Peers and Mental Health*

Lukas Kiessling Jonathan Norris

February 26, 2020

Abstract

In this paper, we study the effects of the school environment on students' mental health. Guided by a theoretical framework, we find that having a better rank in school improves students' immediate mental health. In particular, moving a student from the 25th to the 75th percentile improves her mental health by 10-11% of a standard deviation conditional on ability and the peer composition. These effects are more pronounced for low-ability students, persistent for more than 20 years, and carry over to economic long-run outcomes. Moreover, we document a strong asymmetry in our effects: Our results are driven by individuals receiving negative rather than positive shocks. Our results therefore provide evidence on how features of the school environment can have long-lasting consequences for individuals' well-being.

Keywords: Peer effects, Mental health, 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 students' mental health 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 her own ability, which we interpret as mental health, 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 of depression (Seligman, 1972), 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 cohorts within schools. More specifically, we argue that conditional on a student's school, the ability composition of peers and therefore a student's rank in her cohort is as good as random. We then interpret these as exogenous shocks affecting a student's mental health. 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). In line with previous research and our model, we document 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. This further motivates our analysis on the long-term consequences of shocks occurring during adolescence.

We find that these information shocks measured by students' ordinal ranks in their cohorts affect mental health. Increasing the rank from the 25th to the 75th percentile improves a student's mental health by 10-11% of a standard deviation on average. This effect is similar to a one standard deviation increase in a student's ability, comparable to rank effects estimated for standardized test scores of British students (Murphy and Weinhardt, forthcoming), and confirms the central prediction of our model.

We then explore the robustness of this result. We find that our estimates do not change once we include additional controls such as peer effects in several dimensions, a sensitivity analysis based on Oster (2019) shows that unobservables are unlikely to drive our estimates, and a series of Monte Carlo simulations suggests that measurement error leads us to underestimate the magnitude of the actual effects. Moreover, a second identification strategy with grade-by-school fixed effects yields similar results.[1]

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.

Moving to our third prediction, we find that this effect is larger at the lower end of the ability distribution: Students from the lowest decile of the ability distribution improve their mental health by 35% of a standard deviation when increasing the ordinal rank from the 25th to the 75th percentile, while this effect slowly fades out for higher deciles. When studying further heterogeneities, we find suggestive evidence that rank effects are more pronounced for females, non-white students, and for students with non-college educated parents, as well as for students from larger cohorts with high dispersion in ability, but unfortunately lack statistical power to provide concluding evidence.

These heterogeneous effects in own ability are persistent over time and last from adolescence to adulthood. In fact, we observe that rank effects 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. Using data from wave V, we find this pattern to persist after 23 years, but due to a smaller sample can only provide suggestive evidence.

When studying a range of economic long-run outcomes, we observe that mental health during adolescent is significantly associated to college graduation, employment status, income as well as paying bills on time. Moreover, we document causal effects of ordinal ranks in school on having a college degree as well as income at age 36-42, and mental health during adolescence mediates 8.9-19.7% of these effects. Taken together, our findings provide strong

[1]This strategy 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.

support for the fourth prediction of our theoretical framework that the school environment can have persistent effects on mental health and lend support for models introducing mental health capital similar to general health (Grossman, 1972).

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 many 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., forthcoming) on mental health. Second, mental disorders can be a cause of important economic outcomes such as limited human capital accumulation (Currie and Stabile, 2006; Fletcher, 2010) or reductions in employment and earnings (Bartel and Taubman, 1986; Frank and Gertler, 1991; Ettner, Frank, and Kessler, 1997; Stewart et al., 2003; Fletcher, 2014). Third, it 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, our focus on mental health relates to the growing evidence on the importance of non-cognitive skills in the development of children (Cunha and Heckman, 2008; Cunha, Heckman, and Schennach, 2010) and for subsequent education and labor market outcomes (Heckman, Stixrud, and Urzua, 2006).

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; Krueger and Whitmore, 2001; Chetty et al., 2011) or better teachers (e.g., Chetty, Friedman, and Rockoff, 2014; Rothstein, 2017) are a long-standing and active field 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 Bietenbeck (2019) shows that low-achieving peers can increase non-cognitive skills. Our results highlight the long-term effects of ordinal ranks for psychological well-being that in turn may affect educational attainment, labor market outcomes as well as general health. Our results therefore provide further evidence on how school environment and peer composition shapes the well-being of individuals in the long-term.

Finally, we contribute to the literature studying the consequences of peers in school more generally and the literature on rank effects in specific.[2] 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 individuals' job satisfaction or general well-being (Card et al., 2012;

[2]Rank effects are a specific form of peer effects, in which the ordinal rank rather than the mean (for an overview, see, e.g., Sacerdote, 2011) or the variance (Tincani, 2017) affects outcomes. There exists a small literature on 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.

Luttmer, 2005; Brown et al., 2008).[3] Recent papers have used these ideas in educational contexts to estimate the effect of ordinal ranks on educational outcomes (Murphy and Weinhardt, forthcoming; Elsner and Isphording, 2017; Elsner, Isphording, and Zölitz, 2019; Delaney and Devereux, 2019) 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 as these previous papers and guide our analysis using a theoretical framework based on learning about one's own ability in the presence of noisy signals rather than social comparisons. 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 and hence this yields either positive or negative shocks relative to a rank one might expect. In our application, we find that our rank effects on mental health are driven by negative shocks, in line with what we would expect based on our model. In addition, we also show that ranks affect several long-run outcomes such as college graduation and earnings, which also have a persistent relation to mental health measured during adolescence.

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 based 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 how different shocks – in our empirical application information shocks from quasi-exogenous variation in students' ordinal ranks – may affect students' mental well-being. 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 depressions, our model closely follows de Quidt and Haushofer (2016) and conceptualizes the belief about the returns to one's own ability as a proxy for mental health. In fact, many common symptoms of depressions, 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).

[3]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.

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; Zafar, 2011; Stinebrickner and Stinebrickner, 2012). Second, there are exogenous shocks affecting school performance and students use their performance to update their prior about their ability. Thus, receiving a negative signal reduces her belief about her ability, while positive signals increases her belief. 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 a student's school success depends on the amount she studies, s_t, and her decision to exert high ($e_t = 1$) or low effort ($e_t = 0$). High effort yields a return to studying equal to her 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 $\epsilon_t \sim N(0, \sigma^2_{\epsilon,t})$. 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 signal y_t.[4]

The student is uncertain about her own ability and has a prior or belief about her ability denoted by μ_t. Hence, from a student's perspective, her ability is a random variable $A \sim N(\mu_t, \sigma^2_{A,t-1})$. We assume that a student maximizes her expected utility by allocating time between studying and leisure. While studying increases her educational success, leisure also enters positively into her utility function. Her expected decision utility function is given by

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

where 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 her ability, μ_t, and the decision to exert effort, e_t, as described above.

[4]More specifically, we exploit 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 in would be only a mediocre student in another cohort. We use this variation in the ability distribution as an exogenous shocks affecting students' beliefs.

Given this setup, a students optimal effort decision is

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

and, hence, we can replace $[\mu_t e_t^* + \underline{A}(1 - e_t^*)] = \max\{\mu_t, \underline{A}\}$.[5] Optimal time spend studying therefore equals

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

and is increasing in perceived own ability.

We now want to characterize how a student learns about her ability. Consider a student who wants to update her prior belief μ_{t-1} given she received a signal y_{t-1} about her own ability. If the student only exerts low study effort, $e_{t-1} = 0$, she does not learn new information about her ability A as studying yields a fixed return $\underline{A}$. If she exerts high effort ($e_{t-1} = 1$), she can learn about her 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 $x_{t-1} = As_{t-1} \sim N(\mu_{t-1}s_{t-1}, \sigma^2_{A,t-1}s^2_{t-1})$. Given the signal about school success, y_{t-1}, the student tries to learn about her ability, A. Using the new notation, she wants to infer the expected value of x_{t-1} given y_{t-1}, i.e., the posterior $E[x_{t-1}|y_{t-1}]$:

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

Hence, the corresponding posterior belief μ_t is:

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

Several results emerge. First, a negative shock ($\epsilon_{t-1} < 0$) decreases a student's belief about her own ability (i.e., μ_t decreases) and thus has detrimental effects on her mental health, while a positive shock ($\epsilon_{t-1} > 0$) benefits mental health.

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

[5]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 her beliefs.[6] This implies that negative shocks may have more pronounced consequences relatively to positive shocks as negative shocks decrease the likelihood that a student receives informative signals in the future.

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

Third, a student's study time, s_t, (weakly) decreases in the belief about her 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 a student's school success decrease her belief about her own ability, these effects are more pronounced in the lower part of the ability distribution, and the consequences 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 identification of ordinal rank effect. 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' wellbeing 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 these two key measures in more detail.

[6] 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.

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 disorder 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 depressions (Radloff, 1977). Appendix Table A1 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 A1: 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 comprising 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.[7]

3.2 Constructing a Student's Ordinal Rank Measure

To test the predictions outlined above, we want to lever a shock to 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 a student's ordinal rank in her school cohort as such an information shock, which captures information about the relative standing within a cohort. We construct the ordinal rank based on an ability assessment that 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,

[7]Using data from wave I, Appendix Figure A2 shows that the short and long versions of the CES-D scale indeed are highly correlated. To perform the rescaling, we scale the nine (ten, five) item scales of wave III (IV, V) by 19/9 (19/10, 19/5) to match the 19 item scale.

2007) that was administered as part of wave I and provides us with an objective, age-specific, and standardized ability measure.

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 ability within their cohort by assigning them an absolute rank based on their ability.[8] 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. We illustrate how this ordinal rank varies with a student's ability in Figure 1. 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 her mental health holding her ability constant, we need sufficient variation in ranks for a given ability level. Figure 1 provides some evidence that this is indeed the case: for each ability decile in the global ability distribution, we observe variation in a student's local rank. However, conditioning on various variables, most notably own ability, will likely capture some of the variation depicted in Figure 1. In Section 5, we address this issue and study whether there is sufficient variation in the rank variable after conditioning on a set of controls suggested by our empirical strategy. We find that variation in ranks across the ability distribution, although reduced, continues to hold.

