As a consequence, there exist several studies estimating the causal relationship of, e.g., early-life circumstances (Adhvaryu, Fenske, and Nyshadham, 2019), income shocks from cash transfers (Baird, de Hoop, and Özler, 2013), religion (Fruehwirth, Iyer, and Zhang, 2019), and psychotherapy interventions (Baranov et al., 2020) on mental health. Second, the relationship between economic outcomes and mental health is bi-directional as mental disorders can also be a cause of important economic outcomes. For instance, low mental health limits human capital accumulation (Currie and Stabile, 2006; Fletcher, 2010) and reduces employment and earnings (Bartel and Taubman, 1986; Biasi, Dahl, and Moser, 2019; Ettner, Frank, and Kessler, 1997; Fletcher, 2014; Frank and Gertler, 1991; Stewart et al., 2003). Third, mental health can also be a likely mechanism that potentially explains how other economic outcomes are affected by shocks in early life or adolescence (e.g., Persson and Rossin-Slater, 2018, on maternal stress and the consequences for children).

Moreover, this paper adds to an accumulating evidence base on the long-lasting effects of features of the school environment. Although analyzing the long-run effects of smaller classes (Angrist and Lavy, 1999; Angrist et al., 2019; Chetty et al., 2011; Krueger and Whitmore, 2001) or better teachers (e.g., Chetty, Friedman, and Rockoff, 2014; Rothstein, 2017) are a long-standing and active fields of research, only recently have studies shed light on the long-term effects of peers during school: Carrell, Hoekstra, and Kuka (2018) document that having disruptive peers during childhood decreases earnings, while Olivetti, Patacchini, and Zenou (2020) show that the labor supply of school peers' mothers affects the females' labor force participation in adulthood, and several studies suggest that peers in school have effects on educational attainment and major choices (as in, e.g., Anelli and Peri, 2019; Bifulco et al., 2014; Bifulco, Fletcher, and Ross, 2011; Black, Devereux, and Salvanes, 2013; Gould, Lavy, and Paserman, 2009) and the formation of non-cognitive skills (Bietenbeck, 2020). In addition, there exists a small literature studying peer effects in mental health, but its evidence so far is mixed with strong spillover effects in some studies (e.g., Fowler and Christakis, 2008) and only modest (Eisenberg et al., 2013) or zero effects Zhang, 2018 in others. Our results provide novel evidence on the long-term effects of school peers for mental health proposing mental health as one link how the school environment affects important long-run outcomes in adulthood.

Because we exploit a specific shock based on the students' ability rank in their cohort, our study relates to a small literature on rank effects as a specific form of peer effects. This literature argues that ordinal ranks affect outcomes due to social comparisons and describe these effects as "big fish in a little pond" effects (Festinger, 1954; Marsh, 1987). In particular, there is evidence that such comparisons affect the job satisfaction or general well-being of individuals (Brown et al., 2008; Card et al., 2012; Luttmer, 2005).[2] Recent papers have used these ideas in educational contexts to estimate the effect of ordinal ranks on educational outcomes (Delaney and Devereux, 2019; Elsner and Isphording, 2017; Elsner, Isphording, and Zölitz, 2019; Murphy and Weinhardt, forthcoming) and subsequent earnings (Denning, Murphy, and Weinhardt, 2018). Moreover, there is evidence on rank effects for risky behavior (Elsner and Isphording, 2018) and skill development (Pagani, Comi, and Origo, forthcoming).

We use a similar empirical approach to that employed in these previous papers, and we guide our analysis using a theoretical framework based on learning about one's own ability in the presence of noisy signals. In line with predictions from our model, we document strong, heterogeneous, and persistent rank effects on mental health. Apart from studying a novel margin of rank effects and the persistence of effects over time, we move beyond previous papers by introducing asymmetric effects, differentiating between positive and negative shocks. The idea behind these is that, due to limited cohort sizes, the rank within a school cohort differs from the rank in the cohort across all schools. Hence, this yields either positive or negative shocks relative to a rank one might expect. Confirming our theoretical prediction, we find that

[2]Relative rankings may also affect behavior and outcomes such as consumption, well-being, performance, and effort provision, as in Hopkins and Kornienko (2004), Luttmer (2005), Gill et al. (2019), and Kuziemko et al. (2014), respectively.

rank effects on mental health are driven by negative shocks. In addition, we also show that the same patterns hold several long-run outcomes such as college graduation, employment status, earnings, or arrests, which are also persistently related to mental health measured during adolescence.

Taken together, our results provide evidence on how the school environment can have long-lasting effects on mental health. We show that shocks not only affect immediate mental health – especially for students of low ability – but have persistent effects over at least 14 years. Therefore, our results lend support to models that introduce mental health capital similar to general health (Grossman, 1972), or as a malleable skill in line with the growing evidence on the importance of non-cognitive skills in the development of children (Cunha and Heckman, 2008; Cunha, Heckman, and Schennach, 2010). Such models could explain why we observe long-lasting effects carrying over to educational and labor market outcomes, which is similar to findings for non-cognitive skills (e.g., as shown by Heckman, Stixrud, and Urzua, 2006).

The remainder of this paper is structured as follows. In Section 2, we lay out a theoretical model in which students are uncertain about their ability and base it on noisy signals. We present the data we use to test our predictions in Section 3 and describe mental health patterns over the life-cycle in Section 4. Our empirical strategy and main results for wave I are presented in Sections 5 and 6, before we study the persistence of our effects in Section 7. Finally, Section 8 concludes.

2 Theoretical Framework

Before we turn to our empirical analysis, we outline a simple belief-updating model to formalize our predictions on how different shocks may affect the mental well-being of students.[3] Our model is motivated by cognitive (e.g. Beck, 1967) and attributional theories of depression (e.g., Seligman, 1972), which emphasize the crucial role of biased or incorrect beliefs as a source of depression, as well as cognitive neuropsychological models of depression (Clark, Chamberlain, and Sahakian, 2009; Roiser, Elliott, and Sahakian, 2012), which suggest a central role for negative affective biases. We closely follow de Quidt and Haushofer (2016) and conceptualize the belief about the returns to one's own ability as a proxy for mental health.[4] In fact, many common symptoms of depression, such as pessimism, low self-esteem, lack of motivation, or sadness, can be seen either as a manifestation of biased beliefs (as in the case of pessimism) or as a direct consequence thereof (as for the lack of motivation).

Our model builds on two features. First, students are uncertain about the return to effort and learn about their ability based on their performance in school (for evidence of students' imperfect knowledge about their ability and return to effort, see, e.g., Jensen, 2010; Stine-

[3] In our empirical application, shocks will be information shocks from quasi-exogenous variation in the students' ordinal ranks.

[4] In their model, de Quidt and Haushofer (2016) focus on income shocks in developing countries and additionally consider non-food consumption, food, as well as sleep and other domains entering the utility function. Relative to their model, ours can be seen as a simplification, translating their model to an educational context.

brickner and Stinebrickner, 2012; Zafar, 2011). Second, there are exogenous shocks affecting school performance and students use their performance to update their prior about their ability. Thus, after receiving negative signals, students reduce their belief about their ability, while positive signals increase their beliefs. Yet, updating only occurs if students exert effort to receive a good grade. If they shirk, they will not attribute their educational success to their ability. Thus, if the prior about ability is sufficiently low, students may refrain from exerting any effort to avoid further negative signals. This implies that a low mental health status or low belief about one's own ability may constitute an absorbing state, in which no further updating occurs.

To formalize this intuition and derive more precise predictions, we adopt a simple educational production function, in which "school success" depends on own ability, effort, and exogenous shocks. More specifically, let A_i denote a student's ability or return to effort which is drawn from some distribution F_A. For the ease of exposition, we keep the individual index i implicit and present the model for a single individual with ability A. Let y_t denote the "school success" in period t. We specify the educational production function as

$$y_t = [Ae_t + \underline{A}(1 - e_t)]s_t + \epsilon_t,$$

in which school success depends on the amount the student studies, s_t, and their decision to exert high ($e_t = 1$) or low effort ($e_t = 0$). High effort yields a return to studying equal to their ability A, while shirking yields a low return of $\underline{A}$, assumed to be known to the student. Moreover, school success is subject to exogenous shocks $\epsilon_t \sim N\big(0, \sigma^2_{\epsilon,t}\big)$. In the empirical part of our paper, we will use shifts in the rank of a student due to having better classmates as a shock to the school success y_t.[5]

Students are uncertain about their own ability and have priors or beliefs about their ability denoted by μ_t. Hence, from a student's perspective, their ability is a random variable $A \sim N\big(\mu_t, \sigma^2_{A,t-1}\big)$. We assume that students maximize their expected utility by allocating time between studying and leisure. While studying increases educational success, leisure also enters positively into the utility function. Their expected decision utility function is given by

$$EU(e_t, s_t, l_t | \mu_t) = \underbrace{\big[\mu_t e_t + \underline{A}\big(1 - e_t\big)\big] s_t}_{\text{Exp. school success}} + \underbrace{\phi\big(l_t\big)}_{\text{Utility from leisure}},$$

in which s_t and l_t denote study and leisure time, respectively; total time available is normalized to 1 such that $s_t + l_t = 1$, and $\phi'(\cdot) > 0$, $\phi'' \leq 0$. Expected school success depends on the prior about ability, μ_t, and the decision to exert effort, e_t, as described above.

[5]As we will explain in detail in Section 5, we exploit the fact that schools only have limited size and that the ability composition of students varies across cohorts and schools. This implies that a student with a specific ability may be ranked highly in one cohort, but would only be a student in the middle of the ability distribution in another cohort. We use this variation in the ability distribution as an exogenous shocks affecting students' beliefs after conditioning on own and peer ability.

Given this setup, a student's optimal effort decision is

$$e_t^* = \mathbb{1}\{\mu_t > \underline{A}\},$$

and, hence, we can replace $[\mu_t e_t^* + \underline{A}(1-e_t^*)] = \max\{\mu_t, \underline{A}\}$.[6] The optimal time spent studying therefore equals

$$s_t^* = 1 - \phi'^{-1}\left[\max\left\{\mu_t, \underline{A}\right\}\right]$$

and is increasing in perceived own ability.

We now want to characterize how students learn about their ability. Consider students who want to update their prior beliefs μ_{t-1} given they received a signal y_{t-1} about their own ability. If the students only exerted low study effort, $e_{t-1} = 0$, they do not learn new information about their ability A as studying yields a fixed return $\underline{A}$. However, if they exerted high effort ($e_{t-1} = 1$), they can learn about their ability. For this, rewrite y_{t-1} as follows:

$$y_{t-1} = As_{t-1} + \epsilon_{t-1} =: x_{t-1} + \epsilon_{t-1},$$

where $x_{t-1} = As_{t-1} \sim N(\mu_{t-1}s_{t-1}, \sigma_{A,t-1}^2 s_{t-1}^2)$. Given the signal about school success, y_{t-1}, the students try to learn about their ability, A. Using the new notation, they want to infer the expected value of x_{t-1} given y_{t-1}, i.e., the posterior $E[x_{t-1}|y_{t-1}]$:

$$\begin{aligned}
E[x_{t-1}|y_{t-1}] &= \frac{\sigma_\epsilon^2}{Var(x_{t-1}) + \sigma_\epsilon^2}E[x_{t-1}] + \frac{Var(x_{t-1})}{Var(x_{t-1}) + \sigma_\epsilon^2}y_{t-1} \\
&= E[x_{t-1}] + \frac{Var(x_{t-1})}{Var(x_{t-1}) + \sigma_\epsilon^2}(y_{t-1} - E[x_{t-1}]) \\
&= \mu_{t-1}s_{t-1} + \frac{\sigma_{A,t-1}^2}{\sigma_{A,t-1}^2 + \sigma_\epsilon^2/s_{t-1}^2}\left[\left(A - \mu_{t-1}\right)s_{t-1} + \epsilon_{t-1}\right].
\end{aligned}$$

Hence, the corresponding posterior belief μ_t is:

$$\mu_t = \mu_{t-1} + \frac{\sigma_{A,t-1}^2}{\sigma_{A,t-1}^2 + \sigma_\epsilon^2/s_{t-1}^2}\left[\left(A - \mu_{t-1}\right) + \frac{\epsilon_{t-1}}{s_{t-1}}\right].$$

Several results emerge. First, negative shocks ($\epsilon_{t-1} < 0$) decrease students' beliefs about their ability (i.e., μ_t decreases) and thus have detrimental effects on mental health, while positive shocks ($\epsilon_{t-1} > 0$) benefit mental health.

Prediction 1. *Positive shocks improve mental health, whereas negative shocks decrease mental health.*

[6]Assuming e_t to be continuous on $[0, 1]$ would not affect this results as linearity implies a corner solution.

Second, once the student's belief μ_t decreases below $\underline{A}$, the student withdraws effort and thus stops updating.[7] This implies that negative shocks may have more pronounced consequences relatively to positive shocks, as negative shocks decrease the likelihood of a student receiving informative signals in the future.

Prediction 2. *There are asymmetric effects of positive and negative shocks, with the latter being more pronounced.*

Third, the students' study time, s_t, (weakly) decreases in the belief about their ability and low-ability students have lower priors μ_{t-1}. This implies that shocks have stronger effects for low-ability students as the term ϵ_{t-1}/s_{t-1} becomes larger.

Prediction 3. *The consequences of shocks are more pronounced for low-ability individuals.*

Fourth, given the lower propensity to update after receiving a negative shock and stronger effects for low-ability students, shocks have persistent effects over time and especially so for low-ability students with priors close to $\underline{A}$.

Prediction 4. *The effects of shocks are persistent over time. They are more pronounced for students with low ability.*

In summary, our theoretical framework predicts that if students have imperfect knowledge about their ability, they learn about it by receiving signals through their school success. Negative shocks to students' school success decrease their beliefs about their ability, those effects are more pronounced in the lower part of the ability distribution, and persist over time.

3 Data

In order to test the predictions from the previous section, we use restricted data from the National Longitudinal Study of Adolescent to Adult Health (AddHealth). AddHealth is a longitudinal study of a set of representative middle and high schools in the United States.

For our analysis, AddHealth has several key features. First, it covers multiple cohorts within schools, which we need for our empirical strategy exploiting variation within schools across cohorts. Second, a representative set of students from each cohort is sampled. Third, students were first interviewed in 1994/95, when students were between 13 and 18 years old, and followed for five waves until 2016-2018, when respondents were 36-42 years old. Hence, we can follow the development of adolescents' well-being well into adulthood. Fourth, the dataset has repeated measures of an established mental health self-assessment and a standardized test of cognitive ability. In the following, we discuss the mental health measure in more detail and defer the discussion of the ability measure to Section 5, where we discuss the empirical strategy.

[7]Evidence in line with this mechanism has been found in the psychology literature. Kuppens, Allen, and Sheeber (2010) show that individuals with low self-esteem or depressions display high levels of emotional inertia in response to emotional fluctuations relative to individuals with normal levels of self-esteem and no depressions. Relatedly, Korn et al. (2014) document that depression is related to more pessimistic belief updating.

3.1 Data on Students' Mental Health

We assess mental health of students using the Center of Epidemiologic Studies Depression Scale (CES-D, Radloff, 1977), an established screening measure to test for depression and depressive disorders that is one of the most widely-used instruments in psychiatric epidemiology. The CES-D consists of 19 symptoms (e.g., "You felt sad") and asks respondents how often each symptom applied to them over the course of the past week. Responses are then rated on a scale from 0 ("never or rarely") to 3 ("most of the time or all of the time") and aggregated to a final score ranging from 0 to 57, with higher scores indicating a higher propensity for depressive symptoms. In particular, a score of 16 or higher is commonly interpreted as an indicator for depression (Radloff, 1977). Appendix Table A.1 presents all items of the CES-D score.

The CES-D scale is a widely-used instrument to study mental health: it has been adopted to study how far an individual's mental health status spreads through a social network (Fowler and Christakis, 2008; Rosenquist, Fowler, and Christakis, 2011), the effect of mental health for educational attainment (Fletcher, 2008, 2010), and the consequences of wealth shocks (Schwandt, 2018), cash transfers (Haushofer and Shapiro, 2016), or religion (Fruehwirth, Iyer, and Zhang, 2019) on mental health. Moreover, a rich literature in psychology and psychiatric epidemiology has examined the concurrent validity (i.e., the extent to which the CES-D and a subsequent diagnosis coincide; e.g., Lewinsohn et al., 1997), reliability, and internal consistency of the CES-D scale (e.g., Radloff, 1991; Roberts et al., 1990), and it is frequently used in clinical practice (Murphy, 2011).

We present the distribution of the CES-D in Appendix Figure A.1: The distribution is highly skewed and about 25% of all respondents can be classified as depressive (i.e., have a CES-D score above 16). In the main part of our analysis, we focus on the 19-item CES-D scale as a measure of mental health. Yet, later waves only administered a short scale comprised of a subset of the original items. Thus, when studying the persistence of our results, we scale the CES-D scores of later waves to obtain a comparable measure across waves.[8]

3.2 Summary Statistics

We present summary statistics of our sample in Table 1. After dropping observations from schools with fewer than 20 students in total and 5 students per grade, we observe 18,459 students in wave I. 51% of these students are female and they are on average 15.6 years old. The majority of students are white (53%) and 34% of all students come from college-educated households. Moreover, the mean CES-D score in our sample is 11.3.

[8]Using data from wave I, Appendix Figure A.2 shows that the short and long versions of the CES-D scale are indeed highly correlated. To perform the rescaling, we scale the nine (ten) item scales of wave III (IV) by $19/9$ ($19/10$) to match the 19 item scale.

Table 1. Summary Statistics

	Mean	SD	Min	Max
Female	0.51	0.50	0.00	1.00
Age	15.63	1.70	11.00	19.00
Mental Health (CES-D scores in Wave I)	11.33	7.60	0.00	56.00
Ability (AH PVT scores)	100.17	14.67	13.00	139.00
Ethnicity				
White	0.53	0.50	0.00	1.00
Hispanic	0.17	0.37	0.00	1.00
Black	0.22	0.41	0.00	1.00
Asian	0.07	0.25	0.00	1.00
Other	0.02	0.14	0.00	1.00
Parental Education				
Less than HS	0.18	0.38	0.00	1.00
HS or GED	0.28	0.45	0.00	1.00
Some College	0.20	0.40	0.00	1.00
College	0.22	0.42	0.00	1.00
Postgraduate	0.12	0.32	0.00	1.00
Single Parent Household	0.32	0.47	0.00	1.00
Grade				
Grade 7	0.13	0.34	0.00	1.00
Grade 8	0.13	0.34	0.00	1.00
Grade 9	0.18	0.38	0.00	1.00
Grade 10	0.20	0.40	0.00	1.00
Grade 11	0.19	0.39	0.00	1.00
Grade 12	0.16	0.37	0.00	1.00
Observations	18459			

Notes: This table presents summary statistics for the sample in wave I of AddHealth after dropping observations from schools (grades) with fewer than 20 (5) students.

4 Stylized Facts on Mental Health over the Life-Cycle

We begin by documenting the evolution and persistence of mental health over the life-cycle using the rich information from the AddHealth study. Mental health manifests itself early in life and stays persistent over the life-cycle. These stylized facts highlight the importance of studying the features of the school environment as determinants of mental health, thus motivating our subsequent analysis.

Evolution of Mental Health over the Life-Cycle. AddHealth data at wave I covers respondents from different ages ranging from 12 to 18 years. We lever this feature to investigate the evolution of mental health over time. Although there are several years between the data collections of different waves, there is a partial overlap in ages covered by different waves.

This allows us to aggregate age-specific mental health measures across waves.[9] To increase precision, we aggregate age groups into two-year bins and trace the evolution of mental health measured by CES-D scores from adolescence through mid-life.

The first panel of Figure 1 displays the average evolution of CES-D scores over time. We observe that mental health is deteriorating until age 20, i.e., during the time when respondents are still in school, and stabilizes afterwards. In the remaining panels of Figure 1, we differentiate the evolution of mental health by gender, ethnicity, parental education, and whether respondents were raised in a single-parent household. While there is cross-sectional variation in mental health with females, non-white people, and respondents with lower socioeconomic status who have higher CES-D scores and thus worse mental health, the evolution over the life-cycle is similar across subgroups.[10] In particular, we observe the same steep increase in CES-D scores until age 20 and a relatively flat pattern afterwards.

This observed increase in the prevalence of depressive symptoms over the adolescent period is consistent with a number of observations and hypotheses from the behavioral sciences. Thapar et al. (2012) summarize the literature and describe four broad mechanisms giving rise to the increase in mental illnesses during adolescence: (i) family and genetic factors (e.g., inherited risk factors), (ii) psychosocial risk factors (e.g., exposure to stressful life events), (iii) gene-environment interplay (e.g., genetic predisposition with increasing sensitivity to adversity), and (iv) brain development and hormonal changes with the onset of puberty (e.g., development of emotional regulation during adolescence creating stronger emotional responses; see also Ahmed, Bittencourt-Hewitt, and Sebastian, 2015).

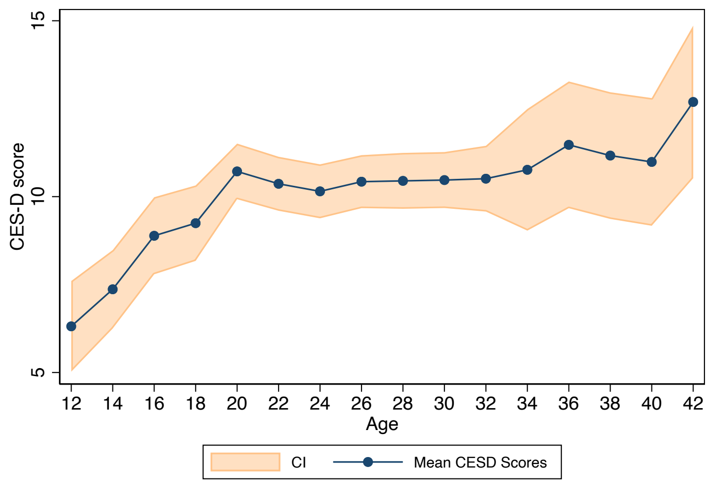
Adverse life events may have an important role in adolescent depression, because of heightened brain activity in the reward and danger-sensitive regions. Exposure to adversity does not necessarily lead to depression. However, there is considerable evidence in the literature that family adversity, bullying, peer rejection, and a wide range of additional possible stressors can prompt the onset of depression and have long-term consequences into adulthood (McCormick and Green, 2013; Thapar et al., 2012). Groups at-risk appear to be those who experience multiple adverse events, as well as along with girls, in particular, who appear to exhibit greater differences in brain activity, which further enhances their risk (Thapar et al., 2012). This is consistent with our observation that girls and groups likely to have experienced multiple adverse events (e.g., minorities and those with less educated parents) tend to have elevated depressive symptoms.

