Peer Motivation and Educational Success*

Jan Bietenbeck†

September 11, 2023

Abstract

I provide evidence of social spillovers of personality by showing that peer motivation affects educational success. I first document that academic motivation, which is a key aspect of personality in the context of education, predicts own achievement, high school GPA, and college-test taking among elementary school students. Exploiting random assignment of students to classes, I then show that exposure to motivated classmates causally affects achievement, an effect that operates over and above spillovers of classmates' past achievement and socio-demographic composition. However, peer motivation in elementary school does not affect own motivation and long-term educational success.

*A previous version of this paper was entitled "Own Motivation, Peer Motivation, and Educational Success." Many thanks to Felix Weinhardt, Sule Alan, Fabian Kosse, Kaveh Majlesi, Therese Nilsson, Tove Nordström, Pia Pinger, Ulf Zölitz, and numerous seminar and conference participants for helpful comments and suggestions.

†Lund University and CESifo, DIW Berlin, IZA.

1 Introduction

A growing literature in economics and psychology documents the importance of personality for success in life (Borghans et al., 2008; Almlund et al., 2011; Heckman, Jagelka, and Kautz, 2019). In particular, aspects of personality such as motivation, preferences, and traits have been shown to predict performance in school and in the labor market (e.g. Duckworth et al., 2007; Steinmayr and Spinath, 2009; Golsteyn, Grönqvist, and Lindahl, 2014). Despite this crucial role played by personality in shaping individuals' own life outcomes, only very little research has examined how it affects other people in their social environment. This is surprising given that there is extensive evidence that peers matter for performance in school and in the workplace (e.g. Guryan, Kroft, and Notowidigdo, 2009; Mas and Moretti, 2009; Sacerdote, 2011).

In this paper, I study the spillover effects of one key aspect of personality in the context of education: academic motivation. I use data from the Tennessee Student-Teacher Achievement Ratio experiment (Project STAR), which followed a single cohort of children from the beginning of kindergarten until the end of third grade. Two features make this setting uniquely suited for my purpose. First, the experiment measured students' academic motivation at the end of grades 1, 2, and 3 using a validated psychological scale. Second, some children entered the experiment in second and third grade and were randomly assigned to existing classes within schools. This randomization generated exogenous, observable variation in the predetermined motivation of entrants' classmates, which I can use to estimate causal spillover effects.

Psychologists define motivation as the conscious and unconscious needs and desires of individuals (Roberts, 2006; Roberts and Yoon, 2022). The scale used in Project STAR applies this definition to the context of learning and conceptualizes academic motivation as consisting of two facets. First, *achievement needs* captures the utility that a child derives from learning and the associated social appreciation. Second, *failure avoidance* captures the disutility from low school achievement and the associated embarrassment. The scale measures these facets using a questionnaire given to children and summarizes the answers in a single academic motivation score for each child.

I begin my empirical analysis by showing that this motivation score predicts children's own educational success. I exploit the fact that participants were followed even after the experiment ended in order to study short- and long-term outcomes. The results reveal that, on average, children with a one standard deviation (SD) higher motivation during grades 1 to 3 score about 0.05 SD higher on standardized reading and math tests in elementary and middle school and are four percent more likely to take a college entrance exam around age 18. Motivation further predicts multiple measures of good classroom behavior, as rated by teachers, in fourth and eighth grade.

I next investigate whether children's academic motivation affects the learning outcomes of their classmates. For this analysis, I focus on a sample of students who first entered Project STAR in second or third grade. These students were randomly assigned to existing classes within school upon entry, which allows me to avoid the selection problems that typically complicate the identification of causal peer effects. Moreover, most new classmates of these entrants had participated in the experiment in the previous school year, which lets me observe their predetermined motivation. My regressions exploit the random variation in classmates' average motivation to identify spillover effects on entrants' short- and long-term educational success.

The results show that students who are randomly assigned to a class with more motivated peers initially perform better in school. Specifically, a 1 SD increase in classmates' average motivation raises performance on a standardized reading test at the end of the school year by 0.08 SD (the effect on math scores is 0.04 SD, but this is imprecisely estimated). This spillover effect is not driven by an improvement in own motivation, which I show is unaffected by peer motivation. More generally, peer motivation does not seem to matter beyond contemporaneous achievement, as it does not affect any of the longer-term outcomes measured after the experiment ended and classes were reorganized at the end of third grade.

Peer motivation is likely correlated with other peer characteristics, which could potentially confound these estimates. In additional regressions, I therefore control for classmates' past achievement and their composition in terms of gender, race, and free-lunch eligibility, the main variables that have been used to study peer effects in education in the previous literature (see Sacerdote, 2011; Paloyo, 2020). This reduces the estimated effect of peer motivation on test scores only slightly, which suggests that it is distinct from spillovers due to peer ability and other peer characteristics.