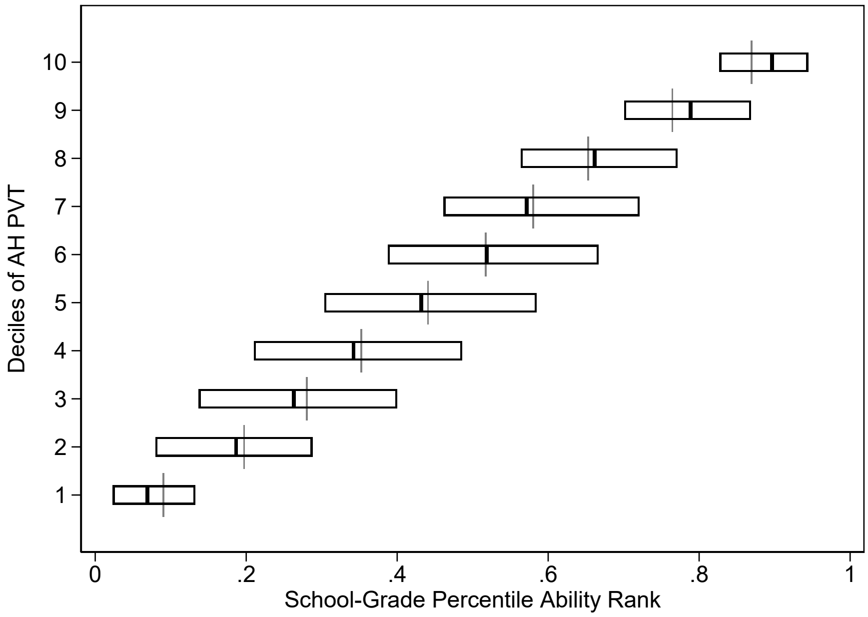
A potential confound for the interpretation of our rank measure based on AddHealth's Picture Vocabulary Test is that neither students nor teachers learn the results of this test.[9] Thus, the question remains how salient is our rank measure. We evaluate this by studying the relationship of ranks based on our ability measure and students' self-assessment about their relative ability as well as their desires and expectations for attending college. We report these results in Appendix Table A3. 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, 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

[8]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 her ability. If two students have the same ability, they are assigned an equal rank.

[9]Alternatively, we could have used a student's GPA to calculate ranks. Yet, this measure would have major 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 students' GPA may be affected by classical peer effects.

Figure 1. 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, black boxes illustrate the 25th, 50th, and 75 percentile of ordinal ranks, while gray lines indicate the mean ranks.

as predetermined and unaffected by features of the school environment and students' own, or their parents', investments.

Finally, we only observe a random sample of students in each school cohort introducing additional sampling variation in our data. In Appendix B, we report results from a simulation study showing that such sampling variation leads us our estimates to be attenuated by approximately a third if we observe only about 10% of the students in a cohort. We therefore think that this concern is relatively small. Taken together, defining ordinal ranks based on AddHealth's Picture Vocabulary Test yields a measure that is pre-determined, salient, and comparable across cohorts as well as schools.

3.3 Summary Statistics

We present summary statistics of our sample in Table 1. In total, we observe 18454 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.

4 Stylized Facts on Mental Health over the Life-Cycle

We begin by documenting evolution and persistence of mental health over the life-cycle using the rich information from the AddHealth study. Mental health manifests early in life and stays

Table 1. Summary Statistics

	Mean	SD	Min	Max
CES-D Wave I	11.33	7.60	0.00	56.00
Ordinal Rank	0.47	0.28	0.00	1.00
Ability (AH PVT scores)	100.17	14.66	13.00	139.00
Female	0.51	0.50	0.00	1.00
Age	15.63	1.70	11.00	19.00
Ethnicity				
White	0.53	0.50	0.00	1.00
Black	0.22	0.41	0.00	1.00
Hispanic	0.17	0.37	0.00	1.00
Asian	0.07	0.25	0.00	1.00
Other	0.02	0.14	0.00	1.00
Parental Background				
Less 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
Post-Graduate	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	18454			

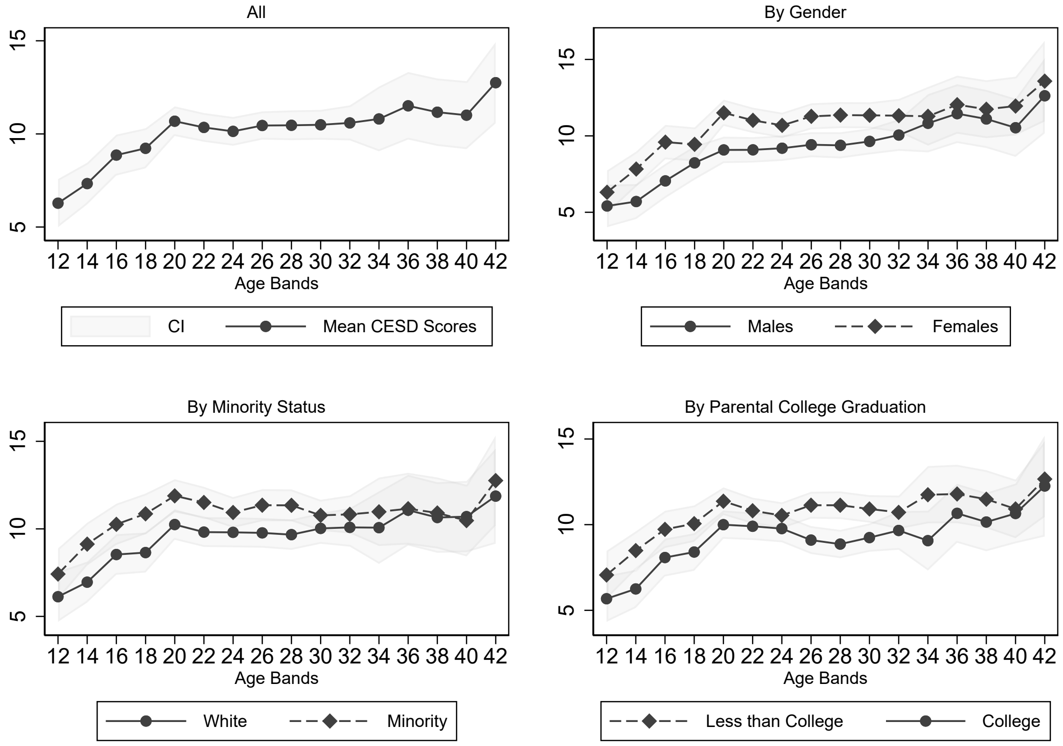
Notes: This table presents summary statistics for the sample used in the analysis.

persistent over the life-cycle. These stylized facts highlight the importance of studying the features of the school environment as determinants of mental health, and thus motivate our subsequent analysis.

Evolution of Mental Health over the Life-Cycle In order to investigate the evolution of mental health over time, we lever one feature of the AddHealth data that they cover respondents from different ages ranging from 12 to 18 in wave I. Although there are several years between the data collections of different waves, the is a partial overlap in ages covered by different waves. This allows us to aggregate age-specific mental health measures across waves.[10] To increase precision, we aggregate age groups into two year bins and trace the evolution of mental health measured by CES-D scores over the life-cycle.

[10]To do so, we need to rescale CES-D scores in later waves, which only elicit a subset of the initial scale as illustrated in Table A1. See the discussion in Footnote 7 on how we perform the scaling.

Figure 2. Evolution of Mental Health over the Life-Cycle

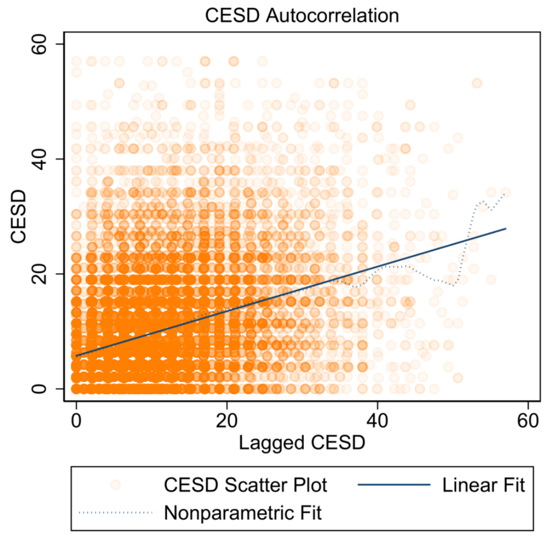


Notes: This figure presents mean CES-D scores at 2-year age bands after pooling over waves I, III, IV, and V and controlling for survey wave effects. Shaded area indicates 90% confidence intervals.

The first panel of Figure 2 displays the evolution of CES-D scores for the whole sample. 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 2, we differentiate the evolution of mental health by gender, ethnicity, and socioeconomic status of the respondents. While there is cross-sectional variation in mental health with females, minorities, and respondents with non-college educated parents having higher CES-D scores and thus worse mental health, the evolution over the life-cycle is similar across subgroups. In particular, we observe the same steep increase in CES-D scores until age 20 and a relatively flat pattern afterwards.

Persistence of Mental Health over Time While Figure 2 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 (we use waves I, III, IV, and V and omit wave II as it is one year after the baseline and only conducted for the subsample of students who have not left high school). Figure 3 presents the distribution of CES-D scores in every wave 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 indicating a strong persistence of mental health over time.

Figure 3. 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 at 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).