[9] To do so, we need to rescale CES-D scores in later waves, which only use a subset of the initial scale, as illustrated in Table A.1. See the discussion in Footnote 8 on how we perform the scaling. Moreover, we restrict the sample to those respondents whom we observe across waves I, III, IV, and V, to present the evolution and persistence of mental health based on a balanced panel.

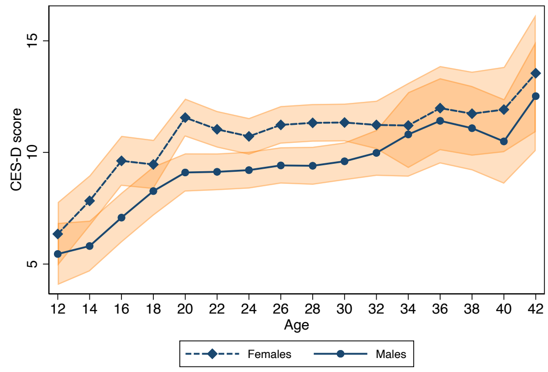
[10] In Appendix Table A.2, we quantify these cross-sectional differences with regressions relating CES-D scores in wave I to observable characteristics of students. Females have 0.25 SD higher CES-D scores than males, while differences between whites and non-whites, between children from college-educated and non-college-educated households, and between children raised in two-parent and single-parent families range between 6 and 10% of a standard deviation.

Figure 1. Evolution of Mental Health over the Life-Cycle

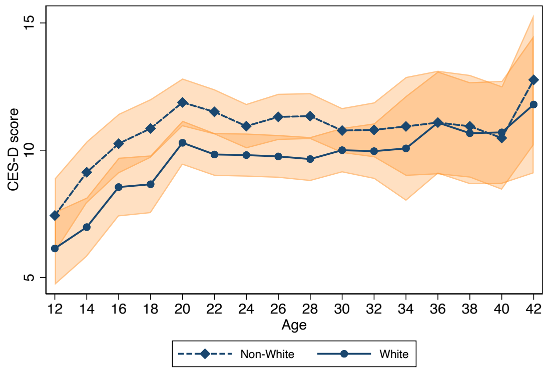
(a) All



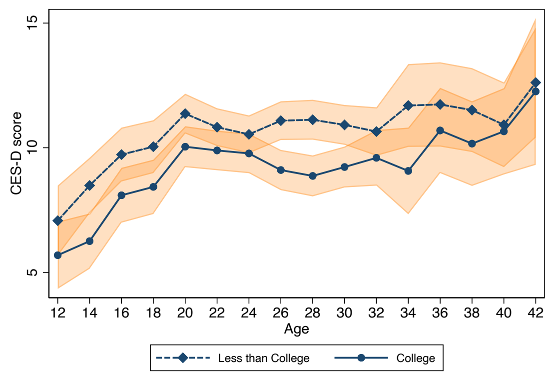
(b) Split by Gender



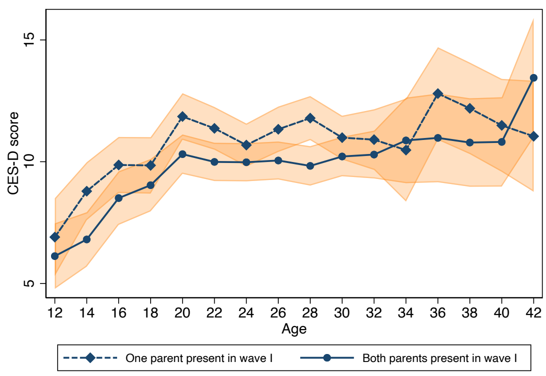
(c) Split by Ethnicity



(d) Split by Parental Education



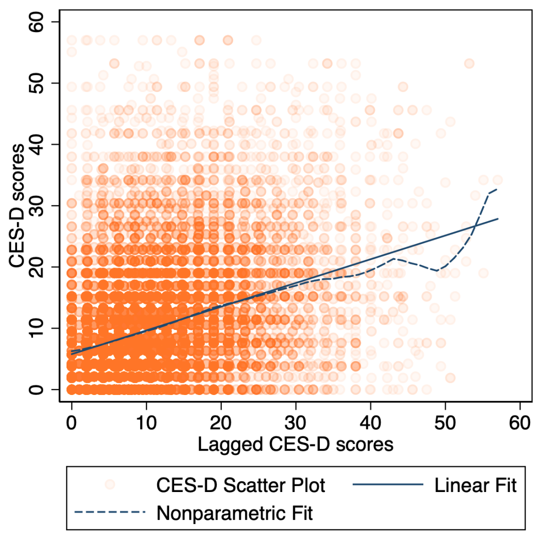
(e) Split by Single-parent Status



Notes: This figure presents mean CES-D scores at 2-year age bands after pooling over waves I, III, IV, and V, and controlling for survey wave effects. The shaded area indicates 90% confidence intervals.

Persistence of Mental Health over Time. While Figure 1 shows that *average* mental health remains relatively stable after the age of 20, it does not tell us about persistence on the individual level. We therefore provide further evidence on the persistence of mental health by studying the relation between CES-D scores in subsequent waves. Figure 2 presents the distribution of CES-D scores in waves I, III, IV, and V, plotted against the corresponding scores in the previous wave including linear and nonparametric fits. We observe a strong autocorrelation of 0.39 in CES-D scores across waves. To put this number into perspective, we compare this autocorrelation to test-retest statistics of CES-D scales. For instance, Roberts et al. (1990) report one-month test-retest correlations of 0.49 and 0.60 for boys and girls. Given that the lags between the AddHealth's waves are about seven to nine years, this persistence in mental health is remarkable.

Figure 2. Persistence in CES-D Scores over Time



Notes: This figure presents a scatter plot of CES-D and lagged CES-D scores, as well as linear and nonparametric fits. We pool across waves I, III, IV, and V (note: wave II is omitted because high-school leavers in wave I are not sampled during wave II), resulting in a time lag of 7 (wave I to III, and III to IV) and 9 years (wave IV to V).

The long-term persistence in mental health documented here is consistent with evidence in the psychological literature. Depressive symptoms during adolescence are linked with a high probability of depressive problems later in life (Thapar et al., 2012). However, to our knowledge, very few studies have provided evidence based on nationally representative longitudinal data. In a representative sample of a cohort in New Zealand followed from early to mid-life, Caspi et al. (2020) document that 59% of their sample experienced an onset of a mental health disorder by the close of adolescence. Moreover, those individuals who experienced an onset during adolescence also exhibited a greater number and variety of symptoms over time. Relatedly, Plana-Ripoll et al. (2019) report strong comorbidity over at least 15 years (i.e., an increased risk of developing further mental disorders after a first one) and Momen et al. (2020) extend these findings, highlighting an increase for subsequent medical conditions, independently of mental health disorders. Again, the strong persistence we observe is consistent with the limited evidence available.

Together, Figures 1 and 2 provide a first indication that the school environment may have a lasting effect on the mental health of students: Mental health seems to be particularly malleable during adolescence and displays a strong persistence over time.

5 Empirical Strategy

In order to test the predictions of the theoretical framework in Section 2, we aim at isolating a shock that may affect students' beliefs about their ability. Ideally, we want to lever a shock that provides information about an individual's relative ability only, holding everything else constant. We use students' ordinal ranks in their school cohort as such information shocks, which – conditional on own and peer ability – capture information about the relative standing within a cohort only. Before we illustrate our identification strategy, we describe how we construct our main variable of interest, the ordinal rank of students.

5.1 Constructing a Student's Ordinal Rank Measure

We construct students ordinal ranks based on an assessment of their cognitive ability, which is comparable across cohorts and schools. More specifically, we use the condensed version of the revised Peabody Picture Vocabulary Test (PPVT-R; Dunn and Dunn, 2007) that was administered as part of wave I and provides us with an objective, age-specific, and standardized measure of ability.

To construct a student's ordinal rank, we follow Murphy and Weinhardt (forthcoming) as well as Elsner and Isphording (2017, 2018). We first rank students based on their cognitive ability within their cohort by assigning them an absolute ability rank.[11] Due to differing school and cohort sizes, we subsequently normalize the absolute rank to an ordinal rank by dividing by the cohort size:

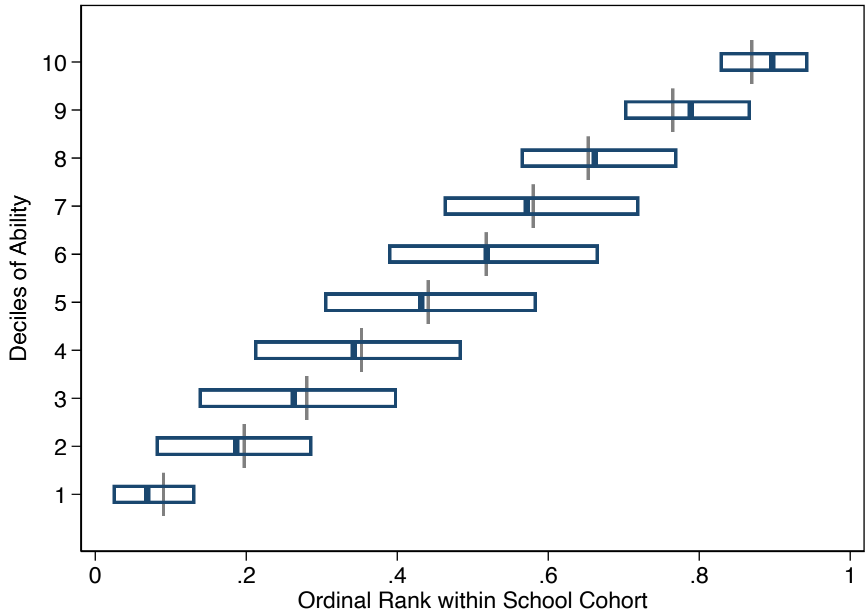
$$\text{ordinal rank} = \frac{\text{absolute rank} - 1}{\text{cohort size} - 1}. \tag{1}$$

This results in an ordinal rank which assigns the value 1 to the highest-ranked student and 0 to the lowest-ranked student. Figure 3 illustrates how this ordinal rank varies with a student's ability. The average ordinal rank increases in a student's ability. Yet, as we are interested in estimating the effect of a student's ordinal rank on their mental health holding ability constant, we need sufficient variation in ranks for a given ability level. Figure 3 provides some evidence that this is indeed the case – for each ability decile in the global ability distribution,

[11] We assign the student with the lowest ability the rank 1 and then increase the absolute rank. Thus, the higher a student's absolute rank, the higher their ability. To define the absolute rank, we count the number of peers with a lower ability, implying that if two students have the same ability they are assigned an equal rank. We relax this definition in robustness checks.

we observe sizable variation in a student's local rank – but we will revisit this question in the following section after formalizing our identification strategy.

Figure 3. Variation in Students' Ordinal Ranks by Ability Decile



Notes: This figure presents the variation in ranks for each ability decile. In particular, for each decile, boxes illustrate the 25th, 50th, and 75th percentiles of ordinal ranks, while gray lines indicate the mean ranks.

A potential confound for the interpretation of our rank measure based on AddHealth's Picture Vocabulary Test is that neither students nor teachers learn the results of this test.[12] Thus, the question remains how salient our rank measure is. We evaluate this by studying the relationship of ranks based on our ability measure and the students' self-assessment about their relative ability, as well as their desires and expectations for attending college. We report these results in Appendix Table A.4. Reassuringly, we find a strong and positive association between ability rank and self-assessed relative ability, mitigating the concern that ranks are not salient to students. Moreover, we observe that those students with a higher rank also have significantly higher expectations regarding their educational attainment.

Another concern is that ability was measured as part of wave I, and hence could be determined simultaneously with students' ranks. Yet, evidence in the literature indicates that cognitive ability is only malleable early in life and is considered as stable from age 10 onward (Jensen, 1998). At the time of AddHealth's wave I, when students were on average 15.6 years old, cognitive ability can therefore be seen as predetermined and unaffected by features of the school environment and the students' own or their parents' investments. In order to provide empirical evidence that ability seems to be fixed, we exploit that the ability test was

[12]Alternatively, we could have used a student's GPA to calculate ranks. Yet, this measure would have considerable limitations. First, GPA may be comparable within a school cohort, but comparisons across cohorts and schools may be difficult. Moreover, teachers have discretion about the grades of students potentially capturing confounding effects, and the students' GPA may be affected by classical peer effects.

administered again in wave III. Appendix Figure A.4 shows that the association between the two ability tests is near-perfect. Yet, one might be concerned with a potential spillover from ranks to ability. Hence, we also test for an effect of ordinal ranks in wave I on ability in wave III. As shown in columns (5) and (6) of Appendix Table A.4, when conditioning on wave I ability, which we do in all of our analyses, the effect of ranks on ability in wave III is essentially zero (increasing a student's rank by 1 SD increases ability in wave III by 0.004 SD). This suggests that cognitive ability is rather stable and pre-determined at the time of observing our sample.

Finally, we only observe a random sample of students in each school cohort, introducing additional sampling variation in our data. In Appendix E, we report results from a simulation study, showing that such sampling variation leads our estimates to be attenuated by approximately a third if we observe only 10% of the students in a cohort. In our sample, for only 3.4% of students do we observe less than 10% of their cohort, while on average 33.8% of all students in each school cohort are part of our sample.[13] Taken together, defining ordinal ranks using AddHealth's Picture Vocabulary Test yields a measure that is based on pre-determined characteristics, salient to students, and comparable across cohorts as well as schools.

5.2 Exploiting Within School-Cross Cohort Variation in the Cohort Composition

We aim to estimate the causal effect of a specific feature of the school environment: how does a student's ordinal rank in their cohort affect their mental health, holding both own and peer ability constant. Before we discuss our empirical strategy more formally, we want to provide some intuition for the identifying variation that we are exploiting. Consider a single school, in which we observe at least two cohorts. To identify a rank effect, we compare two students with the same ability, but different ranks within their respective cohorts. Figure 4 presents an example of this identifying variation. In this example, either the mean or the variance of ability differs across cohorts. This gives rise to different absolute ranks – and thus relative ranks – for students with the same ability of 7 (i.e., their absolute rank varies between 8 and 10 in our example).

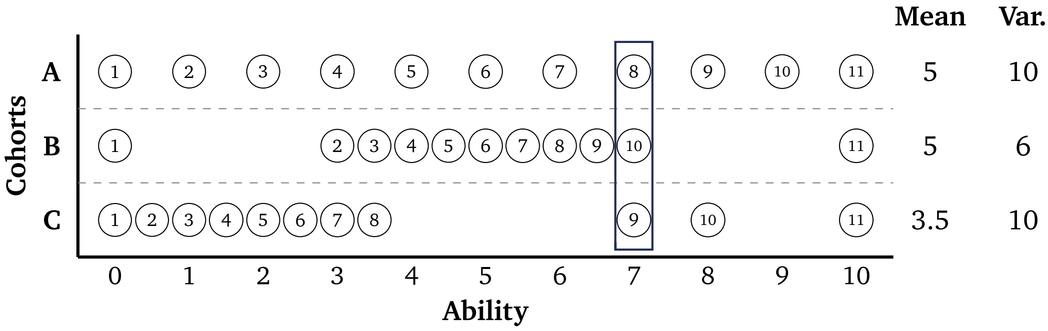
The identifying variation illustrated in Figure 4 describes our basic identification strategy. We follow Hoxby (2000a,b) and exploit the idiosyncratic variation in the ability distribution across cohorts within the same school. This motivates the following empirical specification:

$$y_{ics} = \alpha\, rank_{ics} + f\left(a_{ics}\right) + \mathbf{X}_i'\beta + \theta_{ics} + \epsilon_{ics}, \tag{2}$$

in which y_{ics} denotes the mental health of student i in cohort c and school s. $rank_{ics}$ is this student's ordinal rank within their cohort, as defined in equation (1), and $f\left(a_{ics}\right)$ denotes a flexible functional form of a student's own ability (in our application we use a fourth-

[13] For the majority of the schools, AddHealth samples about 17 students from each grade and gender for the main sample, independently of the size of the cohort. In addition, there are 16 saturated schools, in which all students were interviewed.

Figure 4. Illustrative Example of the Identifying Variation



Notes: This figure illustrates how variations in the ability distribution across cohorts allows us to identify rank effects. In these examples, we fixed the minimum and maximum of the ability distribution and allow either the mean or the variance of the ability distribution to differ across cohorts. Students are ranked according to their ability, as illustrated by the numbers in the circles. A comparison of cohort A and B shows that holding the mean ability constant can give rise to different ranks for individuals of the same ability. A comparison of cohorts A and C illustrates that a variation in mean ability, but constant variance in ability, can also can give rise to different ranks. Empirically, we will exploit any variation in the cohort composition, regardless of the source of variation.

order polynomial, but relax this in robustness checks). $\mathbf{X}_i$ corresponds to a vector of student characteristics which includes gender, age and age squared, indicators for race or ethnicity (Asian, Black, Hispanic, Other), indicators for their parents' highest degree (less than high school, high school/GED, some college, college degree, postgraduate degree), and an indicator for being raised in a single parent household. Finally, θ_{ics} denotes a set of fixed effects – at a bare minimum school (δ_s) and cohort fixed effects (γ_c) – to guide our identification as explained in the following. We cluster standard errors at the school level.

One obvious concern with equation (2) is that a student's ordinal rank is related to the average ability within the cohort. We aim to focus on the pure information shock and do not want our rank measure to be confounded by typical peer effects in ability. We therefore add the leave-one-out peer ability, $\overline{a}_{-ics}$ as an additional control variable, and in our baseline specification control for $\theta_{ics} = \lambda\overline{a}_{-ics} + \delta_s + \gamma_c$. In refinements, we add further linear-in-means peer effects and standard deviations in these peer characteristics to capture other dimensions and potential non-linearities in peer effects.

An additional concern is that parents may select their children's schools based on trends in the school-ability distribution (see, e.g., Hastings, Kane, and Staiger, 2009; Rothstein, 2006, for evidence that parents prefer sending their children to schools with high ability peers). This would potentially bias our results. In a second specification, we therefore add school-specific time trends to capture this potential source of bias. In this case, we identify the rank effect from variation in the ability distribution within schools and across grades after taking school-specific linear trends into account (i.e. $\theta_{ics} = \lambda\overline{a}_{-ics} + \delta_s + \gamma_c + c \times \delta_s$).

In a third and final specification, we go a step further and control for any heterogeneity of a cohort in a given school. We do this by introducing school-specific cohort fixed effects, i.e., $\theta_{ics} = \delta_s \times \gamma_c$, as in Murphy and Weinhardt (forthcoming). Using these school-by-grade fixed effects, we absorb any potential peer effects in terms of means, variances, or any higher moment. In this case, to identify rank effects, we rely on the variation of students' ranks within their cohort compared to cohorts in other schools after all observed and unobserved differences between school-specific cohorts are removed.

In order to identify the causal effect of ranks, α, the ordinal rank has to be as good as randomly assigned. More specifically, this means that we need to assume exogeneity of ranks conditional on a rich set of controls and fixed effects, that is,

$$E\left[\epsilon_{ics}|rank_{ics}, f\left(a_{ics}\right), \mathbf{X}_i, \theta_{ics}\right] = 0.$$

In essence, this assumption implies that ϵ_{ics} is uncorrelated with a student's ordinal rank conditional on her own ability, individual characteristics, and a set of cohort-level controls. In the first specification, we assume that these cohort-level controls are given by separate school and cohort fixed effects, as well as peer effects in student ability, and in the second, we additionally capture school-specific time trends. Using these and individual controls, we compare students in the same school and cohort, with similar peers, and with the same observable characteristics and ability, but who have different ranks.

Nonetheless, there might be other factors that potentially affect a student's mental health and rank that are unobservable to us. If such factors are present, this violates our exogeneity assumption and hence prevents us from estimating unbiased rank effects. Therefore, our third specification with school-specific cohort fixed effects absorbs all observable and unobservable differences between cohorts within and across schools. As mentioned above, we then identify rank effects from variations in ranks within school cohorts or, more specifically, from combinations of different shapes of the ability distribution across school cohorts and own ability that define ordinal ranks.

A natural question is how much variation is left in our rank variable after conditioning on our set of control variables and different fixed effects. We assess this variation in Appendix Table A.3. The standard variation in ranks without controls amounts to 0.28. However, since a student's rank and ability are positively correlated, as indicated by Figure 3, some part of the variation may be due to ability. Moreover, as our analysis will be focused on heterogeneous effects by ability decile, we need to ensure that there is sufficient variation in our variable of interest in each of the deciles.

To assess this condition, we calculate the residual variation in ranks after controlling for background characteristics and different sets of fixed effects and we compare this to the raw standard deviation in ranks. Appendix Table A.3 shows that the raw standard deviation in ranks by decile varies between 0.09 and 0.18. Conditioning on school and grade fixed effects and our set of baseline controls reduces this variation to 0.07-0.12. Using school-specific grade fixed effects reduces the variation slightly further to 0.07-0.11. Hence, our rich set of controls and fixed effects leaves at least 40% of the raw variation. Thus, there remains substantial residual variation in ordinal ranks to study their causal effect on mental health.

Finally, we perform balancing tests on our main treatment variable and other randomized peer variables to provide evidence that the peer composition across cohorts within schools is indeed consistent with quasi-random peer assignment. Each cell of Table 2 presents a regression of our treatment variable of interest – the ordinal ranks of students in their cohort – or another variable that should simultaneously be quasi-randomly assigned – peer ability and variation in peer ability – on pre-determined characteristics of students as well as a fourth-order polynomial in ability and one of the three sets of fixed effects θ_{ics}.[14] Consistent with quasi-random assignment of peers, we observe most characteristics are not related to our treatment variables. Only the indicator whether a student is white seems to be associated with a higher rank. Yet, given that this association does not hold for other quasi-randomly assigned peer variables, the number of tests performed is relatively high, and that the coefficient is small amounting to less than one percentile score, we interpret the balancing check as consistent with quasi-random assignment of peers.