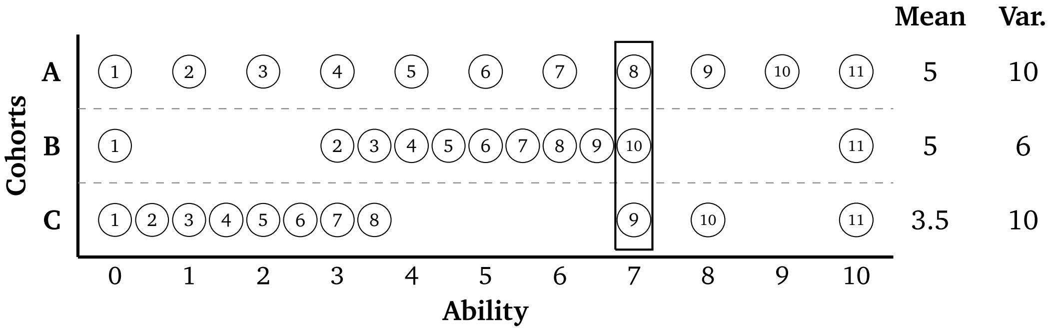
Together, the figures provide a first indication that the school environment may have a lasting effect on the mental health of students: Mental health seems to be malleable during adolescence and remains persistent over time.

5 Empirical Strategy

Our aim is to estimate the causal effect of a specific feature of the school environment: how does a student's ordinal rank in her cohort affects her mental health holding both own and peer ability constant. Before we describe our identification strategy, 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 having 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 within 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 first identification strategy. We follow Hoxby (2000a,b) and exploit the idiosyncratic variation ability across cohorts within

Figure 4. Illustrative Example of the Identifying Variation



Notes: This figure illustrates how variations in the ability distribution across cohorts allows 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. 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 also can give rise to different ranks.

the same school. This motivates the following empirical specification:

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

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 her 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-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, post-graduate degree), and an indicator for being raised in a single parent household. The term θ_{isc} consists of school, δ_s, and cohort fixed effects, γ_c, to capture unobserved heterogeneity by school and cohort. Standard errors are clustered at the school level.

One concern with equation (2) is that a student's ordinal rank is related to the average ability within the cohort. As we are interested in the causal effect of ordinal ranks, we do not want our rank measure to be confounded by typical peer effects. We therefore add the mean peer ability, $\overline{a}_{-ics}$ as an additional control variable.[11] In our first empirical specification, we therefore control for $\theta_{ics} = \lambda\overline{a}_{-ics} + \gamma_c + \delta_s$. In refinements, we add further linear-in-means peer effects and standard deviations in these peer characteristics to capture other peer effect dimensions and potential non-linearities.

In a second 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., $\theta_{ics} = \zeta_{cs} = \gamma_c \times \delta_s$. Using these cohort-by-school fixed effects, we absorb any potential

[11] To calculate the mean and standard deviation of peer ability, we exclude student i.

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 and compare it to other cohorts and 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 our first specification, we assume that that these cohort-level controls are given by separate school and cohort fixed effects, as well as peer effects in students ability. Using these and individual controls, we compare students in the same school and cohort, with similar peers, and with the same observable characteristics and own-ability but who have different ranks.

Nonetheless, there might be other factors potentially affecting a student's mental health and her 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, we adopt a second specification with school-specific cohort fixed effects, which absorb all observable and unobservable differences between cohorts and 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 controls and different fixed effects. The standard variation in ranks without controls is 0.28 as shown in Table 1. However, since a student's rank and ability are positively correlated as indicated by Figure 1, some part of the variation may just 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 potential confounds and compare this to the raw standard deviation in ranks. Appendix Table A2 shows that the raw standard deviation in ranks by decile varies between 0.09 and 0.18 (see also Figure 1 for a graphical representation of the raw variation in ranks by ability decile). 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. We therefore think that the residual variation in ordinal ranks is sufficient to study their causal effect on mental health.

6 Results

How does a student's ordinal rank causally affects her mental health? Our theoretical framework in Section 2 generates four predictions: First, we should observe that positive (negative) shocks, in our application a higher rank, 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 her mental health. More specifically, we relate a student's mental health measured by CES-D scores to her ordinal rank based on our main specification in equation (2) with standard errors clustered on the school-grade level. We present our results in Table 2. 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 students' mental health. Moving a student from the 25th percentile to the 75th improves her mental health by 0.8 CES-D points or 0.11SD. We interpret this as a large effect since we compare students with the same observable characteristics, in particular the same ability, and the same average cohort ability that only happen to be in a cohort where they have different ranks.

Allowing for peer effects in ability only is potentially restrictive. In particular, the literature on linear-in-means peer effects has identified a range of different peer characteristics that causally affect a student's performance and thereby may also affect her mental health. Examples include the share of females, minorities, or students with high socioeconomic status (Lavy and Schlosser, 2011; Hoxby, 2000b; Cools, Fernández, and Patacchini, 2019). 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 nonlinear peer effects. Our estimates show that the rank effect is robust to the inclusion of these additional peer effects and vary only slightly.

In column (4), we adopt our second empirical specification using grade-by-school fixed effects. 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 2 document that the ordinal rank of students causally affect their mental health measured by CES-D scores in line with the central prediction of our model. The estimated effects are comparable to increasing a student's ability by approximately one standard deviation or half of the effect of losing one's job (Marcus, 2013). Taking a one

Table 2. Average Effect of Ordinal Ranks on Mental Health

	Mental Health (CES-D score)			
	(1)	(2)	(3)	(4)
Panel A: Baseline effects				
Rank	-1.58**	-1.58**	-1.60**	-1.68**
	(0.75)	(0.75)	(0.76)	(0.76)
Individual Controls	Yes	Yes	Yes	Yes
Ability Peer Effects (mean)	Yes	Yes	Yes	No
Further Peer Effects (mean)	No	Yes	Yes	No
Ability Peer Effects (SD)	No	No	Yes	No
Further Peer Effects (SD)	No	No	Yes	No
School and Grade FE	Yes	Yes	Yes	No
School $\times$ Grade FE	No	No	No	Yes
N	18454	18454	18454	18454
R^2	0.108	0.108	0.109	0.127
Panel B: Standardized Effects				
Standardized effect	-0.21**	-0.21**	-0.21**	-0.22**
	(0.10)	(0.10)	(0.10)	(0.10)
Panel C: Role of Unobservables				
Oster's δ ($R^2_{max} = 1.3R^2$)	-1.02	-1.2	-1.9	-1.58

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are in parentheses are clustered at the school level. 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-grade 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 below each specification. Oster's δ 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.

standard deviation increase in ranks, the effects correspond to a 0.06 standard deviation increase in CES-D scores and are there about the same size as the effect of ordinal ranks on standardized test scores of British students (0.08SD; Murphy and Weinhardt, forthcoming).

6.2 Robustness Checks

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

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 second identification strategy 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 need 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 unobservable 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 2, δ is larger than one in all specifications. Since we controlled for arguably the most important factors that could bias students' ordinal ranks and that may affect mental health through differences in the cohort composition, we conclude that unobservables are unlikely to drive our estimated rank effects.

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 A4, we therefore examine different polynomials up to a fifth order. We find that using linear or quadratic controls in ability increase our estimated coefficient on ordinal ranks and thus would strengthen our main result. The estimates stabilize for higher order polynomials in ability.

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 proposed by Belloni, Chernozhukov, and Hansen (2014). The PDS Lasso penalizes control variables, but provides allows valid inferences 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 suggesting a relatively linear relationship of CES-D scores and ability as also illustrated in Figure A3. More importantly, however, the estimated effects of ordinal ranks are unaffected and, if anything, become more pronounced.

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 based on ability within school cohorts, which we cannot observe. This would imply that we calculate a students' ranks based on incorrect reference groups. 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 the consequences of these concerns using a series of Monte Carlo simulations reported in Appendix B. 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 within school cohorts. As long as the sorting is not too severe, we underestimate the true effect. When sorting within cohorts becomes sufficiently strong, this can yield estimates with the wrong sign. Given our setting in American schools, where in contrast to many European school systems there is no strict class assignment, as well as strict requirements for tracking given by Title VI of the U.S. Department of Education[12], we think it is unlikely to have sufficient unobserved tracking biasing our estimates. 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 – more or less severe 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 a student experiences a negative shock, her 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 more pronounced rank effects for negative than for 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, Gotlib et al. (2004) provide evidence 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, independent of the ability composition of their school cohort.[13] 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., independent of their school s rather than within a school as in the case of $rank_{ics}$. We then define student i receiving a negative shock if her rank in her school

[12]More specifically, "school districts must be able to demonstrate that there is a valid educational justification for their ability grouping or tracking practices" (URL: `https://www2.ed.gov/about/offices/list/ocr/docs/tviassgn.html`).

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

cohort, $rank_{ics}$ is lower than the rank among all students in a given cohort, $rank_{ic}$:

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

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 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 she has 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.[14]

In Table 3, 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 2 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.

Table 3. Asymmetric Effects of Ordinal Ranks

	Mental Health (CES-D score)		
	(1)	(2)	(3)
Rank	-1.58**		-0.73
	(0.75)		(0.91)
Neg. shock		0.36**	0.61**
		(0.17)	(0.29)
Rank			-0.76
× Neg. shock			(0.47)
N	18454	18454	18454
R^2	0.108	0.108	0.108

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are in parentheses and are 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 2.

[14]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 B, 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, is that 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.

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 have more pronounced effects increasing CES-D scores by 0.61 (0.08 SD). Moreover, rank effects are twice as large for students receiving negative shocks compared to those receiving positive shocks, although the difference between positive and negative shocks is not significant at conventional levels ($p = 0.11$).

These results indicate that the effects of ordinal ranks on mental health we observe are due to negative shocks, i.e., they stem from effects of ranks on those who rank lower in their school cohort than expected. In line with Ertac (2011) and evidence from the psychological literature, but contrasting Eil and Rao (2011) and Möbius et al. (2014), we find that students update stronger in case of negative shocks.[15] The asymmetry in effects documented here implies that rank effects on mental health do not 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 would explain our results, we would expect stronger effects for positive shocks, which is the opposite of what we find.[16]

6.4 Heterogeneous Rank Effects

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 ability decile. In a second step, we want to provide a more comprehensive picture by considering heterogeneities in other observable characteristics that may be of interest when designing policies exploiting the effects of ranks.

6.4.1 Heterogeneity by Ability

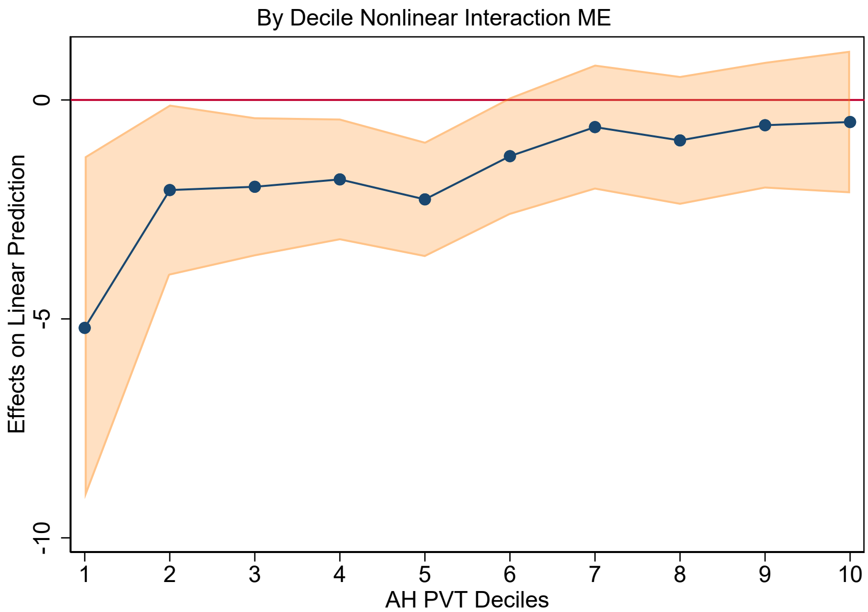
Prediction 3 suggests that the shocks to students at the lower end of the distribution are larger than for higher-ability students because a lower perceived ability decreases study time which subsequently amplifies the consequences of exogenous shocks. If this prediction is correct, we should observe stronger (weaker) rank effects for lower (higher) ability quantiles. We therefore enrich our main specification given in equation (2) by interacting the rank with indicators for each ability decile.

Figure 5 displays the results of this analysis graphically, while Table 4 presents the corresponding regression estimates. We indeed find that rank effects are more pronounced at

[15] 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.

[16] 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 indicating that there mental health is one of potentially several mechanisms affecting economic long-run outcomes.

Figure 5. 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. Shaded area indicates 90% confidence intervals clustered on the school level.

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, corresponds to 0.70SD. 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, which suggests that effects should be more pronounced for low-ability students.

Table 4. Effects of Ordinal Ranks by Ability Decile

	Rank Effect on Mental Health (CES-D scores) by Decile									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Rank	-5.20**	-2.06*	-1.98**	-1.81**	-2.27***	-1.28	-0.62	-0.92	-0.58	-0.50
	(2.36)	(1.18)	(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 are in parentheses and are 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 2.

6.4.2 Further Heterogeneities by Other Individual and Cohort Characteristics

Although our theoretical framework is not aimed at providing predictions for specific subsamples, they are nevertheless important for policy-makers interested in targeting policies and contribute to a better understanding of rank effects on mental health. We focus on three observable characteristics of students – their gender, ethnicity, and socioeconomic status – to identify whether certain groups react more to ranks than others. 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), males are more likely to enter competitions which have rankings as inherent characteristics (Niederle and Vesterlund, 2007). Similarly, individuals from low socio-economic status may be more stressed by social rank concerns (Hackman, Farah, and Meaney, 2010), but students from college-educated parents could also be more receptive to ranks as part of the competition for colleges.

In two further checks, we focus on margins that are subject to frequent debates in the design of schools and classrooms and evaluate their potential consequences for students' mental health. More specifically, we ask whether tracking regimes, or restricting the size of cohorts, impacts rank effects on mental health. There is an active debate about the consequences of tracking regimes (e.g., Duflo, Dupas, and Kremer, 2011; Garlick, 2018), and although we do not observe formal tracking, we exploit one feature of tracking: it creates more homogeneous groups in terms of students' ability. In principle, this increase in homogeneity within cohorts could introduce some more pressure and amplify the consequences of signals about one's own ability. Relatedly, reducing cohort or classroom sizes might have beneficial effects on educational outcomes (Angrist and Lavy, 1999; Krueger, 1999; Hoxby, 2000a; Krueger and Whitmore, 2001), but smaller cohorts and classes also increase the salience of students' ranks as students can easily compare their own performance to that of their peers. In the last two analyses, we therefore consider whether differences in the variance in ability and the cohort size give rise to potential heterogeneous rank effects.

Table 5 presents effects of the ordinal rank by gender (Panel A), minority status (Panel B), as well as socioeconomic status proxied by whether parents attended college or not.[17] Rank effects on mental health seem to be more pronounced among females, non-white students, and students from households without college educated parents, but we lack statistical power to provide concluding evidence (p-values of the differences in coefficients between groups are 0.27, 0.44, and 0.13 for splits by gender, race, and socioeconomic status, respectively).

In the remaining panels of Table 5, we study heterogeneities based on cohort characteristics. More specifically, we perform a median split based on the variance in ability for each cohort (Panel D) and analyze differences by cohort size (Panel E, where we adopt the classification in small, medium, and large as used by Elsner and Isphording, 2017). Consistent with the idea that ranks become more salient in cohorts with higher dispersion in ability, we

[17]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.

Table 5. Heterogeneities by Gender, Race, Socioeconomic Status, and School Characteristics

	Mental Health (CES-D scores)					
	Panel A. By Gender			*Panel B. By Race*		
	Females	Males	p-value (F=M)	White	Non-white	p-value (W=NW)
Rank	-2.22*	-0.68	0.267	-1.03	-2.11**	0.435
	(1.18)	(0.85)		(1.07)	(0.98)	
	Panel C. By Socioeconomic Status			*Panel D. By Ability Variance*		
	College	No College	p-value (C=NC)	Low Var.	High Var.	p-value (L=H)
Rank	0.05	-2.15**	0.130	-1.12	-1.92	0.632
	(1.23)	(0.85)		(1.19)	(1.26)	
	Panel E. By Cohort Size					
	Small (<150)	Medium (150-299)	Large (≥ 300)	p-value (S=M)	p-value (S=L)	p-value (M=L)
Rank	-0.16	-0.67	-2.93***	0.783	0.085	0.157
	(1.17)	(1.41)	(1.05)			

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are in parentheses and are clustered at the school level. We include all base controls, peer ability, school and grade fixed effects as in our preferred baseline specification in column (1) of Table 2.

find suggestive evidence that rank effects are larger in cohorts with high variance, but cannot reject that these coefficients equal each other (p-value = 0.64). When looking at the size of the cohort, we observe that effects are more pronounced for larger cohorts and this difference is significant when comparing small and large cohorts (p-value = 0.09).

Based on these heterogeneous effects, the consequences of being ranked highly or lowly seem to be more pronounced for females, minorities, and students from low socioeconomic backgrounds, but we cannot provide concluding evidence. Policies that try to alleviate rank effects therefore may be targeted at these groups. Perhaps surprisingly, we do not observe that rank effects are more pronounced in small cohorts. To the contrary, we find stronger effects for larger cohorts, maybe because facing more higher or lower ranked students increases the salience of one's own rank.

7 Persistence of Rank Effects

We have established that the ordinal rank exerts a causal effect on students' mental health and this effect is more pronounced for low-ability students. We now want to explore the dynamic effects of ordinal ranks. As there is a significant association between ability and mental health (see Appendix Figure A3 and Appendix Table A4), and students at the lower end of the distribution experience stronger effects, they also 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. In particular, we can study the short-term persistence using wave II, which was conducted one year after wave I, medium-term effects approximately seven years after wave I using data from wave III, and long-term persistence using data from wave IV and V, which were conducted 14 and 23 years after the baseline, when respondents were adults aged 36–42 years. Similar to Section 6.4.1, 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 nine, ten and five of the original items in waves III, IV, and V, respectively. To compare our estimates from all waves to the baseline, we scale the mental health measures from the short scales by $^{19}/_{9}$, $^{19}/_{10}$, and $^{19}/_{5}$ to correspond to the same range from 0 to 57 as the full scale in wave I.[18]