Table 2. Balancing Tests

	Rank			Peer Ability (std.)			SD(Peer Ability) (std.)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.002	-0.001	-0.002	0.001	-0.002	0.000	0.009	0.003	-0.002
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.000)	(0.011)	(0.009)	(0.003)
White	-0.008***	-0.007**	-0.007**	0.003	-0.005	-0.003	0.008	0.003	-0.005
	(0.003)	(0.003)	(0.003)	(0.011)	(0.006)	(0.002)	(0.019)	(0.020)	(0.005)
College-educated Parents	-0.000	-0.000	0.000	0.003	0.003	-0.001	-0.003	-0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)	(0.001)	(0.012)	(0.011)	(0.003)
Raised by a Single Parent	-0.002	-0.002	-0.002	-0.003	0.001	0.000	-0.004	0.004	-0.006**
	(0.002)	(0.002)	(0.001)	(0.005)	(0.004)	(0.001)	(0.015)	(0.012)	(0.003)
Household income (1,000USD)	-0.000	-0.000*	-0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household receives food stamps	-0.004	-0.003	-0.003	0.000	0.001	0.000	-0.046	-0.048*	-0.002
	(0.002)	(0.002)	(0.002)	(0.006)	(0.005)	(0.001)	(0.029)	(0.025)	(0.005)
Household size	0.000	-0.000	-0.000	-0.003	-0.001	-0.000	-0.002	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.004)	(0.004)	(0.001)
First-born child	-0.000	0.000	0.002	0.006	0.005	-0.000	0.020*	0.018*	-0.001
	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.000)	(0.011)	(0.010)	(0.002)
Birth weight (ounces)	-0.000	-0.000	-0.000	0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ability	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School and Grade FE	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
School-specific trends	No	Yes	No	No	Yes	No	No	Yes	No
School × Grade FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses and clustered at the school level. Each cell presents a separate regression of the variable in the column header (rank, standardized peer ability, or standard variation in peer ability) on the variable indicated at the beginning of each row. All specifications include controls for a fourth-order polynomial in ability and fixed effects as indicated at the bottom of the table.

6 Results

How does a student's ordinal rank affect mental health? Our theoretical framework in Section 2 generates four predictions: First, we should observe that positive (negative) shocks,

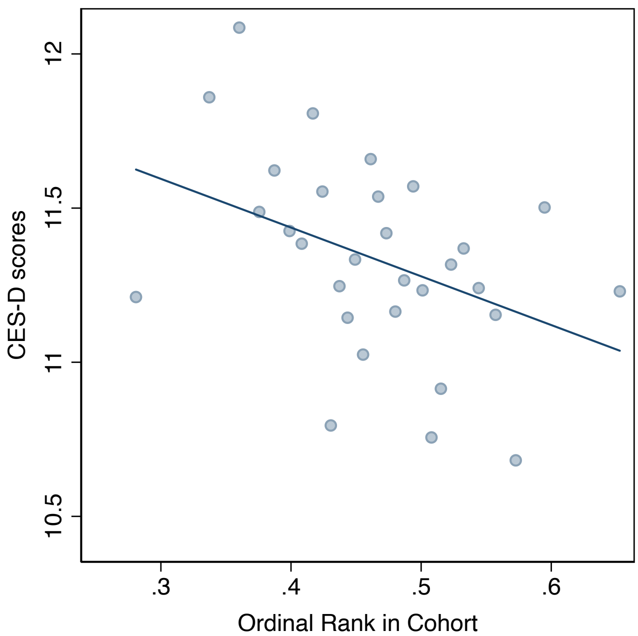
[14]Some information (e.g., information on household income) stem from parental questionnaires that are missing for some individuals. In general, we replace binary variables with zero and non-binary variables to the sample mean, and additionally control for an indicator equal to one if a variable is missing. The results do not change when presenting estimates on the subsample with non-missing information only.

in our application proxied by ability ranks in the school cohort, benefit (worsen) a student's mental health. Second, negative shocks have more pronounced consequences than positive ones. Third, rank effects are predicted to be stronger at the lower end of the ability distribution, where students are more likely to withdraw their study effort in response to negative shocks. Finally, the framework suggests that these effects are persistent over time. In the following, we will test these predictions.

6.1 Average Effect of Students' Ranks on Mental Health

We begin by studying the average effect of a student's rank on their mental health. More specifically, we relate a student's mental health measured by CES-D scores to their ordinal rank based on our main specification in equation (2) with standard errors clustered at the school level. We present our results in Table 3 and Figure 5. Based on our first empirical specification controlling for separate school as well as cohort fixed effects, and ability peer effects, column (1) shows that higher ranks reduce CES-D scores, i.e., they improve the students' mental health. Moving a student from the 25th percentile to the 75th improves their mental health by 0.8 CES-D points, increasing a student's rank by 1 SD (0.28 percentiles) yields a 0.06 SD improvement in mental health.

Figure 5. Average Effect of Ordinal Ranks on Mental Health



Notes: This figure presents the results from a regression of CES-D scores (lower scores corresponding to better mental health) on students' percentile ranks in their cohort (higher ranks correspond to higher relative ability) conditional on a fourth-order polynomial in own ability, gender, ethnicity, age and age-squared, parental education, and being raised by a single parent, average peer ability, as well as school and grade fixed effects as in column (1) of Table 3.

Table 3. Average Effect of Ordinal Ranks on Mental Health

	Mental Health (CES-D score)				
	(1)	(2)	(3)	(4)	(5)
A. Baseline Effects					
Rank	-1.58**	-1.58**	-1.55**	-1.57**	-1.67**
	(0.75)	(0.75)	(0.75)	(0.75)	(0.76)
Ability and Controls	Yes	Yes	Yes	Yes	Yes
Ability Peer Effect (mean)	Yes	Yes	Yes	Yes	No
Further Peer Effects (mean)	No	Yes	Yes	Yes	No
Ability Peer Effect (SD)	No	No	Yes	Yes	No
Further Peer Effects (SD)	No	No	Yes	Yes	No
School and Grade FEs	Yes	Yes	Yes	Yes	No
School-specific time trends	No	No	No	Yes	No
School $\times$ Grade FEs	No	No	No	No	Yes
Mean CES-D score	11.3	11.3	11.3	11.3	11.3
Observations	18459	18459	18459	18459	18459
R^2	0.108	0.108	0.109	0.117	0.127
B. Standardized Effects					
Rank (std.)	-0.06**	-0.06**	-0.06**	-0.06**	-0.06**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
C. Role of Unobservables					
Oster's δ ($R^2_{max} = 1.3R^2$)	-1.01	-1.18	-1.36	-2.65	-1.74

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are in parentheses are clustered at the school level. In Panel A, each coefficient presents a regression of CES-D scores (lower scores corresponding to better mental health) on an individual's percentile rank at the school-level based on equation (2). We include a fourth-order polynomial in own ability, gender, ethnicity, age and age-squared, parental education, and being raised by a single parent as control variables. Peer ability includes the leave-one-out mean and standard deviation of peer ability, peer controls comprises additional peer effect terms in gender, ethnicity, and parental education. We present standardized effects of our main effect in Panel B. Panel C presents the results from a sensitivity analysis based on Oster (2019) and quantifies how severe selection based on unobservables would need to be for zero rank effects (Oster, 2019). To calculate δ, we follow Oster (2019) and assume a maximum R^2_{max} of 1.3 times the actual R^2.

How large are these effects relative to differences in socioeconomic differences in mental health? In Appendix Table A.2, we present associations of several background characteristics on (standardized) CES-D scores. The estimated effect size is similar to the difference in mental health of children from college-educated and non-college-educated households, about two thirds of the difference between white and non-white students, or the difference between students raised by a single parent and those raised by both parents. Another point of comparison can be based on a meta-analysis of several positive psychology interventions (Bolier et al., 2013): effect sizes of these interventions yield effects of 0.20-0.34 SD on outcomes such as psychological well-being, depression, and subjective well-being. Given that these are targeted interventions, we consider the estimated effects of ordinal ranks as large. A fact that is all the more striking as it results from natural variation in the ability composition of school cohorts, conditions on a rich set of demographic characteristics, and compares students with the same ability and the same average cohort ability who only happen to be in cohorts in which they have different ranks.

In the remaining columns of Table 3, we enrich our baseline specification to investigate the robustness of our main results. In particular, allowing for peer effects in ability only is potentially restrictive. For instance, the literature on peer effects has identified a range of different peer characteristics that causally affect students' performance and thereby may also affect their mental health. Examples include the share of females, minorities, or students with high socioeconomic status (Cools, Fernández, and Patacchini, 2019; Hoxby, 2000a; Lavy and Schlosser, 2011). We add these additional peer effect terms in column (2). Furthermore, in column (3), we also add controls for the standard deviation in peer ability, which has been shown to affect school performance (e.g., Tincani, 2017), and other peer characteristics capturing potential non-linear peer effects. Our estimates show that the rank effect is robust to the inclusion of these additional peer effects and varies only slightly.

Our identification is based on quasi-random variation in peer ability across cohorts in a given school. Yet, if parents select schools for their children based on trends in the ability distribution, this might bias our results. To account for such factors that change within a school over time, we further include school-specific time trends in column (4), which neither change the size nor the statistical significance of our results.

In column (5), we adopt an even stricter empirical specification using grade-by-school fixed effects and thus not only account for linear trends in the school-specific ability composition over time, but for any trend. Additionally, this set of fixed effects accounts for all observed and unobserved peer effects and exploits individual-level variation within school-cohorts to identify the effect of ordinal ranks on the mental health status of students. The coefficient of interest slightly increases in magnitude.

Taken together, the results from Table 3 document that the information shock from being ranked high or low exerts a causal effect on students' mental health measured by CES-D scores in line with the central prediction of our model. Decreasing (increasing) a student's rank by 1 SD causes an approximately 0.06 SD decrease (increase) in a their mental health, comparable in magnitude to effects of ranks on test scores (0.08 SD; Murphy and Weinhardt, forthcoming) and to the raw mental health difference between children with college and non-college-educated parents.

6.2 Robustness Checks

In this section, we report a series of additional analyses to probe the robustness of our finding.

Nonlinearity in ability. In our main specification, we adopt a fourth-order polynomial in Peabody scores to take the relation of mental health and ability into account. Yet, one might be worried that this arbitrary choice drives our results. In Appendix Table B.1, we therefore examine different polynomials up to a sixth order. We find that using linear or quadratic controls in ability increases our estimated coefficient on ordinal ranks and thus would strengthen our main result. The rank estimates stabilize for higher-order polynomials in ability. Moreover, when estimating a specification in which we control non-parametrically for different ability levels by including fixed effects for each level of the ability score, our results remain unchanged.

In a second set of specifications, we use a data-driven approach to select the (ability) control variables by employing a post-double selection (PDS) Lasso (Belloni, Chernozhukov, and Hansen, 2014). The PDS Lasso penalizes control variables, but allows valid inference on non-penalized treatment variables. We perform two such specifications, one in which we allow penalization terms of an eight-order polynomial only, and one in which we in addition allow for penalization of the set of baseline control variables (e.g., gender, race indicators, age). Both specifications penalize higher-order ability terms, leaving only a second-degree polynomial or a linear trend in ability, which suggests a relatively linear relationship of CES-D scores and ability, as also illustrated in Figure A.3. More importantly, however, the estimated effects of ordinal ranks are unaffected and, if anything, become more pronounced.

Definition of peer groups. We defined peer groups based on all students in a given cohort. Yet, evidence exists that students form friendships with similar peers (i.e., that friendship networks exhibit homophily; see Graham, 2015; McPherson, Smith-Lovin, and Cook, 2001, for overviews over the literature) and that they systematically select their relevant peers from larger peer groups (Kiessling, Radbruch, and Schaube, 2020a). This raises the question whether the rank effects should be defined at a more local level. Hence, we explore how our results change when we allow for differential rank effects by finer subgroups. More specifically, we enrich our main specification and add a second rank calculated (i) within grade and gender, (ii) within grade and race, or (iii) within grade, gender, and race. Appendix Figure B.1 presents the results from these specifications. As can be seen from the Figure, the baseline effect remains similar across specifications and the additional ranks based on different peer group definitions have small and insignificant effects. Hence, calculating ranks within grades seems appropriate, and this suggests that, in our setting, the effects seem to stem from comparisons to all peers in a cohort rather than from a specific subgroup.

Definition of ranks. In our definition of ranks, we follow Elsner and Isphording (2017, 2018) and calculate absolute ranks based on the number of peers with a strictly lower ability. Yet, other definitions are conceivable with implications for the assignments of ranks for those students who are in the same school cohort and have the same measured ability.[15] For instance, we could have assigned absolute ranks based on the number of peers with a lower or equal ability rather than a strictly lower ability. Alternatively, we could assign the mean of both methods to get at an average rank in case of ties. In Appendix Table B.2, we compare these

[15] In fact, for each school-grade combination, we observe an average of 3.2 students (median: 2) with the same ability and thus the same rank.

different definitions of ranks and find that our estimates are robust to the precise definition of ranks.

Role of unobservables. Our identification strategy assumes that a student's rank is exogenous conditional on own ability and school and cohort fixed effects. It is reassuring that our findings remain nearly unaffected once we control for additional potential confounds and if we adopt a stricter specification using school-specific cohort fixed effects. A more formal approach to test for the role of unobservables is to ask how severe selection based on unobservables would have to be to drive down the estimated rank effects to zero. In order to quantify this, we follow Oster (2019) and calculate δ, a measure for the degree of selection based on unobservables relative to observable characteristics. If δ is larger than one, this indicates that selection on unobservables would need to be at least as important as selection based on observables to explain our effects. As shown in Panel C of Table 3, the magnitude of δ is larger than one in all specifications. Since we control for arguably the most important factors that could bias students' ordinal ranks and that may affect mental health through differences in the cohort composition, these numbers imply that we would have to be missing highly relevant variables in order for unobservables to give rise to our estimated effects. Hence, we conclude that unobservables are unlikely to drive our estimated rank effects.

Sorting based on ranks. One concern is that parents may select schools based on the rank that their children would have, violating our assumption that the rank is as good as randomly assigned. Yet, there is plenty of evidence that parents prefer sending their children to schools with high ability peers (Burgess et al., 2015; Hastings, Kane, and Staiger, 2009; Jackson et al., 2019; Rothstein, 2006). If this is the case, then this is not consistent with positive sorting based on ranks, as ranks and peer ability are inversely related.

Moreover, several patterns in our data suggest that sorting based on ranks is a minor concern. First, in Appendix Figure B.2 we show that if parents sort into specific schools based on average peer ability, there is high uncertainty about the resulting rank of a student with a given ability. This implies that sorting on average ability and rank is rather difficult. In addition, as we will show in Section 6.5.2, the size of the rank effect does not differ by average school ability and is not driven by a specific subset of students or schools having certain characteristics.

Second, we assume that conditional on school and grade (school-by-grade) fixed effects as well as our baseline set of controls, the variation in ranks is as good as random. Yet, as shown in our balancing checks in Table 2, neither ranks nor other peer characteristics seem to be systematically related to ranks, average peer ability or the variation in average peer ability, indicating that sorting based on ranks or other peer characteristics is unlikely to explain our results.

Simulations to assess the role of measurement error. The AddHealth data have several sources of classical and non-classical measurement error. First, only a random subset of all students in each school is sampled introducing potential biases in our main variable as we observe only a fraction of the cohort. Second, our ability measure may suffer from measurement error that translates into a mis-measured rank. Third, although our analysis above suggests that unobservables are unlikely to drive our results, they potentially distort our estimates if there are omitted variables correlated with ability. Fourth, there may be sorting into different classrooms based on ability within school cohorts, which we cannot observe. This would imply that we calculate the students' ranks based on incorrect reference groups. Fifth, and finally, CES-D scores are aggregated from a small number of items that are scored on a scale from 0 to 3 rather than on a continuous scale.

We assess these concerns using a series of Monte Carlo simulations reported in Appendix E. We find that random sampling of students within schools, measurement error in our ability measure, and omitted variables mildly attenuate our estimates. This implies that we would underestimate the true effects of ranks on mental health.

The issue is more complex when we have unobserved sorting into classes within school cohorts. As long as the sorting is not too severe, i.e., is not strongly correlated with ability, we underestimate the true effect. Yet, once sorting within cohorts becomes sufficiently strong, this can yield estimates with the wrong sign. To test whether this might be a source of bias, we estimate heterogeneous effects based on the school administrators' reports of the use of tracking within cohorts. As we show in Section 6.5.2, we find that the effects in schools with tracking are only slightly attenuated and remain statistically significant. We therefore think that it is unlikely to have sufficient unobserved tracking that biases our results. Finally, measurement error from having limited variation in our outcome yields inefficient estimates, i.e., slightly larger standard errors than with continuous measures of mental health, but does not bias our estimates.

Taken together, these simulations suggest that various forms of measurement error lead to – depending on the form of measurement error – mild to moderate attenuation of our estimates. This implies that we are likely to underestimate the true causal effect.

6.3 Exploring Asymmetries in Shocks

We documented large effects of ordinal ranks on mental health. Yet, not all shocks are similar. In particular, Prediction 2 suggests that once students experience a negative shocks, their mental health deteriorates and is more likely to remain in a poor condition. We now want to provide more evidence on the asymmetry of these effects. If our conjecture is right, we should observe that any effect is more pronounced for negative rather than positive shocks.

In the literature, there is some evidence from laboratory experiments on asymmetric updating and information avoidance after negative signals. While some papers find support for the so-called "good news-bad news" effect (e.g., Eil and Rao, 2011; Möbius et al., 2014), in which people react to good news about themselves, but neglect negative signals, others find evidence of asymmetric updating in self-relevant domains along the lines of our conjecture (Ertac, 2011). Focusing on interactions of belief updating and mental health, correlational evidence (Armstrong and Olatunji, 2012; Gotlib et al., 2004; Korn et al., 2014) suggests that depressed individuals have attentional biases for negative, but not positive, information. We thus expect that negative signals, i.e., having a rank that is lower than one might expect, leads to stronger responses in mental health than positive shocks, as Prediction 2 suggests.

In order to differentiate between positive and negative shocks, we calculate the expected rank of students, independently of the ability composition of their school cohort.[16] To do this, we calculate rank measures similar to equation (1), but consider students in a given cohort across all schools. In other words, we calculate individual i's rank, $rank_{ic}$, among all students in a given cohort c, i.e., independently of their school s rather than within a school as in the case of $rank_{ics}$ above. We define student i receiving a negative shock if their rank in their school cohort, $rank_{ics}$ is lower than the rank among all students in a given cohort, $rank_{ic}$:

$$\text{Negative shock} = \mathbb{1}\left\{rank_{ics} < rank_{ic}\right\}. \tag{3}$$

We then extend equation (2) by adding an indicator for negative shocks as well as the interaction of negative shocks and ranks. This allows us to study whether negative shocks differentially affect the mental health of students compared to positive shocks.

The idea behind this is as follows: Consider two students with identical ability. One of them is randomly assigned to a better cohort, where they have a lower rank, whereas the other has a worse cohort and correspondingly a higher rank. Importantly, the distance in their local rank in their respective cohort from the global rank across all cohorts is the same for both individuals. We now investigate whether the effects of (local) ranks are more pronounced for those receiving negative rather than positive shocks.[17]

In Table 4, we study asymmetric responses using our definition of negative shocks from equation (3). Column (1) replicates our baseline result of the first column of Table 3 that ranks significantly reduce CES-D scores. We then study the causal effect of receiving a negative shock on mental health, while abstracting from rank effects. Column (2) shows that negative shocks increase CES-D scores by 0.36 points, corresponding to 0.05 standard deviations. In other words, negative shocks are detrimental to mental health.

Column (3) explores the interaction of ranks and negative shocks by regressing CES-D scores on ranks, an indicator for negative shocks, as well as their interaction. We find that once we account for ranks, negative shocks are more pronounced and increase CES-D scores by 0.62 (0.08 SD). Moreover, rank effects are approximately twice as large for students receiving negative shocks compared to those receiving positive shocks and similar

[16]If students have unbiased beliefs to begin with, this global rank would correspond to the prior in our theoretical framework.

[17]Obviously, the definition of negative and positive shocks can just be an approximation as we observe noisy measurements of both a student's ability as well as the presence of a positive or negative shock. Ex ante, the direction of the bias is not clear. In Appendix E, we therefore study the role of measurement error in a specification with ranks, negative shocks, and their interaction using Monte Carlo simulations (see Simulation F). We find that the rank effect corresponding to positive shocks overestimates the true effect in the presence of small amounts of measurement error, but underestimates for larger measurement errors. More importantly, however, our coefficients of interest – on negative shocks and the interaction of ranks and negative shocks – are attenuated in presence of measurement error, implying that we consistently underestimate the magnitude of these coefficients.

Table 4. Asymmetric Effects of Ordinal Ranks

	Mental Health (CES-D score)		
	(1)	(2)	(3)
Rank	-1.58**		-0.73
	(0.75)		(0.91)
Negative Shock		0.36**	0.62**
		(0.17)	(0.29)
Rank × Negative Shock			-0.76
			(0.47)
Ability	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes
Peer Ability	Yes	Yes	Yes
School and Grade FEs	Yes	Yes	Yes
Mean CES-D score	11.3	11.3	11.3
Observations	18459	18459	18459
R^2	0.108	0.108	0.108

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses and clustered at the school level. We include all base controls, peer ability, school and grade fixed effects, as in our preferred baseline specification of column (1) in Table 3.

to our baseline estimate, although the difference between positive and negative shocks is not significant at conventional levels ($p = 0.11$).

In line with Ertac (2011) and evidence from the psychological literature, but contrasting with Eil and Rao (2011) and Möbius et al. (2014), we find that students update more strongly in case of negative shocks.[18] The asymmetry in effects documented here implies that rank effects on mental health do not seem to stem from positive shocks of unexpectedly being ranked highly, but rather from negative shocks. Note that many social comparison mechanisms posit that having a high rank may open up better opportunities (e.g., having access to better colleges) setting individuals on different trajectories. If this explained our results, we would expect stronger effects for positive shocks, which is the opposite of what we find.[19]

6.4 Different Facets of Mental Health

In our analysis, we use the CES-D score based on the sum of the single items as it is commonly used in the literature. While all of the items are related to depressive symptoms, they cover

[18]Eil and Rao (2011) also show that subjects in their experiment have an aversion to new information after receiving a negative signal, which is in line with what we expect, despite opposite results on asymmetric information processing.

[19]That being said, we think that these social comparison mechanisms are likely important for outcomes other than mental health. In fact, when studying the long-run outcomes such as educational attainment in Section 7.1, we find positive effects of ranks both at the lower as well as the upper end of the ability distribution for some outcomes. This indicates that mental health is one of potentially several mechanisms affecting economic long-run outcomes.

different facets. In order to shed more light onto which facet is driving our results, we perform a principal component analysis on the items. We then apply a Varimax rotation and predict factor scores. This results in four distinct factors corresponding to (i) loneliness, (ii) lack of positive attitudes, (iii) lack of motivation, and (iv) external factors.[20]

Table 5 presents regressions of these different facets on the ordinal rank of students. We find that our results are mainly driven by effects on factors capturing a lack of positive attitudes as well as a lack of motivation rather than loneliness or external factors. These findings support our previous observation. We found that negative shocks drive the rank effect. If the reason for this is disappointment at being ranked lower than expected, then rank should affect the general mood and motivation of students rather than social exclusion and other external factors, which is indeed what we observe.

Table 5. Different Facets of Mental Health

	Mental Health (Principal Components; std.)			
	(1) Loneliness	(2) Lack of pos. attitude	(3) Lack of motivation	(4) External factors
Rank	-0.07	-0.21**	-0.16*	0.04
	(0.09)	(0.09)	(0.09)	(0.10)
Ability and Controls	Yes	Yes	Yes	Yes
Ability Peer Effect (mean)	Yes	Yes	Yes	Yes
School and Grade FEs	Yes	Yes	Yes	Yes
N	18411	18411	18411	18411
R^2	0.078	0.082	0.028	0.028

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses and clustered at the school level. The outcome is a standardized factor (with zero mean and a standard deviation of 1) from a principal component analysis of all 19 items of the CES-D scale. We include all base controls, peer ability, school and grade fixed effects as in our preferred baseline specification of column (1) in Table 3 and hence can compare the results to the standardized average effect in Panel B, column (1) of Table 3.

6.5 Heterogeneous Rank Effects

We now explore heterogeneities in two steps. First, our theoretical framework suggests that shocks have more pronounced effects on students at the lower end of the distribution (cf. Prediction 3). To test this prediction, we study whether effects of ordinal ranks differ by

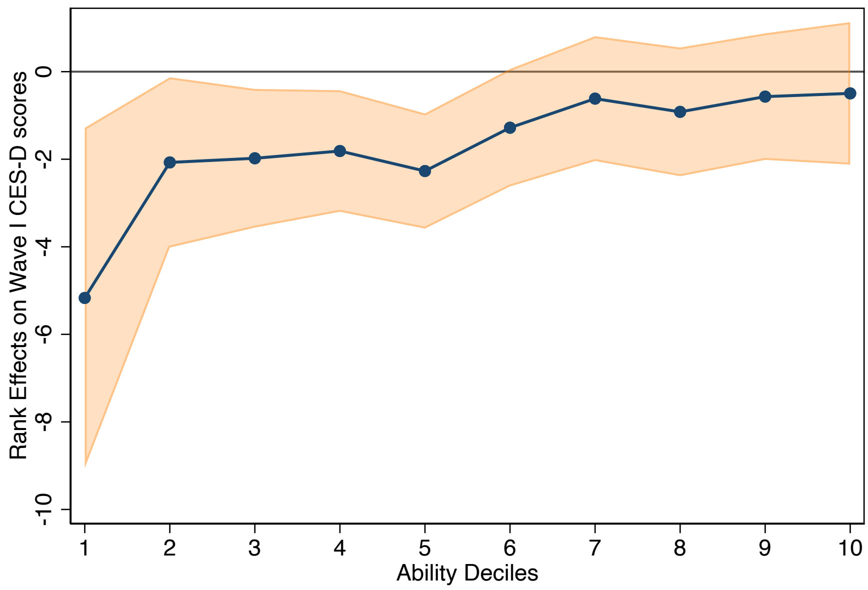
[20]A scree plot in Appendix Figure C.1 shows that four factors have eigenvalues larger than one. The rotated factor loadings of the items are presented in Appendix Table C.1. We assign names to these four factors based on factor loadings larger than 0.6. More specifically, the first factor, loneliness, mainly loads on items 3 (*"You felt that you could not shake off the blues"*), 6 (*"You felt depressed"*), 13 (*"You felt lonely"*), and 16 (*"You felt sad"*); the second factor captures the lack of positive attitudes with items 4 (*"You felt you were just as good as other people"*), 8 (*"You felt hopeful about the future"*), 11 (*"You were happy"*), and 15 (*"You enjoyed life"*); the third factor, lack of motivation, comprises items 5 (*"You had trouble keeping your mind on what you were doing"*), 7 (*"You felt that you were too tired to do things"*), and 18 (*"It was hard to get started doing things"*); and the fourth factor, external factors, consists of items 14 (*"People were unfriendly to you"*) and 17 (*"You felt that people disliked you"*).

ability decile. Second, we aim to provide a more comprehensive picture by considering heterogeneities in other observable characteristics. To do this, we focus on characteristics that may be of interest when designing policies exploiting the effects of ranks.

6.5.1 Heterogeneity by Ability

Prediction 3 suggests that the shocks to students at the lower end of the distribution are larger than for higher-ability students. This results because, on the one hand, a diminished perception of ability decreases study time, which subsequently amplifies the consequences of exogenous shocks, and, on the other, this partly stems from low-ability students holding beliefs closer to the threshold, where they withdraw their study effort. If this prediction is correct, we should observe stronger (weaker) rank effects for lower-ability (higher-ability) quantiles. We therefore enrich our main specification given in equation (2) by interacting the rank with indicators for each ability decile.

Figure 6. Effects of Ordinal Ranks by Ability Decile



Notes: This figure presents the effect of ordinal ranks by ability decile. We estimate the effects using enriching specification (2) and interact a student's rank with indicators for ability deciles. The shaded area indicates 90% confidence intervals clustered at the school level.

Figure 6 displays the results of this analysis graphically, while Table 6 presents the corresponding regression estimates. We indeed find that rank effects are more pronounced at the lower end of the distribution. In particular, the ordinal rank reduces the CES-D score by 5.2 points when moving a student from the bottom to the top rank, which corresponds to 0.70 SD. This effect amounts to three times the average effect and would suffice to move a student diagnosed with a moderate depression according to a threshold of 16 (Radloff, 1977) to the average CES-D score of 11.3 in our sample. While the point estimates are negative for all deciles, the estimated effects slowly fade out and are rather small and not significant at the top end of the distribution (coefficient of -0.50 with a p-value of 0.51 for the tenth decile). These results are therefore consistent with Prediction 3 of our theoretical framework that effects should be more pronounced for low-ability students.

Table 6. Effects of Ordinal Ranks by Ability Decile

	Mental Health (CES-D score) by Decile									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Rank	-5.17**	-2.07*	-1.98**	-1.81**	-2.27***	-1.28	-0.62	-0.92	-0.57	-0.50
	(2.35)	(1.17)	(0.96)	(0.84)	(0.79)	(0.81)	(0.86)	(0.89)	(0.87)	(0.98)

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses and clustered at the school level. We include all base controls, peer ability, school and grade fixed effects, as in our preferred baseline specification presented in column (1) of Table 3.

6.5.2 Further Heterogeneities by Other Individual and School Characteristics

Although our theoretical framework is not aimed at providing predictions for specific subsamples, these are nevertheless important for policy-makers interested in targeting policies. Furthermore, studying heterogeneities over specific subsamples contributes to a better understanding of rank effects on mental health.

We focus on two groups of heterogeneities – based on individual characteristics of students and based on school and cohort characteristics.[21] A priori it is not clear for which subsamples we should observe stronger effects. While previous research has shown that females may be more responsive to features of the environment (Croson and Gneezy, 2009), while males are more likely to enter competitions which have rankings as inherent characteristics (Niederle and Vesterlund, 2007). Similarly, individuals from low socioeconomic status may be more stressed by social rank concerns (Hackman, Farah, and Meaney, 2010), but students with affluent parents could also be more receptive to ranks as part of the competition for colleges.

In further checks, we focus on margins that are subject to frequent education policy debates. First, following the active debates about the consequences of tracking regimes (e.g., Duflo, Dupas, and Kremer, 2011; Garlick, 2018), we study whether tracking moderates shocks on mental health. Second, given the existing evidence on parental preferences for high achievement peers (e.g., Hastings, Kane, and Staiger, 2009; Jackson et al., 2019), we investigate whether heterogeneous effects by average school ability exist. Third, smaller classrooms can have beneficial effects on educational outcomes (Angrist and Lavy, 1999; Hoxby, 2000b; Krueger, 1999; Krueger and Whitmore, 2001), but at the same time increase the salience of students' ranks within classes. Lastly, we look at the number of teachers per student. While having more teachers per student potentially frees some teachers' time to counsel students, a

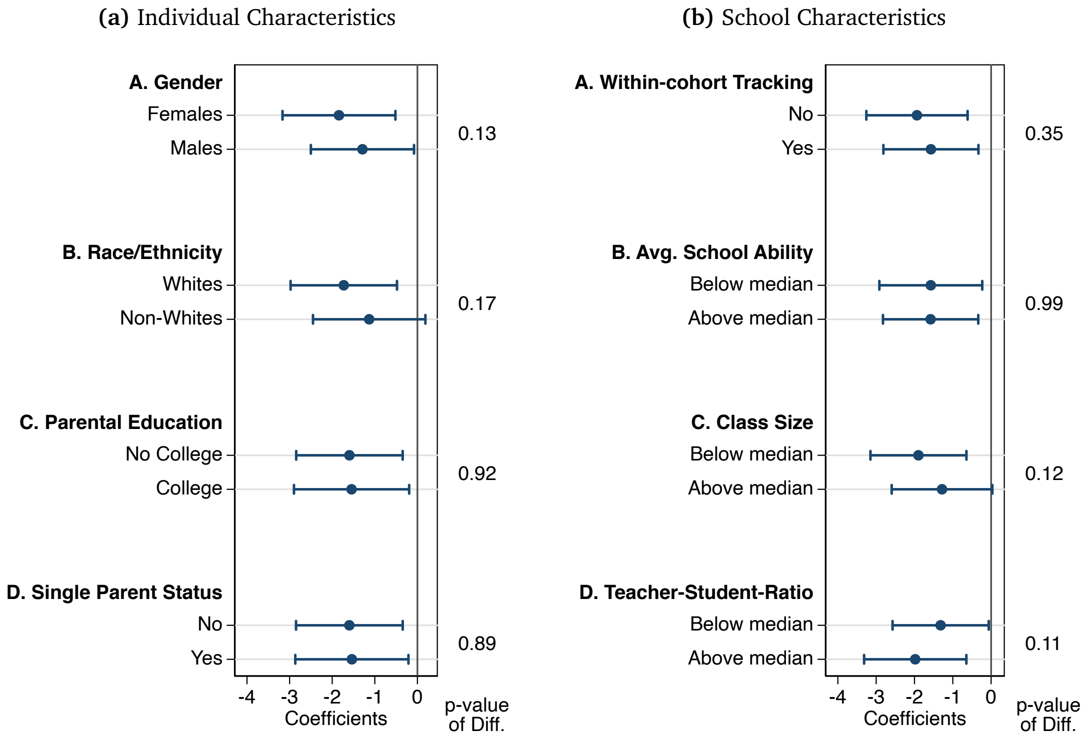
[21]We focus on heterogeneities of the average rank effect. While it would be interesting to study differential patterns in the heterogeneities by ability documented in the previous subsection, we lack power to do this.

lower ratio could suggest that the time a particular teacher teaches a given student increases, which could mediate negative consequences of shocks.

Figure 7 presents the results from our heterogeneity analyses. In Figure 7a, we only observe limited heterogeneity with respect to sociodemographic characteristics. The consequences of ranks, therefore, seem to affect the mental health of all students rather equally.

Turning to school characteristics in Figure 7b, we find homogeneous effects by average peer quality (Panel A), the presence of tracking (Panel B), or smaller classes (Panel C). Although these margins are associated with better educational outcomes, they do not lead to different rank effects. Surprisingly, we find that having more teachers per student is related to larger rank effects (Panel D). This latter finding is consistent with closer student-teacher relationships from lower teacher turnover in a given classroom, but we lack additional data to bolster this claim.

Figure 7. Heterogeneity by Individual and School Characteristics



Notes: This figure presents heterogeneous effects of different subgroups on CES-D scores including 90% confidence intervals clustered at the school level. We interact the rank variable with indicators of the respective variables. Figure 7a presents the heterogeneity with respect to individual characteristics, whereas Figure 7b presents corresponding results for school/cohort characteristics.

7 Persistence of Rank Effects

We have established that the ordinal rank exerts a causal effect on the mental health of students, and this effect is more pronounced for low-ability students. We now want to explore the dynamics of these effects. To derive a prediction about expected patterns, we note two points. First, there is a significant association between ability and mental health (see Appendix Table A.2 and Appendix Figure A.3). Second, students at the lower end of the distribution experience stronger effects and they have a higher risk of becoming depressed as a result of negative shocks. Following our theoretical framework and evidence from the psychological and neuroscience literature (e.g., Holtzheimer and Mayberg, 2011), we think of depressions as absorbing states. If this is the case, we should observe that our effects are persistent for those at-risk students.

To explore this persistence, we use CES-D scores elicited in each of the following waves. We can study short-term persistence using wave II, medium-term persistence using wave III, and long-term persistence using wave IV.[22] Similar to Section 6.5.1, we estimate our main specification (2), but study the heterogeneous effects of ordinal ranks in wave I by ability decile on measures of mental health in later waves.

Unfortunately, not all waves conducted the 19-item version of the CES-D, but adopted a short version comprising a subset of the original items in waves III and IV. To compare our estimates from all waves to the baseline, we scale the mental health measures from the short scales by $^{19}/_{9}$ (Wave III) and $^{19}/_{10}$ (Wave IV) to correspond to the same range from 0 to 57 as the full scale in wave I.[23]

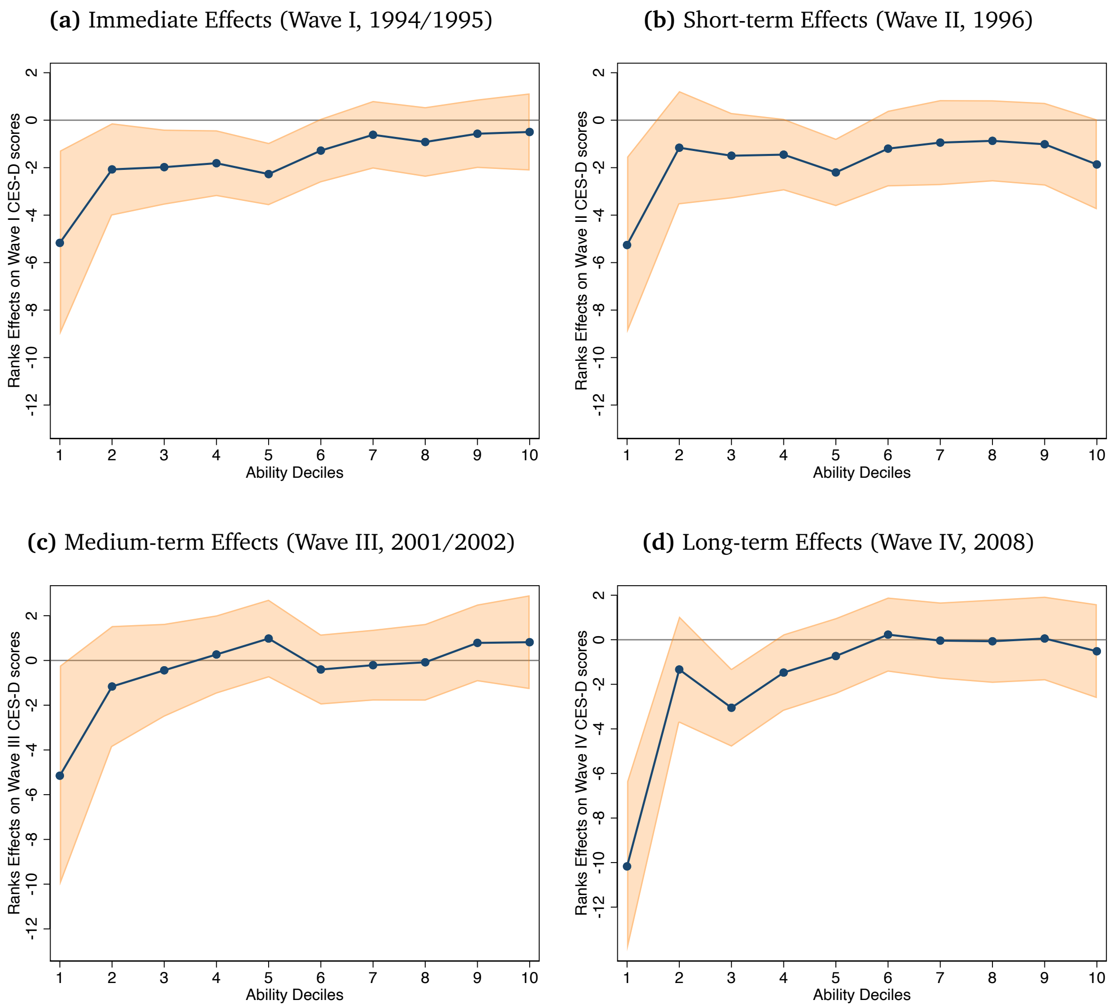
Figure 8 shows that the general pattern persists over time: Across all waves, the effect of ordinal ranks is significant and pronounced at the bottom of the ability distribution and insignificant as well as smaller in magnitude for higher ability deciles. Table 7 quantifies these effects. Panel A replicates the estimates of Section 6.5.1, while Panels B through D consider the short-, medium-, and long-term effects of ordinal ranks in wave I after one, seven, and fourteen years. We find that the the significant effects for the lowest ability decile persist across waves I to IV and amount to -5.2 to -10.2 CES-D points, and fade out for higher ability deciles. This pattern is strikingly similar from wave I, when students are 12-18 years old, to wave IV, when those students are adults of 26-32 years.

Wave V (2016-2018) results. In principle, we could also lever data from wave V, conducted about 23 years after baseline in 2016-2018. Performing the same analysis as for the previous waves, we do not observe any effects as shown in Appendix Table D.1. While this could suggest that the effects fade out 20 years after the initial shock, we think this is rather due to several problems with the CES-D measure in wave V. First, we face selective attrition in wave V. While neither the outcome (CES-D scores) nor the treatment variables (ordinal rank and an

[22]While wave II was conducted one year after wave I, in 1996, respondents were interviewed for wave III approximately seven years after the initial interview (in 2001/2002). Finally, wave IV was conducted 14 years after wave I, when respondents were adults aged 26-32 years.

[23]As shown in Appendix Figure A.2, the short and long versions of the CES-D are highly correlated in wave I. We therefore use scaled version to compare our results to the baseline effects documented in Section 6.5.1. Scales based on fewer items reduce the efficiency of our estimates. Simulation F in Appendix E suggests that the standard errors on our variable of interest increases by about 5%. As an alternative to simple scaling, we also report results from a robustness check that predicts CES-D scores in later waves based on the correlation structure in wave I.

Figure 8. Persistent Effects of Ordinal Ranks by Ability Decile



Notes: This figure presents the effect of ordinal ranks by ability decile for each of the waves as shown in Table 7. The shaded area indicates 90% confidence intervals clustered at the school level.

indicator for negative shocks) are significantly related to attrition status in waves II to IV, they are so in wave V. In particular, those individuals who receive a negative shock and who drive our results as shown in Section 6.3 are more likely to be missing in wave V.

Second, in Section 6.4 we have shown that two of four facets – the lack of a positive attitude and a lack of motivation – drive our results. Yet, the CES-D instrument administered in wave V elicited only a subset of five items rather than the full CES-D scale, with only a single item that loads on these facets (cf. Appendix Tables A.1 and C.1), which reduces the power to detect similar effects as in the previous waves.

Finally, as we will show in the next subsection, we find the same pattern observed for mental health in a range of economic outcomes measured in waves IV and V, suggesting that the effects indeed last for more than 14 years. We therefore think that the mental health measure available in wave V does not allow us to extend our analysis. Note, however, that

Table 7. Effects of Ordinal Ranks by Ability Decile

	Mental Health (CES-D score) by Decile										Obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
A. Immediate Effects (Wave I, 1994/1995)											
Rank	-5.16**	-2.07*	-1.98**	-1.81**	-2.27***	-1.28	-0.61	-0.92	-0.57	-0.49	18459
	(2.35)	(1.17)	(0.96)	(0.84)	(0.79)	(0.81)	(0.86)	(0.89)	(0.87)	(0.98)	
B. Short-term Effects (Wave II, 1996)											
Rank	-5.28**	-1.18	-1.51	-1.48	-2.21**	-1.21	-0.96	-0.88	-1.02	-1.88	13093
	(2.24)	(1.44)	(1.09)	(0.91)	(0.86)	(0.96)	(1.08)	(1.03)	(1.05)	(1.15)	
C. Medium-term Effects (Wave III, 2001/2002)											
Rank	-5.16*	-1.17	-0.44	0.26	0.97	-0.41	-0.21	-0.09	0.78	0.81	13551
	(2.97)	(1.63)	(1.25)	(1.06)	(1.05)	(0.95)	(0.96)	(1.04)	(1.04)	(1.27)	
D. Long-term Effects (Wave IV, 2008)											
Rank	-10.18***	-1.34	-3.05***	-1.48	-0.73	0.23	-0.04	-0.07	0.05	-0.52	14061
	(2.30)	(1.44)	(1.06)	(1.04)	(1.03)	(1.01)	(1.03)	(1.13)	(1.13)	(1.27)	

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses and clustered at the school level. We include all base controls, peer ability, school and grade fixed effects, as in our preferred baseline specification of column (1) in Table 3.

even if the effects indeed fade out between wave IV and V, our results still show that having a low rank in school worsens mental health at least for 14 years.