Table 6. Effects of Ordinal Ranks by Ability Decile

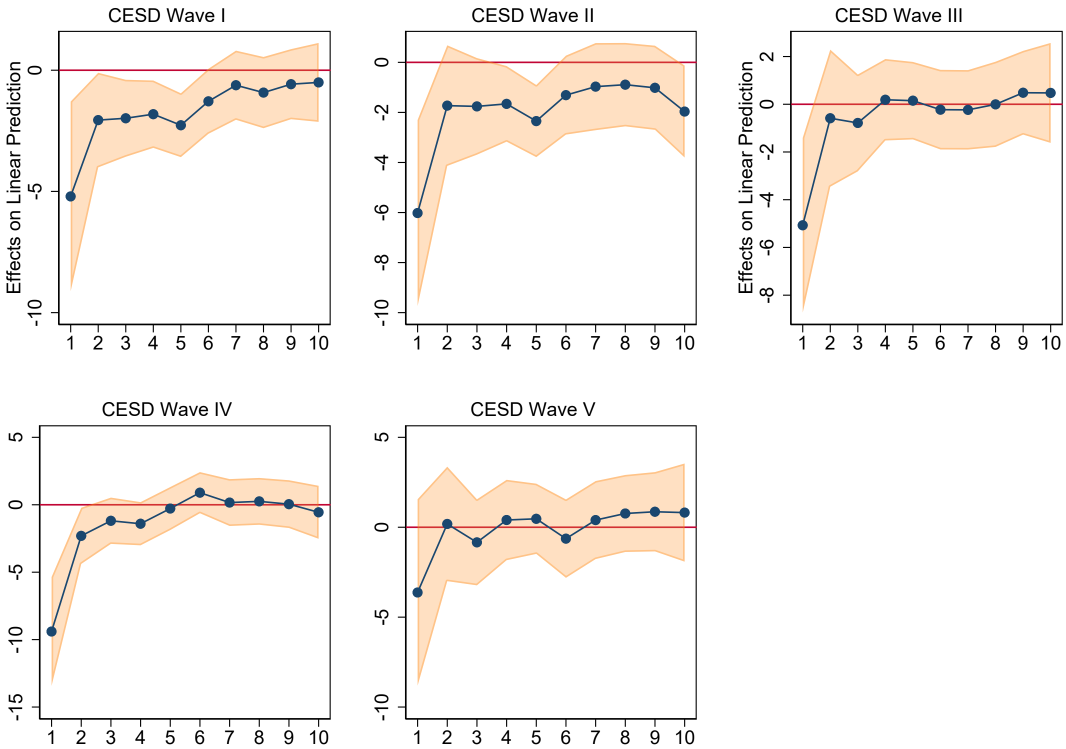
	Effect on Mental Health by Decile										Observations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
Panel A. Immediate Effects (Wave I, 1994/1995)											
Rank	-5.20**	-2.05*	-1.98**	-1.81**	-2.27***	-1.28	-0.62	-0.92	-0.57	-0.50	18454
	(2.36)	(1.18)	(0.96)	(0.84)	(0.80)	(0.81)	(0.86)	(0.89)	(0.87)	(0.98)	
Panel B. Short-term Effects (Wave II, 1996)											
Rank	-6.02***	-1.73	-1.76	-1.66*	-2.35***	-1.31	-0.97	-0.89	-1.02	-1.97*	12892
	(2.24)	(1.45)	(1.17)	(0.91)	(0.87)	(0.95)	(1.05)	(1.01)	(1.01)	(1.11)	
Panel C. Medium-term Effects (Wave III, 2001/2002)											
Rank	-5.07**	-0.58	-0.78	0.19	0.15	-0.23	-0.23	-0.00	0.48	0.48	13541
	(2.21)	(1.73)	(1.22)	(1.03)	(0.98)	(1.01)	(1.01)	(1.08)	(1.06)	(1.26)	
Panel D. Long-term Effects (Wave IV, 2008)											
Rank	-9.40***	-2.30*	-1.19	-1.41	-0.28	0.90	0.16	0.25	0.05	-0.56	14016
	(2.44)	(1.26)	(1.04)	(0.97)	(0.96)	(0.92)	(1.05)	(1.05)	(1.07)	(1.19)	
Panel E. Very Long-term Effects (Wave V, 2016-2018)											
Rank	-3.63	0.18	-0.84	0.40	0.47	-0.63	0.40	0.77	0.86	0.82	11061
	(3.13)	(1.92)	(1.44)	(1.35)	(1.18)	(1.32)	(1.31)	(1.29)	(1.33)	(1.65)	

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses and are 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 2.

Figure 6 shows that the general pattern persists across all five waves: the effect of ordinal ranks in wave I is significant and pronounced at the bottom of the ability distribution and insignificant as well as smaller in magnitude for higher ability deciles. Table 6 quantifies these

[18]As shown in Appendix Figure A2 the short- and long-versions of the CES-D are highly correlated in wave 1. We therefore use scaled version to compare our results to the baseline effects documented in Section 6.4.1. Scaling these short scales reduces the efficiency of our estimates. Simulation F in Appendix B suggests that the standard errors on our variable of interest increases by about 11% for the the 5-item scale.

Figure 6. 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 6. Shaded area indicates 90% confidence intervals clustered on the school-grade level.

effects. Panel A replicates the estimates of Section 6.4.1, while Panels B through E consider the short-, medium-, and long-term effects of ordinal ranks in wave I. We find that the the significant effects for the lowest ability decile persist across waves I to IV and amount to -5.07 to -9.40 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. In wave V, 23 years after baseline, the same pattern persist, but the estimates are insignificant. Partly, this is a power problem due to a smaller sample in wave V and more measurement error in the dependent variable from having only five items in the CES-D scale as shown in simulations in Appendix B. Partly, this may also be due to a fade out of effects. Nevertheless, it is reassuring that the same qualitative pattern holds more than 20 years after the initial shock.[19]

These results are in line with the Prediction 4 suggesting that once a negative shock reduces a student's belief in a her ability sufficiently, she withdraws study effort and therefore avoids new signals. As a consequence, her belief about the returns to ability remains low, positive updating is less likely, and depressions are some form of absorbing states. In other

[19]This also suggests that the effects are unlikely to be driven by outliers in wave I as we would expect mean reversion in subsequent waves.

words, her mental health remains in a poor state and negative shocks may trigger potential vicious cycles. The school environment therefore can have long-lasting effects on the mental well-being of students over the life-cycle.

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). We conduct two analyses to shed light on this issue. First, we assess the correlation between CES-D scores in wave I and 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 ranks are mediated by mental health.

In column (1) of Table 7, we show that a range of economic outcomes – graduating from college, being employed, log income, as well as being late in paying bills – are all significantly related to mental health measured by CES-D scores in wave I controlling for a range of other individual characteristics and, most notably, a fourth order polynomial in ability as in our baseline specification. A one standard deviation increase in CES-D scores, i.e., a decrease in mental health, is associated with a 5 percentage point decrease in the probability of having a college degree, being 2 percentage point more likely in being employed, 11 percentage points lower income, and with being 4 percentage points more likely to be late in paying bills. Taken together, having a better mental health during adolescence is significantly associated with better 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 by 12 percentage points and income by 25 percentage points. Moreover, in line with our previous results, we find that the effects of ranks are most-pronounced for the lowest ability decile as shown in Appendix Figure A4. 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).[20] One potential explanation for the latter finding is that we can study income over 20 years after high school graduation and at least 15 years after graduation from college. Furthermore, if having a low rank sets people on different trajectories compared to those who have a high rank, this difference might increase over time. While we estimate rank effects on employment status and paying bills on time have the sign we would expect, we do not find evidence of ranks affecting these outcomes.

Finally, we ask to what extend the long-run effects of ranks are mediated by mental health. To study this, we calculate the product of the coefficient of ranks on mental health in wave I

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

Table 7. Economic Long-run Outcomes, Their Association with Mental Health, and Rank Effects

	Association with std. CES-D scores	Rank effect	Share of rank effect mediated by CES-D
$\mathbb{1}$\{College graduate\}	-0.05***	0.12**	19.7%
	(0.00)	(0.05)	
$\mathbb{1}$\{Employed\}	-0.02***	0.07	7.1%
	(0.00)	(0.04)	
log(Personal income)	-0.11***	0.25**	8.9%
	(0.01)	(0.13)	
$\mathbb{1}$\{Late paying bills\}	0.04***	-0.01	81.0%
	(0.00)	(0.06)	

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). This specification controls 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses are clustered at the school level. Column (3) presents the share of this effect that is 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 2, 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.

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 8.9-19.7% of the rank effect for having a college degree and income are mediated by mental health.

Mental health during adolescence is therefore associated with a range of economic outcomes. This points towards the long-lasting consequences of a poor mental health state that affects individuals 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 students' mental health? 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 should worsen a student's mental health, 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 we interpret as a signal about her ability, and allows us to identify the causal effect of ranks on students' mental health.

We provide evidence that mental health is malleable during adolescence and document its persistent over time. Leveraging the quasi-random in ordinal ranks, we find that ranks causally affect students' mental health measured by CES-D scores, an established self-assessment of mental health. Increasing a student's rank from the 25th to the 75th percentile, improves her mental health by 10-11% of a standard deviation and is comparable to rank effects estimated for test scores (Murphy and Weinhardt, forthcoming) and are about half of the effect of losing one's job on mental health (Marcus, 2013). Moreover, these effects are driven by negative rather than positive shocks, more pronounced at the lower end of the ability distribution, and persists over time. Even 23 years after the first wave of the survey, when individuals in our data are approximately 36-42 years old, we find the same pattern as in wave I highlighting the persistence of the initial effect over time.

Taken together, our study provides evidence on the long-lasting effects of features of the school environment and raises several avenues for further research. The persistence of our effects helps to understand why mental health affects educational attainment (Currie and Stabile, 2006; Fletcher, 2010), and therefore may be an additional skill valued on the labor market (Heckman, Stixrud, and Urzua, 2006; Fletcher, 2014). 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.

The asymmetry of rank effects also allows to consider potential consequences for policy-makers and school administrators. If effect of ranks on other outcomes apart from mental health are also asymmetric, this could provide a rationale to think about optimal classroom assignment as in Carrell, Sacerdote, and West (2013). Note, however, that the consequences of different assignments may be ambiguous if peer effects are present in multiple dimensions (Kiessling, Radbruch, and Schaube, 2019).

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, policy-makers have high interest in understanding the causes of these issues to design policies alleviating these effects.

References

Adhvaryu, Achyuta, James Fenske, and Anant Nyshadham. 2019. "Early Life Circumstance and Adult Mental Health." *Journal of Political Economy* 127 (4):1516–1549.

Angrist, Joshua D. and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114 (2):533–575.

Angrist, Joshua D., Victor Lavy, Jetson Leder-Luis, and Adi Shany. 2019. "Maimonides' Rule Redux." *American Economic Review: Insights* 1 (3):309–324.

Baird, Sarah, Jacobus de Hoop, and Berk Özler. 2013. "Income Shocks and Adolescent Mental Health." *Journal of Human Resources* 48 (2):370–403.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. forthcoming. "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Control Trial." *American Economic Review* .

Bartel, Ann and Paul Taubman. 1986. "Some Economic and Demographic Consequences of Mental Illness." *Journal of Labor Economics* 4 (2):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." *Econometrica* 80 (6):2369–2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects after Selection Among High-Dimensional Controls." *The Review of Economic Studies* 81 (2):608–650.

Bietenbeck, Jan. 2019. "The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR." *Journal of the European Economic Association* .