Predicting CES-D scores. In our analysis, we used the available items in each wave and scaled the resulting total score up to correspond to the full scale as in wave I and II. While this approach is transparent, it implicitly assumes that every available item has the same weight for the total score. Yet, as argued in Section 6.4, there are different facets captured by the CES-D scale inducing potentially different weights of the items. To acknowledge this unequal weighting, we construct different outcome measures for each wave by regressing the total CES-D score in wave I on the available items in the later wave.[24] In a second step, we then use the coefficients from these regressions as weights for the items in later waves, predict the total CES-D scores, and replicate the regressions in Table 7 using this predicted outcome measure. As shown in Appendix Table D.2, the resulting estimates replicate the findings presented in this section.

These results are in line with Prediction 4, suggesting that once a negative shock reduces a student's belief in their ability sufficiently, they withdraw study effort and therefore avoid new signals. As a consequence, their belief about the returns to ability remain low, positive updating is less likely, and depressions are some form of absorbing states. In other words, their mental health remains in a poor state and negative shocks may trigger potential vicious cycles. Therefore, our results show that the school environment can have long-lasting effects on the mental well-being of students over the life-cycle.

[24]See Appendix Table A.1 for a list of items that are available in each wave.

7.1 Long-Run Effects on Economic Outcomes

How do these long-run effects on mental health translate into other economic outcomes? Previous research suggests that worse mental health reduces educational attainment (e.g., Currie and Stabile, 2006) and lowers employment and earnings (e.g., Fletcher, 2014), and has linked higher ranks to higher educational attainment (Elsner and Isphording, 2017) and income (Denning, Murphy, and Weinhardt, 2018). We conduct two analyses to add to these results. First, we assess the correlation between CES-D scores in wave I and a range of long-term outcomes conditional on a rich set of background characteristics to provide correlational evidence of the importance of mental health in youth for long-term outcomes. Second, we study the causal effect of ranks in wave I on economic outcomes in adulthood. Together with our baseline estimates, we then can calculate how much of the long-run effects of rank are mediated through mental health.

In column (1) of Table 8, we show that a range of economic long-run outcomes – graduating from college, income, being employed and being late paying bills – as well as important non-economic outcomes – ever being married or ever being arrested – are all significantly related to mental health measured by CES-D scores in wave I. Importantly, these regressions control for a range of other individual characteristics and, most notably, a fourth-order polynomial in ability, as well as grade and school fixed effects as in our baseline specification. Increasing CES-D scores by one standard deviation, i.e., worsening mental health, is associated with a 5 percentage points decrease in the probability of having a college degree, being 2 percentage points less likely to be employed, 11% lower income, and with being 4 percentage points more likely to be late paying bills. In addition, we also find that those individuals with worse mental health in adolescence are also less likely to marry and are more likely to get arrested in adulthood. Although these associations are not necessarily causal, they highlight that having better mental health during adolescence predicts better economic and non-economic long-run outcomes.

We then present the effects of ordinal ranks during school on these outcomes. Being ranked higher during school significantly increases the probability of graduating from college, of being employed, and it increases income. More specifically, an increase of one standard deviation in a student's ordinal rank in school increases their likelihood of graduating from college by 4 percentage points, employment by 2 percentage points, and income by 8%. The average results on college graduation mimic the effects found by Elsner and Isphording (2017), while our income results are about double the size to those reported in Denning, Murphy, and Weinhardt (2018).[25] One potential explanation for the latter finding is that we can study income at a later point in life. If having a low rank sets people on different trajectories compared to those who have a high rank, this difference might increase over time explaining the effects that we observe.

[25]Elsner and Isphording (2017) also use AddHealth data, but in contrast to them, we can lever data up to wave V rather than IV, where some individuals might still be enrolled in college. Denning, Murphy, and Weinhardt (2018) use administrative records for students in Texas and earnings earlier in life.

Table 8. Long-run Outcomes, Mental Health, and Rank Effects

	(1) Association standardized CES-D scores	(2) Average Rank Effect	(3) Average Rank Effect (std.)	(4) Share of Rank Effect mediated by CES-D scores
𝟙{College graduate}	-0.05***	0.13***	0.04***	7%
	(0.00)	(0.05)	(0.01)	
log(Personal income)	-0.11***	0.30**	0.08***	6%
	(0.01)	(0.13)	(0.03)	
𝟙{Currently employed}	-0.02***	0.08*	0.02*	6%
	(0.00)	(0.04)	(0.01)	
𝟙{Late paying bills}	0.04***	-0.01	-0.00	74%
	(0.00)	(0.05)	(0.01)	
𝟙{Ever married}	-0.01***	0.05	0.02	5%
	(0.00)	(0.05)	(0.01)	
𝟙{Ever arrested}	0.03***	-0.01	-0.00	43%
	(0.00)	(0.04)	(0.01)	

Notes: This table presents the association of mental health measured by standardized CES-D scores in wave I and several long-run economic outcomes in column (1). These specifications control for all characteristics as our baseline specification apart from the ordinal rank. The second column presents the effect of ranks on economic outcomes based on our main specification, while column (3) presents the corresponding estimates using a standardized rank measure. Column (4) presents shares of these effect that are mediated by mental health. To obtain this share, we run an auxiliary regression where we add CES-D scores as an additional explanatory variable to the specification in column (2). We then calculate the share as the product of the coefficient from CES-D scores in this regression and the rank effect in Table 3, divided by the rank effect in column (2) of the present table. Outcomes are based on wave V data appended by wave IV data if the former data is missing. We add an additional indicator to control for the wave the outcome measure is from. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at the school level.

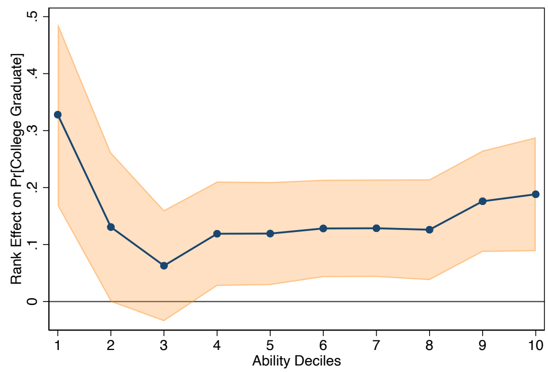
For the remaining outcomes, the rank effect estimates have the sign we would expect, but we do not find evidence of ranks affecting these outcomes on average. However, as we have shown in the previous section, there exists a pronounced heterogeneity with respect to ability. Specifically, the consequences of ranks are more pronounced at the lower end of the ability distribution. In Figure 9, we therefore present the corresponding estimates for economic long-run outcomes. Strikingly, we observe the same qualitative pattern for all outcomes – with being late paying bills as an exception – as for mental health: Rank effects are more pronounced for low-ability individuals and fade out with increasing ability.

We next ask how much of these long-run effects are mediated by mental health. To study this, we calculate the mediated effect as the product of the coefficient of rank on mental health in wave I with the coefficient of mental health on long-run outcomes, which simultaneously controls for the ordinal rank. We then express this mediated effect as a share of the rank effect shown in column (2).

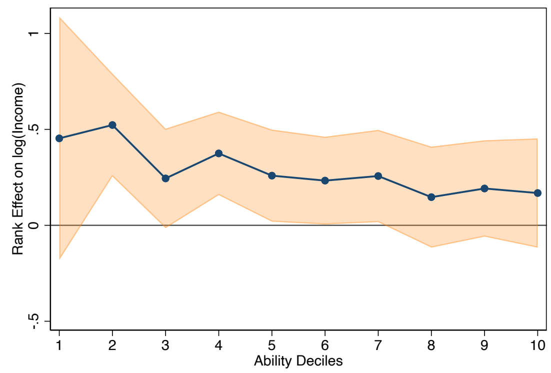
We find that 5-7% of the rank effect for having a college degree, income, employment, and marriage status are mediated by mental health. Interestingly, the mediated effect amounts to

Figure 9. Effects of Ordinal Ranks by Ability Decile on Long-run Outcomes

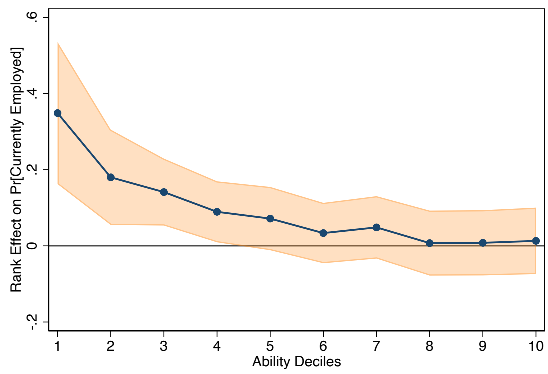
(a) College Degree



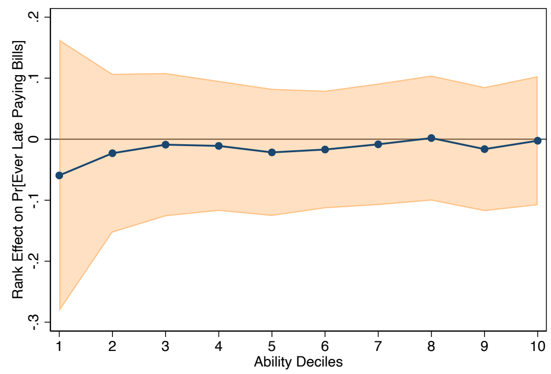
(b) log(income)



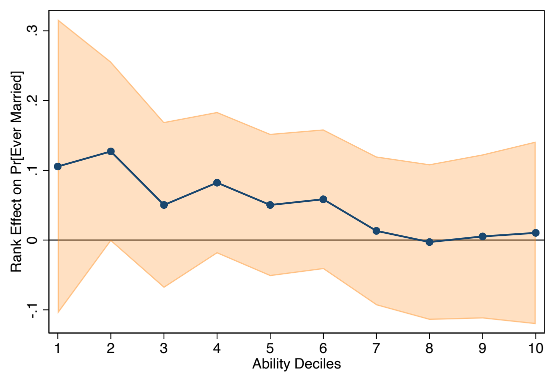
(c) Currently Employed



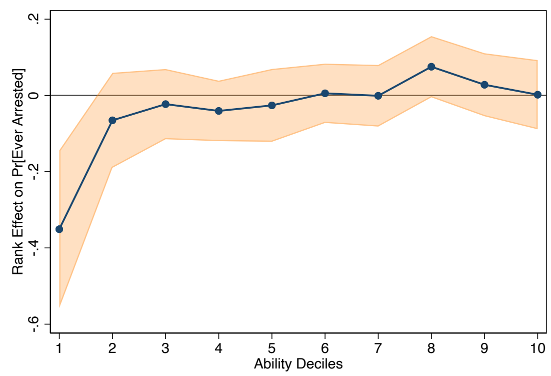
(d) Late Paying Bills



(e) Ever Married



(f) Ever Arrested



Notes: This figure presents the effect of ordinal ranks by ability decile on the outcome indicated in the title using our main specification for heterogeneous effects. The shaded area indicates 90% confidence intervals. Standard errors are clustered at the school level.

about 43% for ever being arrested, suggesting a potentially important link between mental health and criminal behavior.[26]

[26]Since we observe neither a linear nor a non-linear rank for being late paying bills, the mediated effect of 74% is uninformative.

Taken together, these results point in the same direction as our mental health results: experiencing negative shocks in school can have long-lasting negative consequences on many dimensions of life and is particularly pronounced for at-risk students who struggle at school. Moreover, the strong association of mental health during youth and economic long-run outcomes, as well as the finding that long-run effects seem to be partly mediated by mental health in adolescence, both suggest long-lasting consequences from a poor mental-health state during youth. These consequence affect an individual's economic and general well-being over the life-cycle.

8 Conclusion

What are the lasting effects of the school environment in general, and peers more specifically on the mental health of students? We present a theoretical framework in which students face uncertainty over their own ability and need to allocate their time between studying and leisure. Conditional on studying, they can exert effort to learn about it from noisy signals. Based on this framework, we derive several predictions. Negative signals worsen a student's mental health, and these effects are more pronounced at the lower end of the ability distribution and persist over time. In order to test these predictions, we use data from AddHealth, a longitudinal study of a representative set of students in the United States, and lever quasi-exogenous variation in the ability distribution across cohorts. This creates quasi-random variation in a student's ordinal rank, which – conditional on own and peer ability – we interpret as a signal about their ability, and which allows us to identify the causal effect of ranks on mental health of students.

We provide evidence that mental health is malleable during adolescence and document its persistence over time. Leveraging the quasi-random variation in ordinal ranks, we find support for our predictions: first, ordinal ranks causally affect students' mental health, measured by CES-D scores, an established self-assessment of depressions. Increasing a student's rank by one standard deviation improves their mental health by approx. 6% of a standard deviation, which is comparable to rank effects estimated for test scores (Murphy and Weinhardt, forthcoming), are approximately half of the effect of losing one's job on mental health (Marcus, 2013), and are about a sixth to a third of the effects from targeted positive psychology interventions on depression and psychological well-being (Bolier et al., 2013). Moreover, these effects are driven by negative rather than positive shocks, are more pronounced at the lower end of the ability distribution, and persist over time for at least 14 years. In addition, we find the same qualitative pattern in several economic outcomes measured in adulthood and show that these effects are partly mediated by mental health during adolescence.

Taken together, our study provides evidence on the long-lasting effects of features of the school environment. Also, it raises several avenues for further research. The persistence of our effects helps us to understand why mental health affects educational attainment (Currie and Stabile, 2006; Fletcher, 2010), and may be an additional skill valued on the labor market

(Fletcher, 2014; Heckman, Stixrud, and Urzua, 2006). Our results therefore point towards a mental health formation process similar to the accumulation of general health (Grossman, 1972) or the formation of other cognitive and non-cognitive skills (Cunha and Heckman, 2008; Cunha, Heckman, and Schennach, 2010). Extending these models would shed light on the relationship between these skills and mental health more generally. Moreover, such a model could be used to quantify the role of mental health to explain rank effects on educational outcomes.

What are the practical implications of our results? First, although the asymmetry of rank effects could provide a rationale to implement tracking within schools to create more homogeneous groups of students, we want to caution against these approaches. As shown in Carrell, Sacerdote, and West (2013), we do not understand the consequences of reassigning students well enough to design policies that strategically exploit peer effects. Moreover, we isolated a very specific peer effect, but other effects may co-exist. In fact, the consequences of different assignment rules may be ambiguous if peer effects are present in multiple dimensions (Kiessling, Radbruch, and Schaube, 2020b). Second, our results highlight that investments to compensate for negative shocks during school potentially have larger benefits than expected. They may not only compensate for immediate consequences of shocks, but protect students from being set on worse life trajectories, thus paying off in the long run.

We think that studying different causes of mental health and their long-term effects are a fruitful area for future research. Given the rise of mental health issues in the developed world and the wide-spread prevalence in developing countries (for a review of the relationship of poverty and mental health, see Ridley et al., 2020), policy-makers have a high interest in understanding the causes of these issues to design policies that alleviate mental illnesses.

References

Adhvaryu, Achyuta, James Fenske, and Anant Nyshadham (2019). “Early Life Circumstance and Adult Mental Health”. In: *Journal of Political Economy* 127.4, pp. 1516–1549.

Ahmed, Saz P., Amanda Bittencourt-Hewitt, and Catherine L. Sebastian (2015). “Neurocognitive Bases of Emotion Regulation Development in Adolescence”. In: *Developmental Cognitive Neuroscience* 15, pp. 11–25.

Anelli, Massimo and Giovanni Peri (2019). “The Effects of High School Peers’ Gender on College Major, College Performance and Income”. In: *Economic Journal* 129.618, pp. 553–602.

Angrist, Joshua D. and Victor Lavy (1999). “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement”. In: *Quarterly Journal of Economics* 114.2, pp. 533–575.

Angrist, Joshua D., Victor Lavy, Jetson Leder-Luis, and Adi Shany (2019). “Maimonides’ Rule Redux”. In: *American Economic Review: Insights* 1.3, pp. 309–324.

Armstrong, Thomas and Bunmi O. Olatunji (2012). “Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis”. In: *Clinical Psychological Review* 32.8, pp. 704–723.

Baird, Sarah, Jacobus de Hoop, and Berk Özler (2013). “Income Shocks and Adolescent Mental Health”. In: *Journal of Human Resources* 48.2, pp. 370–403.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko (2020). “Maternal Depression, Women’s Empowerment, and Parental Investment: Evidence from a Randomized Control Trial”. In: *American Economic Review* 110.3, pp. 842–859.

Bartel, Ann and Paul Taubman (1986). “Some Economic and Demographic Consequences of Mental Illness”. In: *Journal of Labor Economics* 4.2, pp. 243–256.

Beck, Aaron T. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. University of Pennsylvania Press.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen (2012). “Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain”. In: *Econometrica* 80.6, pp. 2369–2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014). “Inference on Treatment Effects after Selection Among High-Dimensional Controls”. In: *The Review of Economic Studies* 81.2, pp. 608–650.

Biasi, Barbara, Michael S. Dahl, and Petra Moser (2019). “Career Effects of Mental Health”. Working Paper.

Bietenbeck, Jan (2020). “The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR”. In: *Journal of the European Economic Association* 18.1, pp. 392–426.

Bifulco, Robert, Jason M. Fletcher, Sun Jung Oh, and Stephen L. Ross (2014). “Do high school peers have persistent effects on college attainment and other life outcomes?” In: *Labour Economics* 29, pp. 83–90.

Bifulco, Robert, Jason M. Fletcher, and Stephen L. Ross (2011). “The Effect of Classmate Characteristics on Post-secondary Outcomes: Evidence from the Add Health”. In: *American Economic Journal: Economic Policy* 3.1, pp. 25–53.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2013). “Under Pressure? The Effect of Peers on Outcomes of Young Adults”. In: *Journal of Labor Economics* 31.1, pp. 119–153.

Bolier, Linda, Merel Haverman, Gerben J. Westerhof, Helen Riper, Filip Smit, and Ernst Bohlmeijer (2013). “Positive psychology interventions: a meta-analysis of randomized controlled studies”. In: *BMC Public Health* 13.119.

Brown, Gordon D. A., Jonathan Gardner, Andrew J. Oswald, and Jing Qian (2008). “Does Wage Rank Affect Employees’ Well-being?” In: *Industrial Relations* 47.3, pp. 355–389.

Burgess, Simon, Ellen Greaves, Anna Vignoles, and Deborah Wilson (2015). “What Parents Want: School Preferences and School Choice”. In: *Economic Journal* 125.587, pp. 1262–1289.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez (2012). “Inequality at Work: The Effect of Peer Salaries on Job Satisfaction”. In: *American Economic Review* 102.6, pp. 2981–3003.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka (2018). “The Long-Run Effects of Disruptive Peers”. In: *American Economic Review* 108.11, pp. 3377–3415.

Carrell, Scott E., Bruce Sacerdote, and James West (2013). “From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation”. In: *Econometrica* 81.3, pp. 855–882.

Caspi, Avshalom, Renate M. Houts, Antony Ambler, Andrea Danese, Maxwell L. Elliott, Ahmad Hariri, HonaLee Harrington, Sean Hogan, Richie Poulton, Sandhya Ramrakha, et al. (2020). “Longitudinal Assessment of Mental Health Disorders and Comorbidities Across 4 Decades Among Participants in the Dunedin Birth Cohort Study”. In: *JAMA Network Open* 3.4, e203221–e203221.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2011). “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star”. In: *Quarterly Journal of Economics* 126.4, pp. 1593–1660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014). “Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood”. In: *American Economic Review* 104.9, pp. 2633–2679.

Clark, Luke, Samuel R. Chamberlain, and Barbara J. Sahakian (2009). “Neurocognitive Mechanisms in Depression: Implications for Treatment”. In: *Annual Review of Neuroscience* 32.1, pp. 57–74.

Cools, Angela, Raquel Fernández, and Eleonora Patacchini (2019). “Girls, Boys, and High Achievers”. Working Paper.

Croson, Rachel and Uri Gneezy (2009). “Gender Differences in Preferences”. In: *Journal of Economic Literature* 47.2, pp. 448–474.

Cunha, Flavio and James J. Heckman (2008). “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Journal of Human Resources* 43.4, pp. 738–782.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach (2010). “Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Econometrica* 78.3, pp. 883–931.

Currie, Janet and Mark Stabile (2006). “Child mental health and human capital accumulation: The case of ADHD”. In: *Journal of Health Economics* 25.6, pp. 1094–1118.

Delaney, Judith and Paul Devereux (2019). “The Effect of High School Rank in English and Math on College Major Choice”. Working Paper.

Denning, Jeffrey T., Richard Murphy, and Felix Weinhardt (2018). “Class Rank and Long-Run Outcomes”. Working Paper.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2011). “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya”. In: *American Economic Review* 101.5, pp. 1739–1774.

Dunn, Lloyd M. and Leota M. Dunn (2007). *The Peabody Picture Vocabulary Test, 4th edition*. NCS Pearson, Inc.

Eil, David and Justin M. Rao (2011). “The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself”. In: *American Economic Journal: Microeconomics* 3.2, pp. 114–138.

Eisenberg, Daniel, Ezra Golberstein, Janis L. Whitlock, and Marilyn F. Downs (2013). “Social Contagion of Mental Health: Evidence from College Roommates”. In: *Health Economics* 22.8, pp. 965–986.

Elsner, Benjamin and Ingo E. Isphording (2017). “A Big Fish in a Small Pond: Ability Rank and Human Capital Investment”. In: *Journal of Labor Economics* 35.3, pp. 787–828.

— (2018). “Rank, Sex, Drugs, and Crime”. In: *Journal of Human Resources* 52.2, pp. 356–381.

Elsner, Benjamin, Ingo E. Isphording, and Ulf Zölitz (2019). “Achievement rank affects performance and major choices in college”. Working Paper.

Ertac, Seda (2011). “Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback”. In: *Journal of Economic Behavior & Organization* 80.3, pp. 532–545.

Ettner, Susan L., Richard G. Frank, and Ronald C. Kessler (1997). “The Impact of Psychiatric Disorders on Labor Market Outcomes”. In: *Industrial and Labor Relations Review* 51.1, pp. 64–81.

Festinger, Leon (1954). “A Theory of Social Comparison Processes”. In: *Human Relations* 7.2, pp. 117–140.

Fletcher, Jason M. (2008). “Adolescent depression: diagnosis, treatment, and educational attainment”. In: *Health Economics* 17.11, pp. 1215–1235.

— (2010). “Adolescent depression and educational attainment: results using sibling fixed effects”. In: *Health Economics* 19.7, pp. 855–871.

Fletcher, Jason M. (2014). “The Effects of Childhood ADHD on Adult Labor Market Outcomes”. In: *Health Economics* 23.2, pp. 159–181.

Fowler, James H. and Nicholas A. Christakis (2008). “Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study”. In: *BMJ* 337.

Frank, Richard and Paul Gertler (1991). “An Assessment of Measurement Error Bias for Estimating the Effect of Mental Distress on Income”. In: *Journal of Human Resources* 26.1, pp. 154–164.

Fruehwirth, Jane Cooley, Sriya Iyer, and Anwen Zhang (2019). “Religion and Depression in Adolescence”. In: *Journal of Political Economy* 127.3, pp. 1178–1209.

Garlick, Robert (2018). “Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment”. In: *American Economic Journal: Applied Economics* 10.3, pp. 345–369.

Gill, David, Zdenka Kissová, Jaesun Lee, and Victoria Prowse (2019). “First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision”. In: *Management Science* 65.2, pp. 494–507.

Gotlib, Ian H., Elena Kranoperova, Dana N. Yue, and Jutta Joorman (2004). “Attentional Biases for Negative Interpersonal Stimuli in Clinical Depression”. In: *Journal of Abnormal Psychology* 113.1, pp. 121–135.

Gould, Eric D., Victor Lavy, and Daniele M. Paserman (2009). “Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence”. In: *Economic Journal* 119.540, pp. 1243–1269.

Graham, Bryan S. (2015). “Methods of Identification in Social Networks”. In: *Annual Review of Economics* 7.1, pp. 465–485.

Grossman, Michael (1972). “On the Concept of Health Capital and the Demand for Health”. In: *Journal of Political Economy* 80.2, pp. 223–255.

Hackman, Daniel A., Martha J. Farah, and Michael J. Meaney (2010). “Socioeconomic status and the brain: mechanistic insights from human and animal research”. In: *Nature Reviews Neuroscience* 11.9, pp. 651–659.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger (2009). “Heterogeneous Preferences and the Efficacy of Public School Choice”. Working Paper.

Haushofer, Johannes and Jeremy Shapiro (2016). “The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya”. In: *Quarterly Journal of Economics* 131.4, pp. 1973–2042.

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006). “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior”. In: *Journal of Labor Economics* 24.3, pp. 411–482.

Holtzheimer, Paul E. and Helen S. Mayberg (2011). “Stuck in a rut: rethinking depression and its treatment”. In: *Trends in Neurosciences* 34.1, pp. 1–9.

Hopkins, Ed and Tatiana Kornienko (2004). “Running to Keep in the Same Place: Consumer Choice as a Game of Status”. In: *American Economic Review* 94.4, pp. 1085–1107.

Hoxby, Caroline M. (2000a). “Peer Effects in the Classroom: Learning from Gender and Race Variation”. Working Paper.

— (2000b). “The Effects of Class Size on Student Achievement: New Evidence from Population Variation”. In: *Quarterly Journal of Economics* 115.4, pp. 1239–1285.

Jackson, C. Kirabo, Diether W. Beuermann, Laia Navarro-Sola, and Francisco Pardo (2019). “What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output”. Working Paper.

Jensen, Arthur R. (1998). *The g Factor: The Science of Mental Ability*. Praeger.

Jensen, Robert (2010). “The (Perceived) Returns to Education and the Demand for Schooling”. In: *Quarterly Journal of Economics* 125.2, pp. 515–548.

Kahneman, Daniel and Angus Deaton (2010). “High income improves evaluation of life but not emotional well-being”. In: *Proceedings of the National Academy of Sciences* 107.38, pp. 16489–16493.

Kessler, Ronald C., Matthias Angermeyer, and James C. et al. Anthony (2007). “Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization’s World Mental Health Survey Initiative”. In: *World Psychiatry* 6.3, pp. 168–176.

Kiessling, Lukas, Jonas Radbruch, and Sebastian Schaube (2020a). “Determinants of Peer Selection”.

— (2020b). “Self-selection of peers and performance”.

Korn, Christoph W., Tali Sharot, Hendrik Walter, Hauke R. Heekeren, and Raymond J. Dolan (2014). “Depression is related to an absence of optimistically biased belief updating about future life events”. In: *Psychological Medicine* 44.3, pp. 579–592.

Krueger, Alan B. (1999). “Experimental Estimates of Education Production Functions”. In: *Quarterly Journal of Economics* 114.2, pp. 497–532.

Krueger, Alan B. and Diane M. Whitmore (2001). “The Effect of Attending a Small Class in the Early Grades on College-test Taking and Middle School Test Results: Evidence from Project Star”. In: *Economic Journal* 111.468, pp. 1–28.

Kuppens, Peter, Nicholas B. Allen, and Lisa B. Sheeber (2010). “Emotional Inertia and Psychological Maladjustment”. In: *Psychological Science* 21.7, pp. 984–991.

Kuziemko, Ilyana, Ryan W. Buell, Taly Reich, and Michael I. Norton (2014). ““Last-Place Aversion”: Evidence and Redistributive Implications”. In: *Quarterly Journal of Economics* 129.1, pp. 105–149.

Lavy, Victor and Analia Schlosser (2011). “Mechanisms and Impacts of Gender Peer Effects at School”. In: *American Economic Journal: Applied Economics* 3.2, pp. 1–33.

Lewinsohn, Peter M., John R. Seeley, Robert E. Roberts, and Nicholas B. Allen (1997). “Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults”. In: *Psychology and Aging* 12.2, pp. 277–287.

Luttmer, Erzo F. P. (2005). “Neighbors as Negatives: Relative Earnings and Well-Being”. In: *Quarterly Journal of Economics* 120.3, pp. 963–1002.

Marcus, Jan (2013). “The effect of unemployment on the mental health of spouses – Evidence from plant closures in Germany”. In: *Journal of Health Economics* 32.3, pp. 546–558.

Marsh, Herbert W. (1987). “The Big-Fish-Little-Pond Effect on Academic Self-Concept”. In: *Journal of Educational Psychology* 79.3, pp. 280–295.

McCormick, C. M. and M. R. Green (2013). “From the Stressed Adolescent to the Anxious and Depressed Adult: Investigations in Rodent Models”. In: *Neuroscience* 249, pp. 242–257.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001). “Birds of a Feather: Homophily in Social Networks”. In: *Annual Review of Sociology* 27.1, pp. 415–444.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat (2014). “Managing Self-Confidence”. Working Paper.

Mojtabai, Ramin, Mark Olfson, and Beth Han (2016). “National Trends in the Prevalence and Treatment of Depression in Adolescents and Young Adults”. In: *Pediatrics* 138.6.

Momen, Natalie C. et al. (2020). “Association between Mental Disorders and Subsequent Medical Conditions”. In: *New England Journal of Medicine* 382.18, pp. 1721–1731.

Murphy, Jane M. (2011). “Symptom Scales and Diagnostic Schedules in Adult Psychiatry”. In: *Textbook of Psychiatric Epidemiology*. John Wiley & Sons, Ltd. Chap. 13, pp. 199–219.

Murphy, Richard and Felix Weinhardt (forthcoming). “Top of the Class: The Importance of Ordinal Rank”. In: *Review of Economic Studies*.

Niederle, Muriel and Lise Vesterlund (2007). “Do Women Shy Away From Competition? Do Men Compete Too Much?” In: *Quarterly Journal of Economics* 122.3, pp. 1067–1101.

OECD (2015). *Fit Mind, Fit Job*. Mental Health and Work, OECD Publishing.

Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou (2020). “Mothers, Peers, and Gender-Role Identity”. In: *Journal of the European Economic Association* 18.1, pp. 266–301.

Oster, Emily (2019). “Unobservable Selection and Coefficient Stability: Theory and Evidence”. In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Pagani, Laura, Simona Comi, and Federica Origo (forthcoming). “The Effect of School Rank on Personality Traits”. In: *Journal of Human Resources*.

Persson, Petra and Maya Rossin-Slater (2018). “Family Ruptures, Stress, and the Mental Health of the Next Generation”. In: *American Economic Review* 108.4-5, pp. 1214–1252.

Plana-Ripoll, Oleguer et al. (2019). “Exploring Comorbidity Within Mental Disorders Among a Danish National Population”. In: *JAMA Psychiatry* 76.3, pp. 259–270.

Quidt, Jonathan de and Johannes Haushofer (2016). “Depression for Economists”. Working Paper.

Radloff, Lenore S. (1977). “The CES-D scale: A self report depression scale for research in the general population”. In: *Applied Psychological Measurements* 1.3, pp. 385–401.

— (1991). “The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults”. In: *Journal of Youth and Adolescence* 20.2, pp. 149–166.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel (2020). “Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms”. Working Paper.

Roberts, Robert E., Judy A. Andrews, Peter M. Lewinsohn, and Hyman Hops (1990). "Assessment of depression in adolescents using the Center for Epidemiologic Studies Depression Scale". In: *Psychological Assessment: A Journal of Consulting and Clinical Psychology* 2.2, pp. 122–128.

Roiser, Jonathan P., Rebecca Elliott, and Barbara J. Sahakian (2012). "Cognitive Mechanisms of Treatment in Depression". In: *Neuropsychopharmacology* 37, pp. 117–136.

Rosenquist, James N., James H. Fowler, and Nicholas A. Christakis (2011). "Social network determinants of depression". In: *Molecular Psychiatry* 16, pp. 273–281.

Rothstein, Jesse M. (2006). "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions". In: *American Economic Review* 96.4, pp. 1333–1350.

— (2017). "Measuring the Impacts of Teachers: Comment". In: *American Economic Review* 107.6, pp. 1656–1684.

Schwandt, Hannes (2018). "Wealth Shocks and Health Outcomes: Evidence from Stock Market Fluctuations". In: *American Economic Journal: Applied Economics* 10.4, pp. 349–377.

Seligman, Martin E. P. (1972). "Learned Helplessness". In: *Annual Review of Medicine* 23.1, pp. 407–412.

Stewart, Walter F., Judith A. Ricci, Elsbeth Chee, Steven R. Hahn, and David Morganstein (2003). "Cost of Lost Productive Work Time Among US Workers With Depression". In: *JAMA* 289.23, pp. 3135–3144.

Stinebrickner, Todd and Ralph Stinebrickner (2012). "Learning about Academic Ability and the College Dropout Decision". In: *Journal of Labor Economics* 30.4, pp. 707–748.

Thapar, Anita, Stephan Collishaw, Daniel Pine, and Ajay Thapar (2012). "Depression in Adolescence". In: *The Lancet* 379.9820, pp. 1056–1067.

Tincani, Michela M. (2017). "Heterogeneous Peer Effects and Rank Concerns: Theory and Evidence". Working Paper.

Zafar, Basit (2011). "How Do College Students Form Expectations?" In: *Journal of Labor Economics* 29.2, pp. 301–348.

Zhang, Anwen (2018). "Peer Effects on Mental Health: Evidence from Random Assignment into Classrooms". Working Paper.

Appendix – For Online Publication

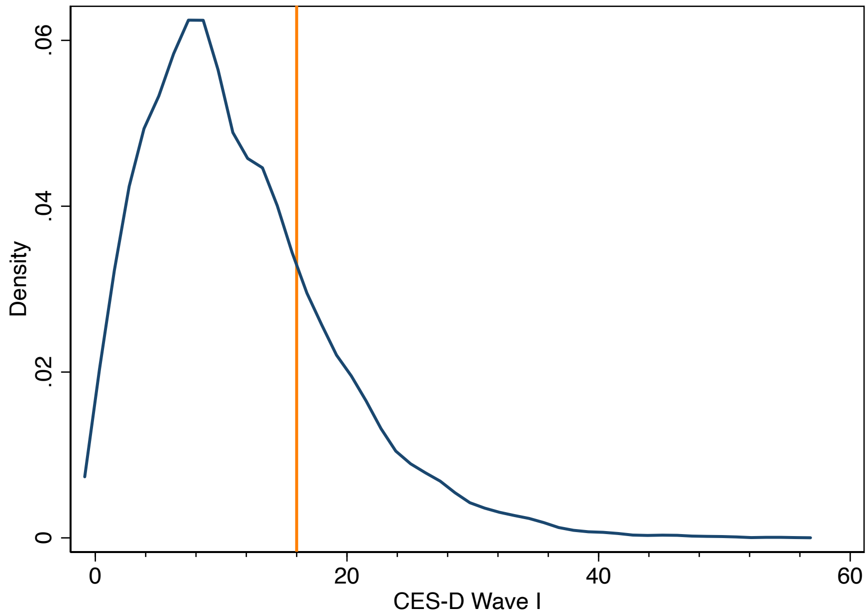
A Additional Tables and Figures

Table A.1. Items of the CES-D Scale

How often was the following true during the past week?	Wave I	Wave II	Wave III	Wave IV	Wave V
1. You were bothered by things that don't usually bother you.	X	X	X	X	
2. You didn't feel like eating, your appetite was poor.	X	X			
3. You felt that you could not shake off the blues, even with help from your family and your friends.	X	X	X	X	X
4. You felt you were just as good as other people.	X	X	X	X	
5. You had trouble keeping your mind on what you were doing.	X	X	X	X	
6. You felt depressed.	X	X	X	X	X
7. You felt that you were too tired to do things.	X	X	X	X	
8. You felt hopeful about the future.	X	X			
9. You thought your life had been a failure.	X	X			
10. You felt fearful.	X	X			
11. You were happy.	X	X		X	X
12. You talked less than usual.	X	X			
13. You felt lonely.	X	X			
14. People were unfriendly to you.	X	X			
15. You enjoyed life.	X	X	X	X	
16. You felt sad.	X	X	X	X	X
17. You felt that people disliked you.	X	X	X	X	
18. It was hard to get started doing things.	X	X			
19. You felt life was not worth living.	X	X			X
Number of items	19	19	9	10	5

Notes: This table presents all items of the CES-D scale and in which wave they were elicited. Responses are rated on a scale from 0 ("never or rarely") to 3 ("most of the time or all of the time") and aggregated to a final score ranging from 0 to 57, with higher scores indicating a higher propensity for depressive symptoms. For our main analysis, CES-D scores in wave III, IV, and V are scaled by $^{19}/_{9}$, $^{19}/_{10}$, and $^{19}/_{5}$, respectively.

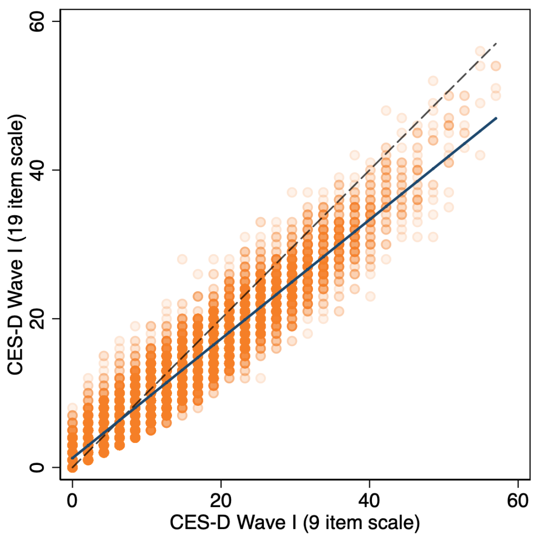
Figure A.1. Distribution of CES-D Scores at Wave I



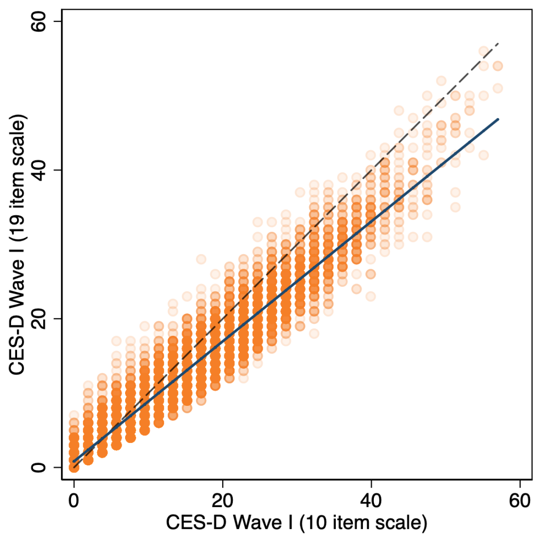
Notes: This figure presents the distribution of the our mental health measure (CES-D score) in wave I. The vertical line indicates a threshold of 16 often used as an indicator of depressions (Radloff, 1977).

Figure A.2. Relationship of Long- and Short-Scale of the CES-D Score

(a) 9-item scale (Wave III)



(b) 10-item scale (Wave IV)



Notes: This figure presents the relationship of the CES-D using 19 items as used in wave I and a 9-item (10-item) short version adopted in wave III (IV) including a linear fit in Figure A.2a (Figure A.2b). The dashed line indicates the 45-degree line.

Table A.2. Associations of Covariates with CES-D Scores

	CES-D score		CES-D score (std.)	
	(1)	(2)	(3)	(4)
Ability (std.)	-1.21***	-1.20***	-0.16***	-0.16***
	(0.06)	(0.06)	(0.01)	(0.01)
Female	1.94***	1.92***	0.25***	0.25***
	(0.14)	(0.14)	(0.02)	(0.02)
White	-0.66***	-0.70***	-0.09***	-0.09***
	(0.19)	(0.19)	(0.03)	(0.03)
College-educated parents	-0.53***	-0.49***	-0.07***	-0.06***
	(0.15)	(0.15)	(0.02)	(0.02)
Single-parent household	0.80***	0.79***	0.10***	0.10***
	(0.14)	(0.15)	(0.02)	(0.02)
Age (in years)	0.79***	0.76***	0.10***	0.10***
	(0.08)	(0.08)	(0.01)	(0.01)
School and Grade FEs	Yes	No	Yes	No
School × Grade FEs	No	Yes	No	Yes
Observations	18432	18432	18432	18432
R^2	0.101	0.121	0.101	0.121

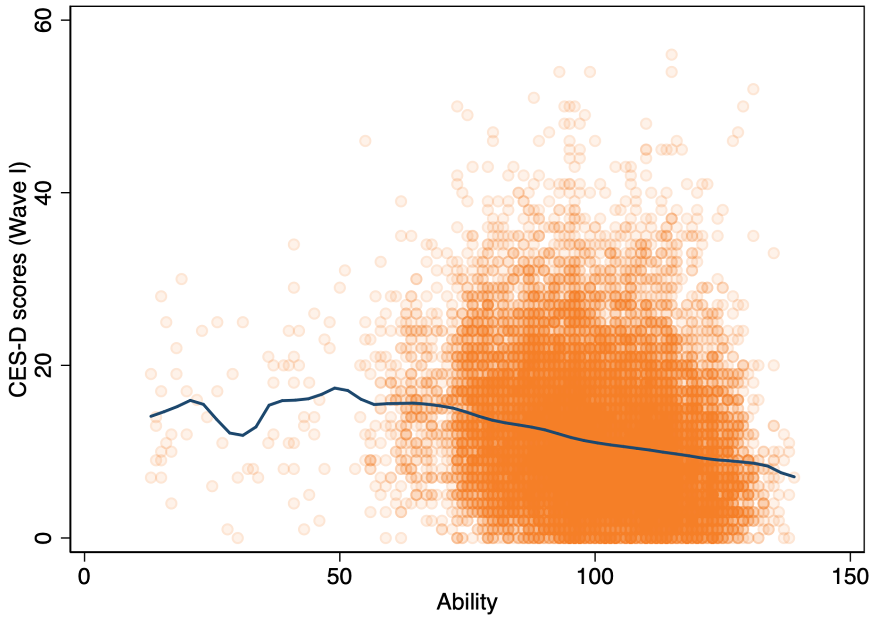
Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses and clustered at the school level. Odd-numbered columns are specifications with seperate school and grade fixed effects, while even-numbered columns employ school-specific grade fixed effects. The first two columns use raw CES-D scores, while the last two columns present estimates using standardized CES-D scores. In contrast to our specifications for the analysis of rank effects, here we use just a linear specifications for ability and age to facilitate interpretation.

Table A.3. Variation in Ranks

	Standard Deviation in Rank Variable										
	Full	By Decile									
	Sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
No Controls	0.28	0.09	0.14	0.17	0.18	0.18	0.18	0.17	0.16	0.14	0.11
Controls, School and Grade FE	0.08	0.09	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.12
Controls, School-by-Grade FE	0.08	0.09	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.11
Observations	18459	1872	1984	2006	2141	1340	2006	1630	1904	2003	1573

Notes: This table presents the variation in our variable of interest for the full sample and by ability decile. The first row presents the raw variation. The second row takes out all variation from individual controls, school and grade fixed effects as in our baseline specification and presents the standard deviation in the rank residuals. The third row additionally controls for school-by-grade fixed effects similar to our alternative identification strategy.