Brown, Gordon D. A., Jonathan Gardner, Andrew J. Oswald, and Jing Qian. 2008. "Does Wage Rank Affect Employees' Well-being?" *Industrial Relations* 47 (3):355–389.

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

Carrell, Scott, Bruce Sacerdote, and James West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3):855–882.

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

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." *Quarterly Journal of Economics* 126 (4):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.” *American Economic Review* 104 (9):2633–2679.

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.” *Journal of Economic Literature* 47 (2):448–474.

Cunha, Flavio and James J. Heckman. 2008. “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation.” *Journal of Human Resources* 43 (4):738–782.

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

Currie, Janet and Mark Stabile. 2006. “Child mental health and human capital accumulation: The case of ADHD.” *Journal of Health Economics* 25 (6):1094 – 1118.

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

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.” *American Economic Review* 101 (5):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.” *American Economic Journal: Microeconomics* 3 (2):114–138.

Eisenberg, Daniel, Ezra Golberstein, Janis L. Whitlock, and Marilyn F. Downs. 2013. “Social Contagion of Mental Health: Evidence from College Roommates.” *Health Economics* 22 (8):965–986.

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

———. 2018. “Rank, Sex, Drugs, and Crime.” *Journal of Human Resources* 52 (2):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.” *Journal of Economic Behavior & Organization* 80 (3):532 – 545.

Ettner, Susan L., Richard G. Frank, and Ronald C. Kessler. 1997. "The Impact of Psychiatric Disorders on Labor Market Outcomes." *Industrial and Labor Relations Review* 51 (1):64–81.

Festinger, Leon. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7 (2):117–140.

Fletcher, Jason M. 2008. "Adolescent depression: diagnosis, treatment, and educational attainment." *Health Economics* 17 (11):1215–1235.

———. 2010. "Adolescent depression and educational attainment: results using sibling fixed effects." *Health Economics* 19 (7):855–871.

———. 2014. "The Effects of Childhood ADHD on Adult Labor Market Outcomes." *Health Economics* 23 (2):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." *BMJ* 337.

Frank, Richard and Paul Gertler. 1991. "An Assessment of Measurement Error Bias for Estimating the Effect of Mental Distress on Income." *Journal of Human Resources* 26 (1):154–164.

Fruehwirth, Jane Cooley, Sriya Iyer, and Anwen Zhang. 2019. "Religion and Depression in Adolescence." *Journal of Political Economy* 127 (3):1178–1209.

Garlick, Robert. 2018. "Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment." *American Economic Journal: Applied Economics* 10 (3):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." *Management Science* 65 (2):494–507.

Gotlib, Ian H., Elena Kranoperova, Dana N. Yue, and Jutta Joorman. 2004. "Attentional Biases for Negative Interpersonal Stimuli in Clinical Depression." *Journal of Abnormal Psychology* 113 (1):121–135.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80 (2):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." *Nature Reviews Neuroscience* 11 (9):651–659.

Haushofer, Johannes and Jeremy Shapiro. 2016. "The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *Quarterly Journal of Economics* 131 (4):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." *Journal of Labor Economics* 24 (3):411–482.

Holtzheimer, Paul E. and Helen S. Mayberg. 2011. "Stuck in a rut: rethinking depression and its treatment." *Trends in Neurosciences* 34 (1):1–9.

Hopkins, Ed and Tatiana Kornienko. 2004. “Running to Keep in the Same Place: Consumer Choice as a Game of Status.” *American Economic Review* 94 (4):1085–1107.

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

———. 2000b. “Peer Effects in the Classroom: Learning from Gender and Race Variation.” 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.” *Quarterly Journal of Economics* 125 (2):515–548.

Kahneman, Daniel and Angus Deaton. 2010. “High income improves evaluation of life but not emotional well-being.” *Proceedings of the National Academy of Sciences* 107 (38):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.” *World Psychiatry* 6 (3):168–176.

Kiessling, Lukas, Jonas Radbruch, and Sebastian Schaube. 2019. “Self-selection of peers and performance.”

Krueger, Alan B. 1999. “Experimental Estimates of Education Production Functions.” *Quarterly Journal of Economics* 114 (2):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.” *Economic Journal* 111 (468):1–28.

Kuppens, Peter, Nicholas B. Allen, and Lisa B. Sheeber. 2010. “Emotional Inertia and Psychological Maladjustment.” *Psychological Science* 21 (7):984–991.

Kuziemko, Ilyana, Ryan W. Buell, Taly Reich, and Michael I. Norton. 2014. “"Last-Place Aversion": Evidence and Redistributive Implications.” *Quarterly Journal of Economics* 129 (1):105–149.

Lavy, Victor and Analia Schlosser. 2011. “Mechanisms and Impacts of Gender Peer Effects at School.” *American Economic Journal: Applied Economics* 3 (2):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.” *Psychology and Aging* 12 (2):277–287.

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

Marcus, Jan. 2013. “The effect of unemployment on the mental health of spouses – Evidence from plant closures in Germany.” *Journal of Health Economics* 32 (3):546 – 558.

Marsh, Herbert W. 1987. “The Big-Fish-Little-Pond Effect on Academic Self-Concept.” *Journal of Educational Psychology* 79 (3):280–295.

Mojtabai, Ramin, Mark Olfson, and Beth Han. 2016. “National Trends in the Prevalence and Treatment of Depression in Adolescents and Young Adults.” *Pediatrics* 138 (6).

Murphy, Jane M. 2011. *Symptom Scales and Diagnostic Schedules in Adult Psychiatry*, chap. 13. John Wiley & Sons, Ltd, 199–219.

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

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

Niederle, Muriel and Lise Vesterlund. 2007. “Do Women Shy Away From Competition? Do Men Compete Too Much?” *Quarterly Journal of Economics* 122 (3):1067–1101.

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

Oster, Emily. 2019. “Unobservable Selection and Coefficient Stability: Theory and Evidence.” *Journal of Business & Economic Statistics* 37 (2):187–204.

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

Persson, Petra and Maya Rossin-Slater. 2018. “Family Ruptures, Stress, and the Mental Health of the Next Generation.” *American Economic Review* 108 (4-5):1214–1252.

Radloff, Lenore S. 1977. “The CES-D scale: A self report depression scale for research in the general population.” *Applied Psychological Measurements* 1 (3):385–401.

———. 1991. “The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults.” *Journal of Youth and Adolescence* 20 (2):149–166.

Roberts, 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.” *Psychological Assessment: A Journal of Consulting and Clinical Psychology* 2 (2):122–128.

Rosenquist, James N., James H. Fowler, and Nicholas A. Christakis. 2011. “Social network determinants of depression.” *Molecular Psychiatry* 16:273–281.

Rothstein, Jesse. 2017. “Measuring the Impacts of Teachers: Comment.” *American Economic Review* 107 (6):1656–1684.

Sacerdote, Bruce. 2011. “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?” chap. 4. Elsevier, 1 ed., 249–277.

Schwandt, Hannes. 2018. “Wealth Shocks and Health Outcomes: Evidence from Stock Market Fluctuations.” *American Economic Journal: Applied Economics* 10 (4):349–377.

Seligman, Martin E. P. 1972. “Learned Helplessness.” *Annual Review of Medicine* 23 (1):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.” *JAMA* 289 (23):3135–3144.

Stinebrickner, Todd and Ralph Stinebrickner. 2012. "Learning about Academic Ability and the College Dropout Decision." *Journal of Labor Economics* 30 (4):707–748.

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

Zafar, Basit. 2011. "How Do College Students Form Expectations?" *Journal of Labor Economics* 29 (2):301–348.

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

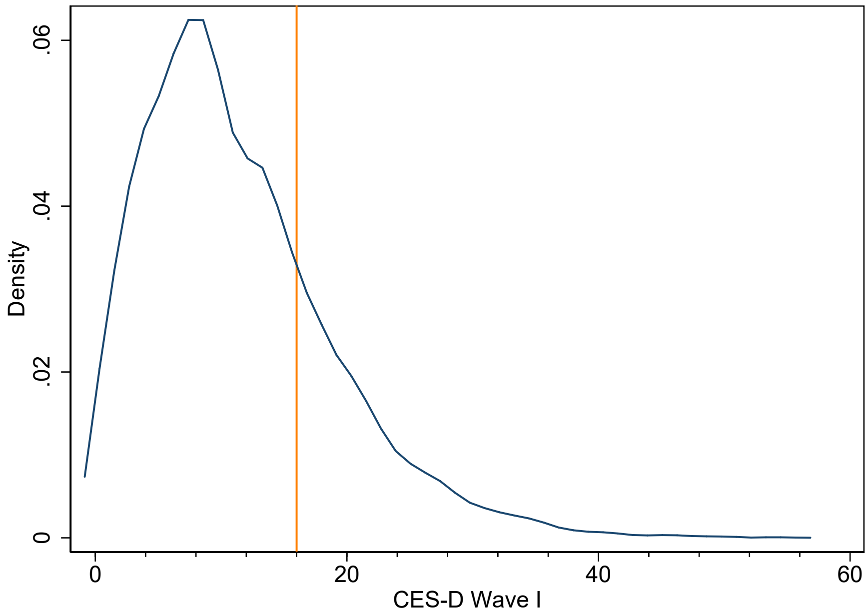
A Additional Tables and Figures

Table A1. 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 felt depressed.	X	X	X	X	X
2. It was hard to get started doing things.	X	X			
3. People were unfriendly to you.	X	X			
4. You didn't feel like eating, your appetite was poor.	X	X			
5. You enjoyed life.	X	X	X	X	
6. You felt fearful.	X	X			
7. You felt hopeful about the future.	X	X			
8. You felt life was not worth living.	X	X			X
9. You felt lonely.	X	X			
10. You felt sad.	X	X	X	X	X
11. You felt that people disliked you.	X	X	X	X	
12. You felt that you could not shake off the blues, even with help from your family and your friends.	X	X	X	X	X
13. You felt that you were too tired to do things.	X	X	X	X	
14. You felt you were just as good as other people.	X	X	X	X	
15. You had trouble keeping your mind on what you were doing.	X	X	X	X	
16. You talked less than usual.	X	X			
17. You thought your life had been a failure.	X	X			
18. You were bothered by things that don't usually bother you.	X	X	X	X	
19. You were happy.	X	X		X	X
Number of items	19	19	9	10	5