Figure A.3. Relationship of CES-D Score and Ability



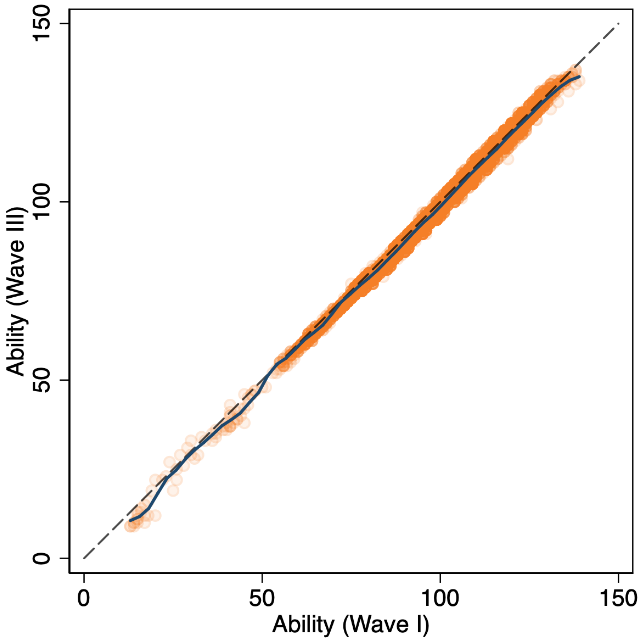
Notes: This figure presents a scatter plot of the CES-D score in wave I on our ability measure, which is constructed from a regression of CES-D on AH PVT scores including a local polynomial fit.

Table A.4. Salience of Ranks and Reverse Causality

	Intelligence (1–6)		College Exp. (1–5)		Ability (Wave III)	
	(1)	(2)	(3)	(4)	(5)	(6)
Rank	1.05***	0.28**	1.15***	0.44**	44.00***	0.20*
	(0.04)	(0.13)	(0.07)	(0.18)	(0.90)	(0.11)
Ability and Controls	No	Yes	No	Yes	No	Yes
School and Grade FEs	Yes	Yes	Yes	Yes	Yes	Yes
N	18434	18434	18389	18389	18459	18459
R^2	0.123	0.159	0.095	0.162	0.882	0.998

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification. Column headers denote the dependent variable. "Intelligence" is how intelligent the adolescent feels compared to other people their age (1-6 with 1 moderately below average and 6 extremely above average). "College expectations" is a scale based on the sum of the adolescents' report on how much they want to go to college and how likely it is they will go to college (each is 1-5 with 1 low and 5 high). "Ability (Wave III)" measures AH PVT scores in wave III and is standardized to have a mean of 100 and a standard deviation of 15.

Figure A.4. Stability of Ability Across Time



Notes: This figure presents a scatter plot and nonlinear fit of the ability measure in wave III on the ability measure in wave I. The dashed line indicates the 45-degree line.

B Robustness Checks for Average Results

Table B.1. Robustness to Ability Nonlinearity

	Iterations of Ability Polynomial Controls						FEs	PDS Lasso	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rank	-1.83***	-2.20***	-1.47**	-1.59**	-1.59**	-1.58**	-1.50**	-2.20***	-1.69***
	(0.63)	(0.68)	(0.73)	(0.75)	(0.75)	(0.76)	(0.76)	(0.68)	(0.62)
Ability	-0.04***	-0.07**	0.25**	0.40	-0.43	-1.64		-0.07**	-0.05***
	(0.01)	(0.03)	(0.10)	(0.28)	(0.56)	(1.58)		(0.03)	(0.01)
$(\text{Ability})^2$		0.00	-0.00***	-0.01	0.02	0.07		0.00	
		(0.00)	(0.00)	(0.01)	(0.02)	(0.07)		(0.00)	
$(\text{Ability})^3$			0.00***	0.00	-0.00	-0.00			
			(0.00)	(0.00)	(0.00)	(0.00)			
$(\text{Ability})^4$				-0.00	0.00	0.00			
				(0.00)	(0.00)	(0.00)			
$(\text{Ability})^5$					-0.00	-0.00			
					(0.00)	(0.00)			
$(\text{Ability})^6$						0.00			
						(0.00)			
Ability FEs	No	No	No	No	No	No	Yes	No	No
Penalize Ability	No	No	No	No	No	No	No	Yes	Yes
Penalize Controls	No	No	No	No	No	No	No	No	Yes
#Pen. Variables incl.								8	41
#Pen. Variables sel.								2	11
Observations	18459	18459	18459	18459	18459	18459	18459	18459	18459

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification. Column (7) includes fixed effects for every level of ability (AH PVT score) to control for a fully non-parametric specification of ability. In columns (8) and (9), we report results from the post-double selection (PDS) Lasso method by Belloni, Chernozhukov, and Hansen (2014), using the theory driven penalizer selection of Belloni et al. (2012). More specifically, column (8) includes up to an 8-degree polynomial in ability (AH PVT scores) and allows the Lasso to select only over these (baseline controls are not penalized, thus always included). Only the linear and quadratic ability polynomials are selected. In column (9), we again include up to an 8-degree polynomial in ability and allow selection on both these and our baseline control set (excludes school and grade fixed effects, which are always included). Only a linear ability term is selected of the ability polynomials and 10 additional controls from the remaining control set. Under the PDS Lasso method, standard errors and statistics are only valid for the rank coefficient.

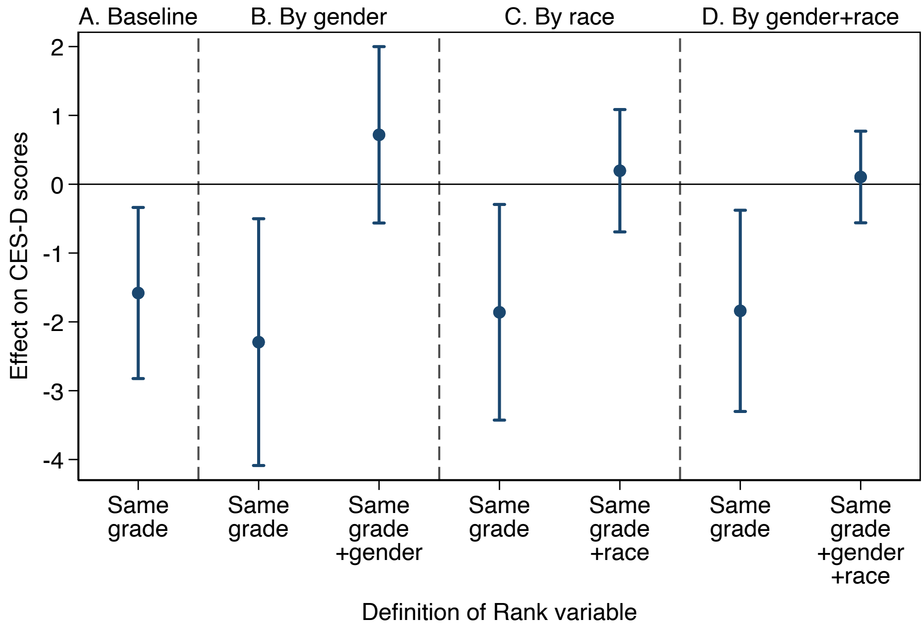
Table B.2. Comparison of Different Methods for Calculating Ranks

	Mental Health (CES-D score)					
	Bottom Rank		Top Rank		Mean Rank	
	(1)	(2)	(3)	(4)	(5)	(6)
Rank	-1.58**	-1.67**	-1.23*	-1.37*	-1.45*	-1.58**
	(0.75)	(0.76)	(0.73)	(0.74)	(0.74)	(0.75)
Ability	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Peer Ability	Yes	No	Yes	No	Yes	No
School and Grade FEs	Yes	No	Yes	No	Yes	No
School × Grade FEs	No	Yes	No	Yes	No	Yes
Observations	18459	18459	18459	18459	18459	18459
R^2	0.108	0.127	0.108	0.127	0.108	0.127

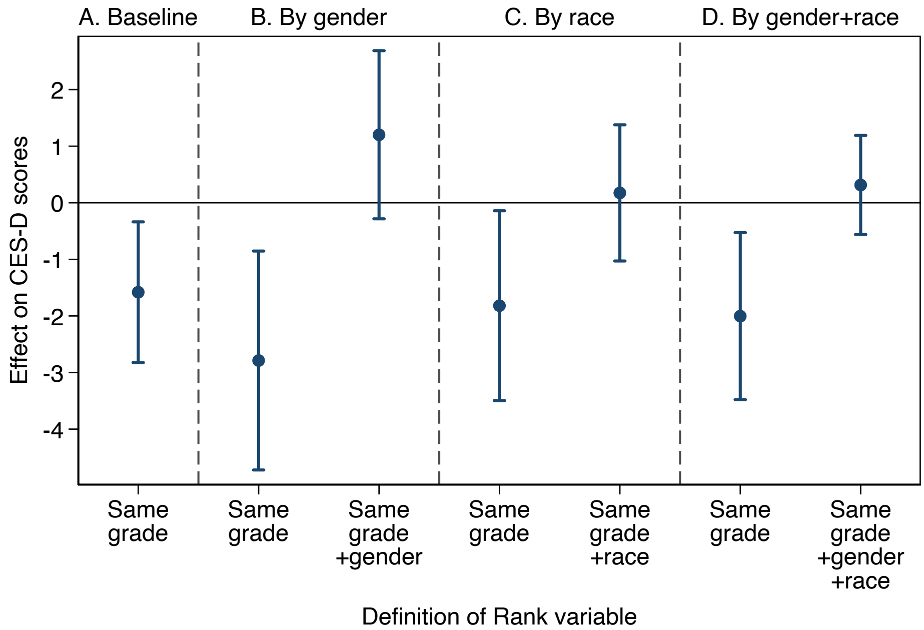
Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses and clustered at the school level. We include all base controls, peer ability, school and grade fixed effects, as in our preferred baseline specification of column (1) in Table 3. The column headers indicate how we calculated ranks, in particular how we break ties for individuals that have the same ability. “Bottom rank” in columns (1) and (2) are calculated by counting the number of peers in the same grade that have a strictly lower ability and corresponds to our main specification in Table 3. “Top rank” in columns (3) and (4) assign a rank based on the number of individuals having a lower or equal ability, whereas “mean rank” in columns (5) and (6) assigns the average of both approaches as the rank.

Figure B.1. Different Definitions of Peers Groups

(a) No Controls for Peer Ability of Newly Defined Peer Groups



(b) Including Controls for Peer Ability of Newly Defined Peer Groups

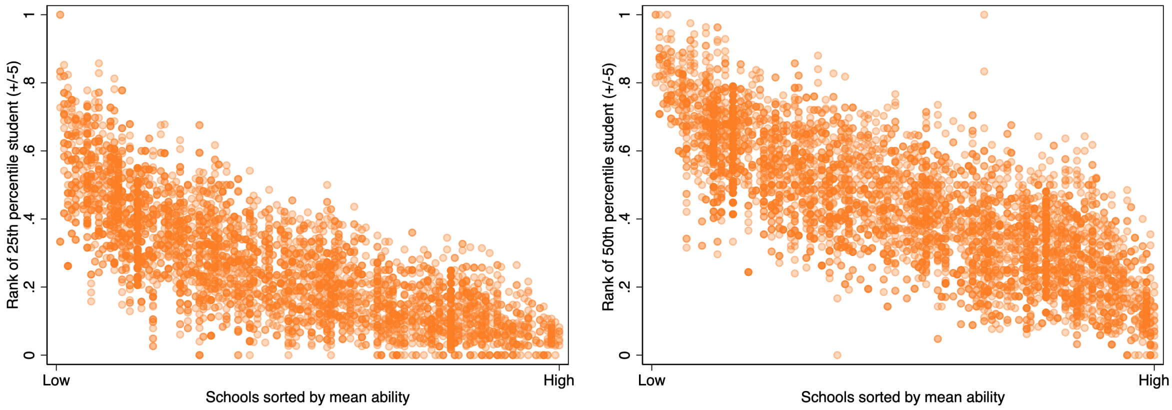


Notes: These figures show how different definitions of peer groups affect our baseline effects of ordinal ranks on CES-D scores. Panel A presents our baseline estimate. Panels B-D add an additional term for ranks defined based on peers in the same grade, as well as having the same gender (Panel B), the same race/ethnicity (Panel C), or the same gender and race/ethnicity (Panel D). Panels B-D of Figure B.1a only include the rank for the subgroup, while the corresponding panels in Figure B.1b additionally control for average peer ability of the subgroup. Whiskers indicate 90% confidence intervals clustered at the school level.

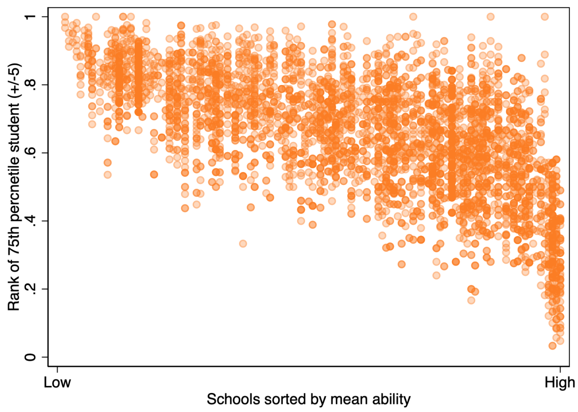
Figure B.2. Difficulty of Sorting into Schools based on Ranks

(a) 25th Ability Percentile

(b) 50th Ability Percentile



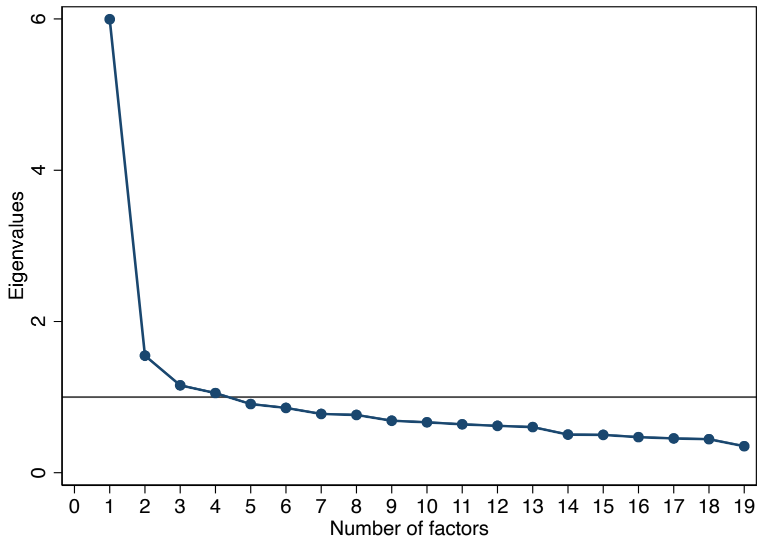
(c) 75th Ability Percentile



Notes: These figures present ranks for students at the 25th (50th, 75th) percentile of the ability distribution ±5 ability scores by schools. Schools are sorted based on the mean ability of students from the lowest to the highest.

C Principal Component Analysis of CES-D Items

Figure C.1. Screeplot of a Principal Component Analysis of CES-D Items



Notes: This figure presents a screeplot of a principal component factor analysis, i.e., the eigenvalues of each of the recovered factors, using all 19 items of the CES-D scale.

Table C.1. Factor Loadings of CES-D Items for Principal Component Factors

	Mental Health (Principal Components)			
	Loneliness	Lack of Pos. Attitude	Lack of Motivation	External Factors
1. You were bothered by things that don't usually bother you.	0.54	0.10	0.32	0.07
2. You didn't feel like eating, your appetite was poor.	0.46	0.08	0.33	-0.09
3. You felt that you could not shake off the blues, even with help from your family and your friends.	**0.74**	0.14	0.20	0.05
4. You felt you were just as good as other people.	0.09	**0.68**	0.05	0.12
5. You had trouble keeping your mind on what you were doing.	0.31	0.10	**0.60**	0.12
6. You felt depressed.	**0.76**	0.18	0.19	0.13
7. You felt that you were too tired to do things.	0.21	0.11	**0.69**	0.13
8. You felt hopeful about the future.	-0.00	**0.76**	0.09	0.01
9. You thought your life had been a failure.	0.56	0.22	-0.00	0.34
10. You felt fearful.	0.45	0.02	0.17	0.30
11. You were happy.	0.32	**0.67**	0.09	0.06
12. You talked less than usual.	0.30	0.13	0.31	0.08
13. You felt lonely.	**0.66**	0.12	0.16	0.21
14. People were unfriendly to you.	0.09	0.02	0.15	**0.81**
15. You enjoyed life.	0.31	**0.68**	0.07	0.11
16. You felt sad.	**0.70**	0.14	0.16	0.21
17. You felt that people disliked you.	0.23	0.13	0.12	**0.78**
18. It was hard to get started doing things.	0.12	0.08	**0.71**	0.22
19. You felt life was not worth living.	0.54	0.20	-0.07	0.33

Notes: This table presents loadings of all 19 CES-D items for the four principal component factors. Bold entries indicate items with loadings larger than 0.6, which we use to assign interpretations to these factors.

D Robustness Checks for Persistence Results

Table D.1. Effects of Ordinal Ranks by Ability Decile in Wave 5

	Mental Health (CES-D score) by Decile										Obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
Rank	0.64	0.02	0.61	0.10	0.80	0.33	0.35	1.49	1.22	1.17	11085
	(2.99)	(2.22)	(1.52)	(1.38)	(1.33)	(1.17)	(1.25)	(1.29)	(1.27)	(1.52)	

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses and clustered at the school level. This table presents results for wave V corresponding to the results in waves I through IV in Table 7. We include all base controls, peer ability, school and grade fixed effects, as in our preferred baseline specification presented in column (1) of Table 3.

Table D.2. Effects of Ordinal Ranks by Ability Decile – Using Predicted CES-D Scores

	Mental Health (CES-D score) by Decile										Obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
A. Immediate Effects (Wave I, 1994/1995)											
Rank	-5.16**	-2.07*	-1.98**	-1.81**	-2.27***	-1.28	-0.61	-0.92	-0.57	-0.49	18459
	(2.35)	(1.17)	(0.96)	(0.84)	(0.79)	(0.81)	(0.86)	(0.89)	(0.87)	(0.98)	
B. Short-term Effects (Wave II, 1996)											
Rank	-5.28**	-1.18	-1.51	-1.48	-2.21**	-1.21	-0.96	-0.88	-1.02	-1.88	13093
	(2.24)	(1.44)	(1.09)	(0.91)	(0.86)	(0.96)	(1.08)	(1.03)	(1.05)	(1.15)	
C. Medium-term Effects (Wave III, 2001/2002)											
Rank	-4.28*	-0.89	-0.26	0.31	0.84	-0.28	-0.10	-0.06	0.63	0.65	13551
	(2.36)	(1.29)	(0.99)	(0.84)	(0.84)	(0.75)	(0.76)	(0.82)	(0.82)	(1.01)	
D. Long-term Effects (Wave IV, 2008)											
Rank	-8.19***	-1.09	-2.49***	-1.19	-0.59	0.16	-0.09	-0.06	-0.02	-0.45	14061
	(1.85)	(1.16)	(0.84)	(0.84)	(0.82)	(0.81)	(0.82)	(0.90)	(0.90)	(1.01)	
E. Very Long-term Effects (Wave V, 2016–2018)											
Rank	0.71	-0.12	0.26	-0.09	0.42	0.10	0.18	1.01	0.82	0.82	11085
	(2.11)	(1.61)	(1.08)	(0.98)	(0.97)	(0.84)	(0.90)	(0.93)	(0.91)	(1.08)	

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses and clustered at the school level. We include all base controls, peer ability, school and grade fixed effects, as in our baseline specifications of column (1) in Table 3. In contrast to Table 7, we construct CES-D scores using weights from OLS-regressions in wave I, in which we regress the total CES-D score on the items of the CES-D scale that are available in later waves. We use the coefficients from these regressions as weights to predict CES-D scores.

Table D.3. Attrition Analysis by Wave

	𝟙{Attrited in Wave II}			𝟙{Attrited in Wave III}			𝟙{Attrited in Wave IV}			𝟙{Attrited in Wave V}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Rank	0.05			-0.03			0.03			-0.05		
	(0.03)			(0.04)			(0.04)			(0.04)		
Negative Shock		-0.01			0.02			0.01			0.03***	
		(0.01)			(0.01)			(0.01)			(0.01)	
CES-D score (Wave I)			-0.00			-0.00			-0.00			0.00***
			(0.00)			(0.00)			(0.00)			(0.00)
Ability	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer Ability	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School and Grade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share Attrited	.29	.29	.29	.27	.27	.27	.24	.24	.24	.4	.4	.4
N	18459	18459	18459	18459	18459	18459	18459	18459	18459	18459	18459	18459
R^2	.29	.29	.29	.059	.059	.059	.054	.054	.054	.074	.075	.075

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses and clustered at the school level. Each specification includes all controls as in our preferred baseline specification. The dependent variable is an indicator equal to one if an individual has attrited in wave II/III/IV/V and zero otherwise.

E Simulations to Assess Various Forms of Measurement Error

In the following, we present different simulations to assess the role of various forms of measurement error. Our point of departure is the following data-generating process (DGP):

$$y = -1.8r - 0.6a \tag{4}$$

in which y denotes our outcome, mental health as assessed by the CES-D scale, a denotes a student's ability, which is randomly drawn from a standard normal distribution ($a \sim N(0, 1)$), and r denotes a students' rank based on the ability distribution in their school cohort. For the simulations, we abstract from the fact that we observe several cohorts per school. The parameters for the simulations ($\beta = -1.8$ and $\gamma = -0.6$) are based on the simple specification shown in column (1) of Appendix Table B.1 and scaled up as the AHPVT scores have a standard deviation of 15.

Given the data-generating process in equation (4), we assess the consequences of several forms of measurement error using Monte Carlo simulations.[1] For each of the simulations reported below, we run 1000 repetitions with 500 schools/cohorts each and 180 students per school, and estimate specifications of $y = \beta r + \delta a$. For each specification, we report the average estimate $\hat{\beta}$ as well as the ratio of estimated effect and true coefficient in parentheses.

A. Random sampling of students per school. We begin by assessing the consequences of observing a random sample of students per school. Hence, in our first exercise, illustrated in Table E.1 and Figure E.1, we assess what happens if we only observe a subset of students in each school. To do this, we simulate schools of 180 students and decrease the share of students in our sample from full saturation, i.e., sampling all students in a school/cohort, to a situation in which we only observe 10% of all students. The simulations demonstrate that random sampling within cohorts biases the coefficient towards zero: if we observe half of all students, our estimates would be attenuated by 10%; in schools for which we only observe 10% of the sample, attenuation is more severe and the estimated effects correspond to approximately 50% of the original effect. Hence, random sampling of students implies that we underestimate the true effect.

B. Measurement error in ability measure. In our second set of simulations, we introduce measurement error in our ability measure. In particular, we assess how our estimates change once we introduce noise into our measurement, i.e., we measure $\tilde{a} = a + \phi z$ rather than a, where $a, z \sim N(0, 1)$ and $\phi \in [0, 1]$. Thus, $\phi = 0$ corresponds to situations in which we have no measurement error, whereas $\phi = 1$ corresponds to a situation where we have as much measurement error as noise in our ability measure. This measurement error in our ability measure translates into measurement error in the rank that we assign students in their respective cohorts ($r(\tilde{a})$) as measurement error pertubates the ranks. Table E.2 and Figure E.2

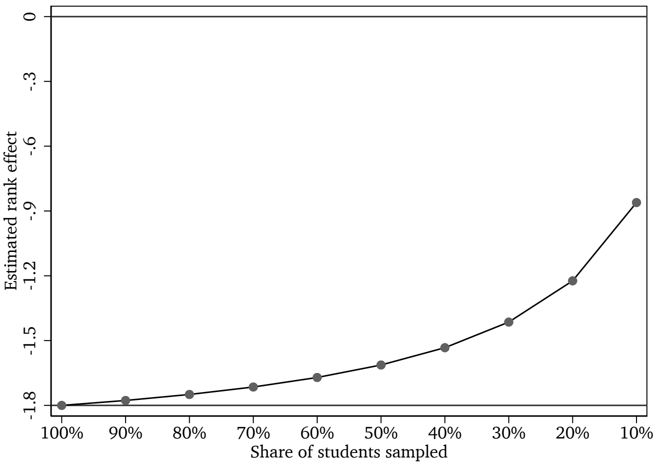
[1] Simulations A through C closely follow analogous simulations reported by Elsner and Isphording (2017).

Table E.1. Simulations to Assess the Consequences of Measurement Error A

Simulation A: Random sampling of students							
DGP: $y = -1.8r - 0.6a$; $a \sim N(0,1)$							
Estimate: $y = \beta r + \gamma a$; select x% of students per school							
	Share of students sampled						
	100%	80%	60%	50%	40%	20%	10%
Rank effect	-1.80	-1.75	-1.67	-1.61	-1.53	-1.22	-0.86
	(100%)	*(97%)*	*(93%)*	*(90%)*	*(85%)*	*(68%)*	*(48%)*

Notes: This table presents the results of Monte Carlo simulations with 1000 repetitions of 500 schools each. Shares in parentheses report the ratio of the estimate to the true coefficient from the data-generating process.

Figure E.1. Simulations to Assess Bias due to Random Sampling within Schools



Notes: This figure presents results from Monte Carlo simulations with 1000 repetitions of 500 schools, as summarized in Table E.1.

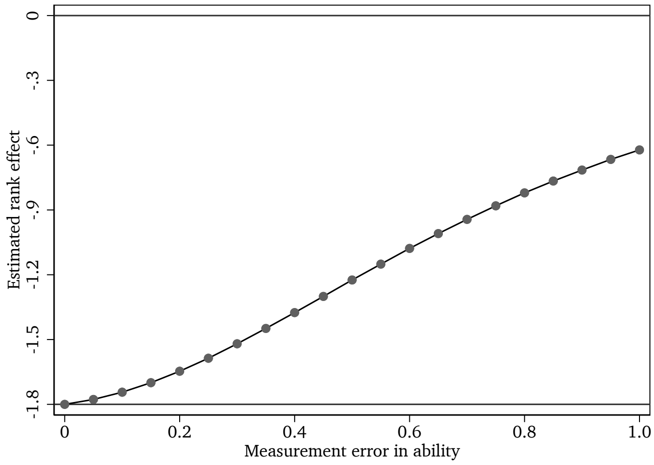
demonstrate that this also leads to classical attenuation bias yielding an underestimation of the true effect.

Table E.2. Simulations to Assess the Consequences of Measurement Error B

Simulation B: Measurement error in ability						
DGP: $y = -1.8r - 0.6a$; $\tilde{a} = a + \phi z$; $a, z \sim N(0,1)$						
Estimate: $y = \beta r(\tilde{a}) + \gamma \tilde{a}$						
	Measurement error (ϕ)					
	0	0.2	0.4	0.6	0.8	1.0
Rank effect	-1.80	-1.65	-1.38	-1.08	-0.82	-0.62
	(100%)	*(91%)*	*(76%)*	*(60%)*	*(46%)*	*(35%)*

Notes: This table presents the results of Monte Carlo simulations with 1000 repetitions of 500 schools each. Shares in parentheses report the ratio of the estimate to the true coefficient from the data-generating process.

Figure E.2. Simulations to Assess Measurement Error in Ability



Notes: This figure presents results from Monte Carlo simulations with 1000 repetitions of 500 schools, as summarized in Table E.2.

C. Omitted variables correlated with ability. Measurement error can also be more complex. For example, there could be an omitted variable z correlated with a that also exerts a direct effect on mental health, y. We model this by extending the data-generating process in equation (4) as follows:

$$y = -1.8r - 0.6\tilde{a} + \rho z \text{ with } \tilde{a} = a + \phi z \text{ and } z \sim N(0, 1). \tag{5}$$

Here, z has a direct effect on y, measured by ρ, and z is correlated with ability $\tilde{a}$. In our estimations, z is unobserved and hence potentially induces a bias in our estimates of r. For the simulations, we change both the strength of the direct effect, ρ, as well as the correlation induced by ϕ. Moreover, we differentiate between cases in which the rank in the data-generating process is based on measured ability ($r = r(\tilde{a})$) or is based on actual ability ($r = r(a)$). The simulations in Table E.3, as well as Figures E.3a and E.3b, reveal that if the rank is based on measured ability $\tilde{a}$ and we control for $\tilde{a}$, then the estimates of r are unbiased. If r is based on actual ability a, then we observe attenuated estimates similar to the measurement error in ability considered in B.

D. Unobserved ability sorting within cohorts. A next set of simulations considers that we only observe students at the cohort level, but have no information on class assignments. Yet, students may be allocated into classes based on their ability. That is, schools may employ tracking into different classrooms. A key assumption for our simulations is that peers affecting y are those peers students interact with, i.e., only those who are in the same classroom. In other words, the rank in the data-generating process depends on the rank within the class, while we as econometricians only observe the rank in the cohort. Table E.4 and Figure E.4 consider how the estimated rank effect varies with classroom allocations. We start with purely

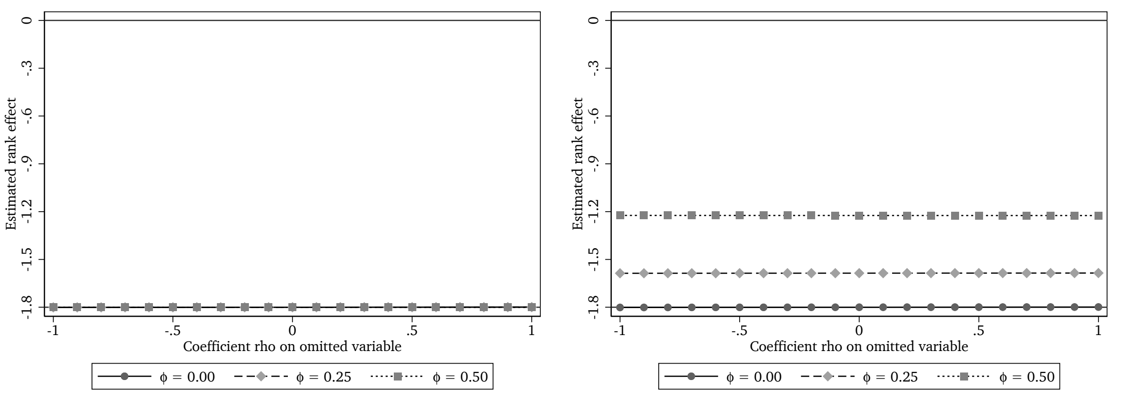
Table E.3. Simulations to Assess the Consequences of Measurement Error C

Simulation C: Omitted variables correlated with ability					
DGP: $y = -1.8r - 0.6\tilde{a} + \rho z$; $\tilde{a} = a + \phi z$; $a, z \sim N(0,1)$					
Estimate: $y = \beta r(\tilde{a}) + \gamma \tilde{a}$					
	Direct effect ρ of omitted variable				
	-1.00	-0.50	0.00	0.50	1.00
(i) Rank based on measured ability ($r = r(\tilde{a})$)					
Rank effect ($\phi = 0.00$)	-1.80	-1.80	-1.80	-1.80	-1.80
	(100%)	*(100%)*	*(100%)*	*(100%)*	*(100%)*
Rank effect ($\phi = 0.25$)	-1.80	-1.80	-1.80	-1.80	-1.80
	(100%)	*(100%)*	*(100%)*	*(100%)*	*(100%)*
Rank effect ($\phi = 0.50$)	-1.80	-1.80	-1.80	-1.80	-1.80
	(100%)	*(100%)*	*(100%)*	*(100%)*	*(100%)*
(ii) Rank based on actual ability ($r = r(a)$)					
Rank effect ($\phi = 0.00$)	-1.80	-1.80	-1.80	-1.80	-1.80
	(100%)	*(100%)*	*(100%)*	*(100%)*	*(100%)*
Rank effect ($\phi = 0.25$)	-1.59	-1.59	-1.59	-1.59	-1.59
	(88%)	*(88%)*	*(88%)*	*(88%)*	*(88%)*
Rank effect ($\phi = 0.50$)	-1.22	-1.22	-1.22	-1.22	-1.22
	(68%)	*(68%)*	*(68%)*	*(68%)*	*(68%)*

Notes: This table presents the results of Monte Carlo simulations with 1000 repetitions of 500 schools each. Shares in parentheses report the ratio of the estimate to the true coefficient from the data-generating process.

Figure E.3. Simulations to Assess Biases from Omitted Variables ($r = r(\tilde{a})$)

(a) $r = r(\tilde{a})$ **(b)** $r = r(a)$



Notes: This figure presents results from Monte Carlo simulations with 1000 repetitions of 500 schools, as summarized in Table E.3.

random ($\omega = 0$) allocation and gradually move towards perfect tracking ($\omega = 1$). In addition, we check how the estimates vary if ability is measured with error (i.e., we observe $\tilde{a} = a + \phi z$, $z \sim N(0, 1)$), while schools may have better information and base their tracking on true ability a.

Our simulations show that tracking policies within cohorts strongly bias the estimated coefficients when tracking becomes sufficiently strong. Random assignment to classrooms on average leads to unbiased rank effect estimates, and the effects are attenuated for small and moderate weights on tracking. Yet, if within-school tracking is sufficiently strong, the sign of the coefficient flips and perfect tracking yields a coefficient of the same size, but the opposite sign compared to the original coefficient. These relationships are dampened once we allow for measurement error in the ability measure, but the same pattern persists.

Table E.4. Simulations to Assess the Consequences of Measurement Error D

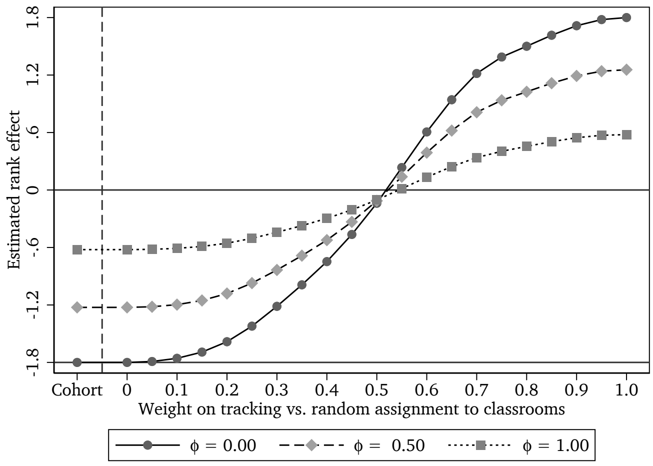
Simulation D: Sorting within cohorts
DGP: $y = -1.8r_{class} - 0.6a$; $a \sim N(0,1)$; r depends on rank in each of 6 classrooms; assignment to classrooms partly based on tracking
Estimate: $y = \beta r_{cohort} + \gamma a$

		Weight ω on tracking vs. random assignment				
	Cohort	0.00	0.25	0.50	0.75	1.00
Rank effect (ϕ = 0.00)	-1.80	-1.80	-1.42	-0.14	1.39	1.80
	(100%)	*(100%)*	*(79%)*	*(8%)*	*(-77%)*	*(-100%)*
Rank effect (ϕ = 0.25)	-1.59	-1.59	-1.25	-0.11	1.26	1.64
	(88%)	*(88%)*	*(69%)*	*(6%)*	*(-70%)*	*(-91%)*
Rank effect (ϕ = 0.50)	-1.23	-1.23	-0.97	-0.11	0.94	1.26
	(68%)	*(68%)*	*(54%)*	*(6%)*	*(-52%)*	*(-70%)*
Rank effect (ϕ = 0.75)	-0.88	-0.88	-0.71	-0.11	0.63	0.87
	(49%)	*(49%)*	*(39%)*	*(6%)*	*(-35%)*	*(-48%)*
Rank effect (ϕ = 1.00)	-0.62	-0.62	-0.50	-0.10	0.40	0.58
	(35%)	*(35%)*	*(28%)*	*(6%)*	*(-22%)*	*(-32%)*

Notes: This table presents the results of Monte Carlo simulations with 1000 repetitions of 500 schools each. Shares in parentheses report the ratio of the estimate to the true coefficient from the data-generating process.

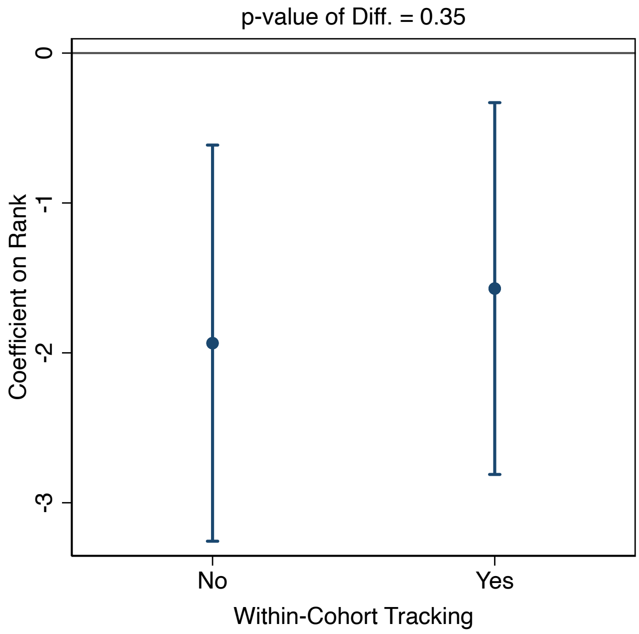
It is ultimately an empirical question how severe the bias is and whether the coefficient flips its sign. To get a sense of the effects of sorting, we lever information from the school administrators' questionnaire, which asked about ability stratification in English and Language Arts classes ("For English or language arts, does your school group classes according to ability or achievement?"). We use the administrators' responses to this question as proxies for tracking. In fact, about 50% of all schools in our sample report using some form of tracking. To gauge the extent of tracking, we estimate our main specification again, but interact the rank variable with an indicator for tracking to separate tracking and non-tracking schools. Figure E.5 (replicating Panel A of Figure 7b) shows that the effect for ranks in schools that employ tracking is indeed attenuated, but still of the same sign and significant. Moreover, we cannot reject that the difference between the rank effects of tracking and non-tracking schools is zero.

Figure E.4. Simulations to Assess Bias due to Sorting within Cohorts



Notes: This figure presents results from Monte Carlo simulations with 1000 repetitions of 500 schools, as summarized in Table E.4.

Figure E.5. Heterogeneous rank effects by within-school tracking



Notes: This figure presents heterogeneous effects by within-school tracking on CES-D scores, including 90% confidence intervals clustered at the school level. We interact the rank variable with an indicators equal to one if the school administrator reports that the school employs tracking and zero otherwise.

E. Measurement error in dependent variable. We now assess the extent to which having access to short scales for the dependent variable y affects our estimates. While mental health is a continuous concept, we observe several noisy measures (i.e., different facets) coded on a discrete 0-3 scale and aggregate them to a composite CES-D score. Table E.5 and Figure E.6 show that, while this does not bias our estimates, it increases the standard errors by about

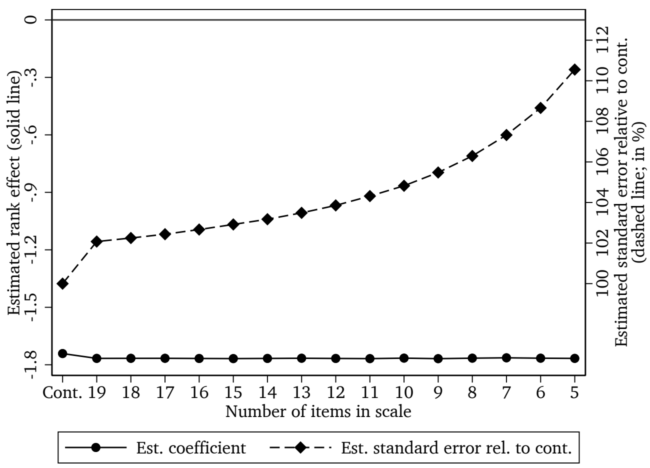
10% when moving from a full scale of 19 items to a short scale of 5 items, as used in Wave V of AddHealth.

Table E.5. Simulations to Assess the Consequences of Measurement Error E

Simulations E: Measurement error in dep. variable due to short scales					
DGP: $y = \sum_i^I y_i$, where $y_i = -0.14r - 0.003a + x + \epsilon_i$; $a \sim N(0, 1)$;					
$x \sim LN$ with $E[x] = 0.65$, $SD(x) = 1$; $\epsilon \sim N(0, 0.8)$;					
y_i rounded to $\{0, 1, 2, 3\}$					
Estimate: $y = \beta r + \gamma a$					
	Number of items in scale				
	Cont.	19	10	9	5
Rank effect	-1.74	-1.77	-1.77	-1.77	-1.77
	(100%)	*(101%)*	*(101%)*	*(102%)*	*(101%)*
Relative standard error	(100%)	(102%)	(105%)	(104%)	(111%)

Notes: This table presents the results of Monte Carlo simulations with 1000 repetitions of 500 schools each. Shares in parentheses report the ratio of the estimate to the true coefficient from the data-generating process.

Figure E.6. Simulations to Assess Loss of Efficiency due to Measurement Error in Dependent Variable



Notes: This figure presents results from Monte Carlo simulations with 1000 repetitions of 500 schools, as summarized in Table E.5.

F. Interaction of rank and negative shocks in the presence of measurement error. Finally, we want to extend the simulations of measurement error in our ability measure (see Simulations B above) and explore its role when we study the interaction of ranks and negative shocks as in Section 6.3. Both the rank as well as our definition of a negative shock are based on the potentially noisy measure of ability. As this measurement error affects each of the variables as well as their interactions, the consequences for our effects of interest (the interaction of rank and the indicator of negative shocks) is ambiguous. In Table E.6 and Figure E.7, we therefore study the consequences of measurement error on our estimates. We find that, for small to medium-sized measurement errors ($\phi \in [0, 0.5]$), we overestimate the rank effect and only underestimate the effect for larger measurement errors ($\phi \in (0.5, 1.0]$). Our estimates of the effects of negative shocks, as well as the interaction of ranks and negative shocks are consistently attenuated towards zero, implying that we underestimate the true effect.

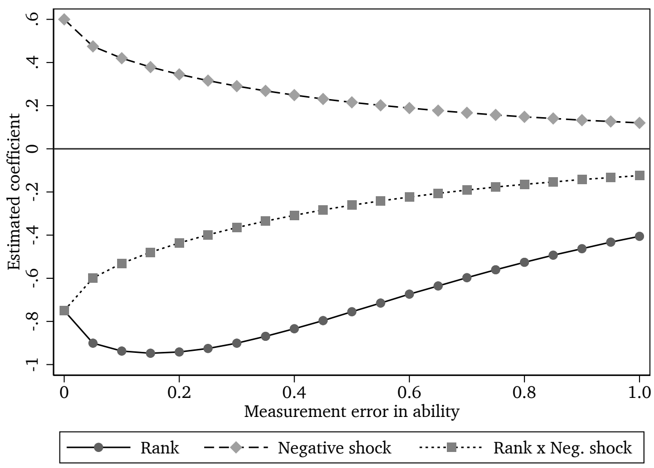
Table E.6. Simulations to Assess the Consequences of Measurement Error F

Simulation F: Measurement error in ability II
DGP: $y = -0.75r + 0.6ns - 0.75(r \times ns) - 0.6a$; $\tilde{a} = a + \phi z$; $a, z \sim N(0, 1)$; $ns = \mathbb{1}\{r(a;\ \text{local}) < r(a;\ \text{global})\}$
Estimate: $y = \beta r(\tilde{a}) + \lambda ns(\tilde{a}) + \mu[r(\tilde{a}) \times ns(\tilde{a})] + \gamma \tilde{a}$

	Measurement error (ϕ)					
	0	0.2	0.4	0.6	0.8	1.0
Rank effect	-0.75	-0.94	-0.83	-0.67	-0.53	-0.41
	(100%)	*(126%)*	*(111%)*	*(90%)*	*(70%)*	*(54%)*
Negative shock	0.60	0.35	0.25	0.19	0.15	0.12
	(100%)	*(58%)*	*(41%)*	*(31%)*	*(25%)*	*(20%)*
Rank × Neg. shock	-0.75	-0.44	-0.31	-0.22	-0.16	-0.12
	(100%)	*(58%)*	*(41%)*	*(30%)*	*(22%)*	*(17%)*

Notes: This table presents the results of Monte Carlo simulations with 1000 repetitions of 500 schools each. Shares in parentheses report the ratio of the estimate to the true coefficient from the data-generating process.

Figure E.7. Simulations to Assess Measurement Error in Ability in Regressions with Interactions



Notes: This figure presents results from Monte Carlo simulations with 1000 repetitions of 500 schools, as summarized in Table E.6.