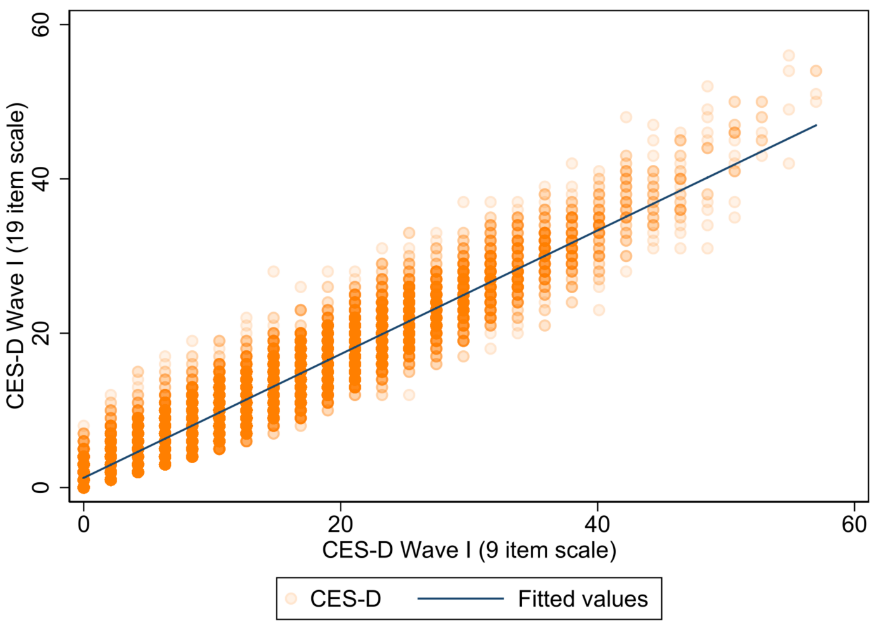
Notes: This table presents all items of the CES-D scale and in which wave they were elicited. Responses are rated on a scale from 0 ("never or rarely") to 3 ("most of the time or all of the time") and aggregated to a final score ranging from 0 to 57, with higher scores indicating a higher propensity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

Figure A1. Distribution of CES-D Scores at Wave I



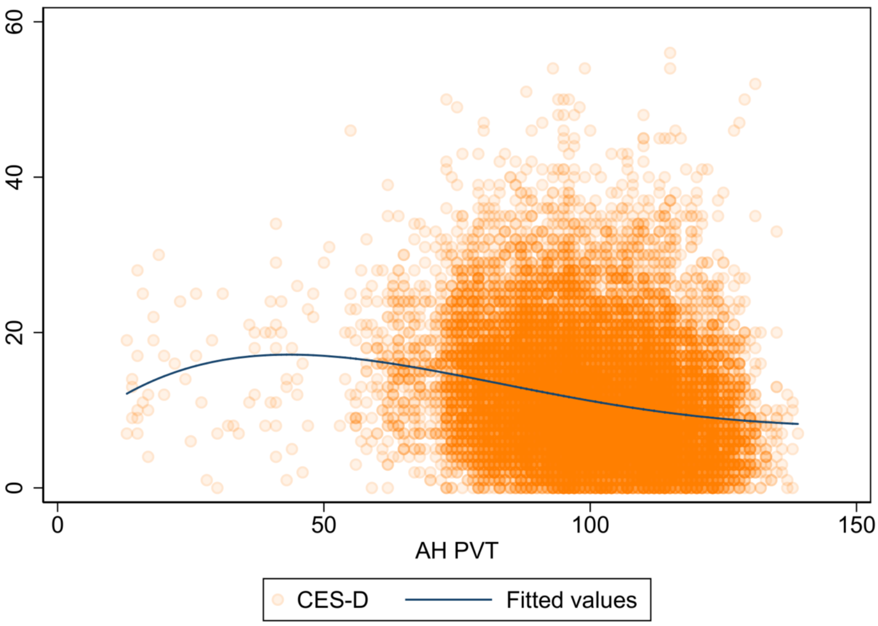
Notes: This figure presents the distribution of the our mental health measure (CES-D score) in wave I. The vertical line indicates a threshold of 16 often used as an indicator of depressions (Radloff, 1977).

Figure A2. Relationship of Long- and Short-Scale of the CES-D Score



Notes: This figure presents the relationship of the CES-D using 20 items as used in wave I and the short version adopted in later waves.

Figure A3. Relationship of CES-D Score and Ability



Notes: This figure presents a scatter plot and nonlinear fit 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 up to a 4th order polynomial.

Table A2. 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	0.09	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
N	18454	1870	1983	2006	2141	1340	2004	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.

Table A3. Ability Rank Salience Check

	Intelligence	College Expectations
Rank	0.28**	0.44**
	(0.13)	(0.18)
N	18429	18385
R^2	0.159	0.163

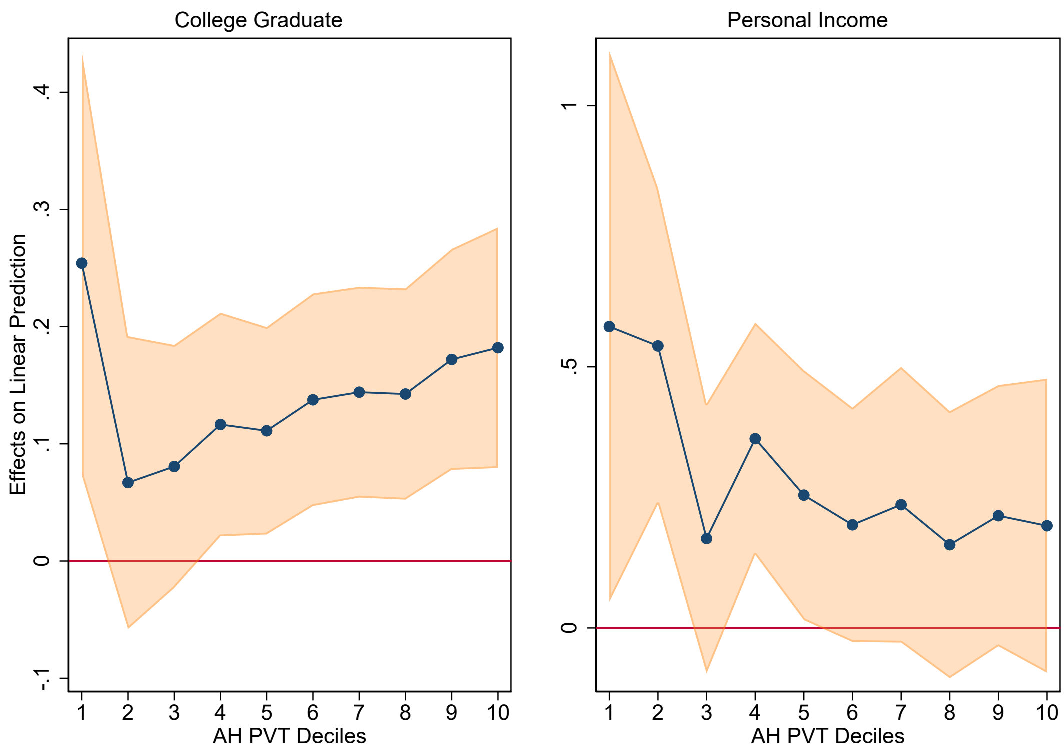
Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are in parentheses are clustered at the school level. Each specification includes all controls as in our preferred baseline specification. Column headers denote the dependent variable. Intelligence is how intelligent the adolescent feels compared 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 adolescent's 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).

Table A4. Robustness to Ability Nonlinearity

	Iterations of AH PVT Polynomial Controls					PDS Lasso	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rank	-1.81***	-2.19***	-1.45**	-1.58**	-1.57**	-2.19***	-1.69***
	(0.63)	(0.68)	(0.73)	(0.75)	(0.75)	(0.67)	(0.63)
AH PVT	-0.04***	-0.07**	0.25**	0.41	-0.41	-0.07**	-0.05***
	(0.01)	(0.03)	(0.10)	(0.28)	(0.56)	(0.03)	(0.01)
$(\text{AH PVT})^2$		0.00	-0.00***	-0.01	0.02	0.00	
		(0.00)	(0.00)	(0.01)	(0.02)	(0.00)	
$(\text{AH PVT})^3$			0.00***	0.00	-0.00		
			(0.00)	(0.00)	(0.00)		
$(\text{AH PVT})^4$				-0.00	0.00		
				(0.00)	(0.00)		
$(\text{AH PVT})^5$					-0.00		
					(0.00)		
Penalized Controls	No	No	No	No	No	Yes	Yes
Number of Penalized Controls Included						8	32
Number of Penalized Controls Selected						2	11
N	18454	18454	18454	18454	18454	18454	18454

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are in parentheses are clustered at the school level. Each specification includes all controls as in our preferred baseline specification. In columns 6-7, 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). In column 6, we include up to an 8-degree polynomial in AH PVT scores and allow the Lasso to select only over these (baseline controls are not penalized, thus always included). Only the degree 1 and 2 AH PVT polynomials are selected. In column 7, we again include up to an 8-degree polynomial in AH PVT and allow selection on both these and our baseline control set (excludes school fixed effects, which are always included). Only AH PVT is selected (degree 1) of the ability polynomials and 10 additional controls from the remaining control set. Under the Lasso standard errors and statistics are only valid for the rank coefficient.

Figure A4. Effects of Ordinal Ranks by Ability Decile on Education-Level and Income



Notes: This figure presents the effect of ordinal ranks by ability decile having obtained a college degree (left panel) and log income (right panel). Shaded area indicates 90% confidence intervals. Standard errors are clustered on at the school level.

B 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):

$$y = -1.8r - 0.6a \tag{B4}$$

in which y denotes our outcome, mental health as assessed by the CES-D scale, a denotes a students 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 her 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 A4 and scaled up as the AHPVT scores have a standard deviation of 15.

Given the data generating process in equation (B4), we assess the consequences of several forms of measurement error using Monte Carlo simulations. 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 B5 and Figure B5, 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 show 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.

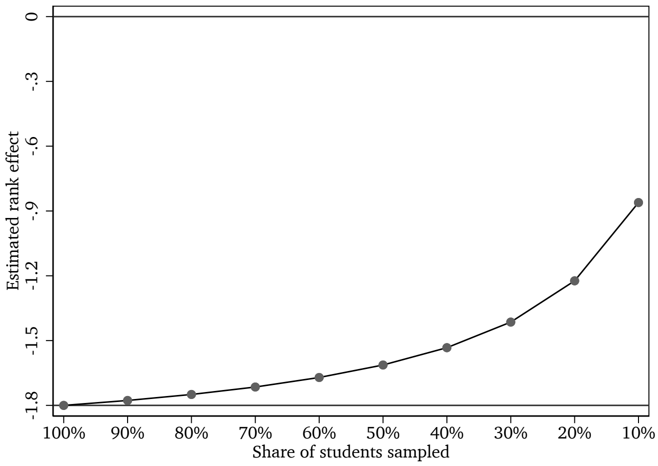
B. Measurement error in ability measure. In our second set of simulations, we introduce measurement error in our ability measure. In particular, we assess how our estimates change once we introduce noise into our measurement, i.e., we measure $\tilde{a} = a + \phi z$ rather than a, where $a, z \sim N(0,1)$ 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 a student in her cohort ($r(\tilde{a})$) as measurement error pertubates the ranks. Table B6 and Figure B6 demonstrate that this also leads to classical attenuation bias yielding an underestimation of the true effect.

Table B5. 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 parenthesis report the ratio of the estimate to the true coefficient from the data generating process.

Figure B5. 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 B5.

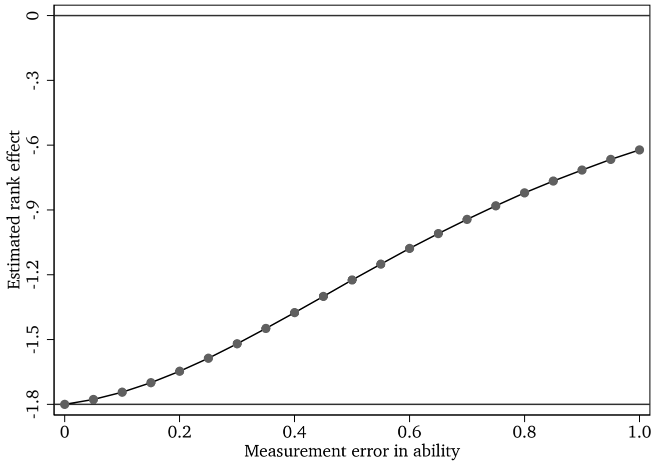
Table B6. Simulations to Assess the Consequences of Measurement Error B

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 parenthesis report the ratio of the estimate to the true coefficient from the data generating process.

C. Omitted variables correlated with ability. Measurement error can also be more complex. For example, there could be an omitted variables z correlated with a that also exert a direct effect on mental health, y. We model this by extending the data generating process in

Figure B6. 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 B6.

equation (B4) as follows:

$$y = -1.8r - 0.6\tilde{a} + \rho z \text{ with } \tilde{a} = a + \phi z \text{ and } z \sim N(0,1). \tag{B5}$$

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 B7 as well as Figures B7a and B7b 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.

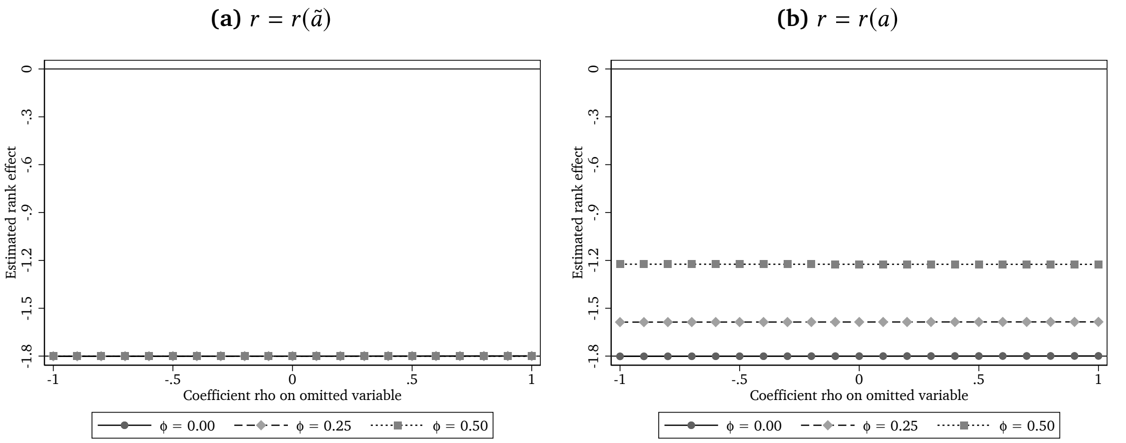
D. Unobserved ability sorting within cohorts. A next set of simulations considers that we only observe students at the cohort level, but have no information on class assignments. Yet, students may be allocated into classes based on their ability. That is, schools may employ tracking into different classrooms. 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 B8 and Figure B8 consider how the estimated rank effect varies with classroom allocations that are either purely random ($\omega = 0$) and gradually moves towards perfect tracking ($\omega = 1$). In addition, we check how the estimates vary once we 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

Table B7. 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 parenthesis report the ratio of the estimate to the true coefficient from the data generating process.

Figure B7. Simulations to Assess Biases from Omitted Variables ($r = r(\tilde{a})$)



Notes: This figure presents results from Monte Carlo simulations with 1000 repetitions of 500 schools as summarized in Table B7.

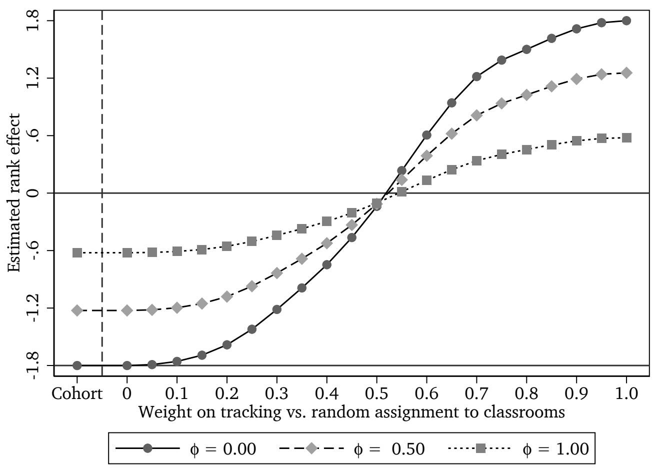
a. Our simulations show that tracking policies within cohorts strongly biases the estimated coefficients. Having random assignment to classroom on average leads to unbiased estimates. Yet, having perfect tracking yields a coefficient of the same size, but the opposite sign. These relationships are dampened, once we allow for measurement error in the ability measure, but the same pattern persists.

Table B8. 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 parenthesis report the ratio of the estimate to the true coefficient from the data generating process.

Figure B8. 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 B8.

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 B9 and Figure B9 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 B9. Simulations to Assess the Consequences of Measurement Error E

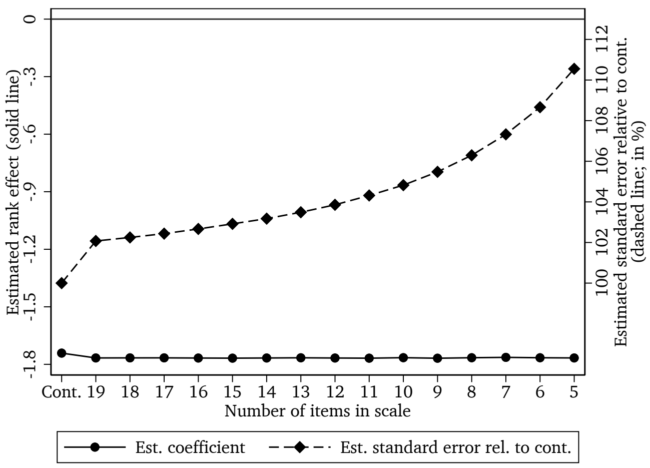
Simulations E: Measurement error in dep. variable due to short scales

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 parenthesis report the ratio of the estimate to the true coefficient from the data generating process.

Figure B9. 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 B9.

F. Interaction of rank and negative shocks in 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 B10 and Figure B10 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.

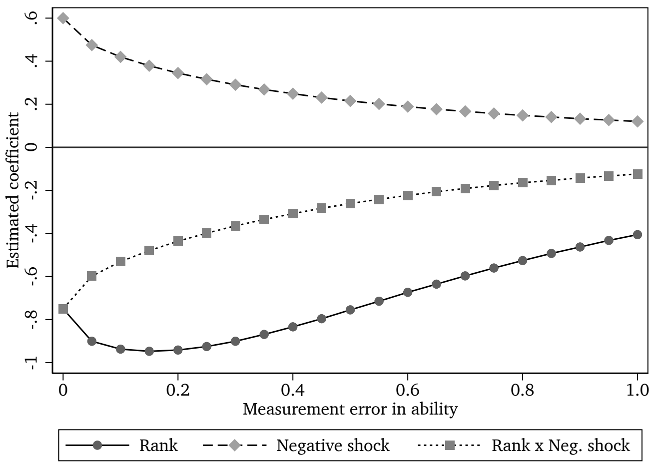
Table B10. Simulations to Assess the Consequences of Measurement Error F

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 parenthesis report the ratio of the estimate to the true coefficient from the data generating process.

Figure B10. 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 B10.