What are the mechanisms behind these results? I argue that the spillovers on contemporaneous achievement are most likely due to an improved learning environment in school, as motivated peers show better classroom behavior and distract their classmates less. As for the lack of longer-term effects, previous research has found that childhood interventions are particularly successful at changing future outcomes if they affect children's personality (e.g. Heckman, Pinto, and Savelyev, 2013). Thus, the absence of longer-term impacts might be due to the fact that peer motivation does not change own motivation. It appears that the contemporaneous effect on reading scores by itself is simply not large enough to generate measurable longer-term impacts. I briefly discuss the implications of these findings in the conclusion.

This paper contributes to a large literature on peer effects in education (for surveys, see Sacerdote, 2011; Paloyo, 2020). One strand of this research focuses on spillovers from peer demographic composition, as measured, for example, by the share of female peers or the share of black peers (e.g. Hoxby, 2000; Hoxby and Weingarth, 2005; Whitmore, 2005; Lavy and Schlosser, 2011; Brenoe and Zölitz, 2019). Another strand exploits random assignment of students to groups in order to examine spillovers from peer ability, which is often proxied by past achievement (e.g. Lavy, Paserman, and Schlosser, 2012; Sojourner, 2013; Booij, Leuven, and Oosterbeek, 2017; Feld and Zölitz, 2017).[1] In Bietenbeck (2020), I add to this latter line of research by studying the impacts of being exposed to a very low-achieving repeater during kindergarten in Project STAR. While the current paper uses the same data, the treatment is very different, and I show below that the effects of peer motivation are robust to controlling for repeater exposure.

A few innovative recent papers extend the research on peer effects by studying spillovers from peer personality. Golsteyn, Non, and Zölitz (2021) exploit data on personality traits and random assignment to classes in a university setting and find that students perform better in the presence of persistent peers, an effect that operates over and above spillovers from peer ability and peer demographic composition. Related work by Shure (2021) and Hancock and Hill (2022) shows that peer conscientiousness positively affects performance in early secondary school and in college, respectively. Ballis (2023) studies a policy-driven increase in the returns to schooling for undocumented youths in the United States. She shows that U.S.-born high school peers of these youths, who do not benefit from the policy directly, perform better in school after its implementation, an effect that could be due to undocumented youths' increased motivation. I contribute to this research by studying spillovers from peer motivation in elementary school, when cognitive and non-cognitive skills are still highly malleable (Kautz et al., 2014). Unlike previous studies, I can estimate effects on long-term outcomes. Moreover, I examine how peer motivation affects own motivation; to the best of my knowledge, this is the first evidence on whether peer personality affects own personality.

This paper also adds to the large literature in economics and psychology on the importance of personality (for surveys, see Borghans et al., 2008; Almlund et al., 2011; Heckman, Jagelka, and Kautz, 2019). This research has shown that motivation (e.g. Wong and Csikszentmihalyi, 1991; Steinmayr and Spinath, 2009), preference parame-

[1] Some related papers examine impacts of peers who likely exhibit disruptive behavior, such as children exposed to domestic violence (Carrell and Hoekstra, 2010; Carrell, Hoekstra, and Kuka, 2018) and boys with female-sounding names (Figlio, 2007).

ters, such as patience (e.g. Golsteyn, Grönqvist, and Lindahl, 2014; Cadena and Keys, 2015), and personality traits, such as conscientiousness (e.g. Poropat, 2009; Gensowski, 2018), grit (e.g. Duckworth et al., 2007), and locus of control (e.g. Piatek and Pinger, 2016), predict educational success. Related recent work documents that school-based interventions can boost favorable aspects of personality in children and thereby improve their school performance (e.g. Alan and Ertac, 2018; Alan, Boneva, and Ertac, 2019; Sorrenti et al., 2020). I complement this research by showing that academic motivation, one important aspect of personality in the context of education, not only predicts children's own educational success but also affects the learning outcomes of their peers.

The remainder of this paper is organized as follows. Section 2 gives an overview of the research on motivation in personality psychology. Section 3 presents details on Project STAR and the data. In Section 4, I document how own motivation in early elementary school relates to short- and longer-term educational success. Section 5 presents estimates of spillovers from motivated peers and tests the robustness of these effects. Section 6 discusses potential mechanisms for and policies based on spillovers from peer motivation. Section 7 concludes.

2 Motivation in personality psychology

The prototypical model of personality in psychology conceives of a core of personality which is made up of four domains: traits, motives, abilities, and narratives (Roberts, 2006; Roberts and Yoon, 2022). Traits capture the relatively stable patterns of thoughts, feelings, and behaviors of an individual and are often represented using the well-known Big Five taxonomy.[2] Motives are defined as what an individual desires, needs, and strives for. Abilities capture things such as intelligence, and narratives are the stories that an individual tells herself in order to make sense of her life. How exactly these four

[2]The Big Five traits are openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. Almlund et al. (2011) give an overview of different taxonomies of personality traits and their relation to the widely-studied concepts of grit and locus of control.

domains relate each other is the subject of an ongoing debate in psychology (Roberts and Yoon, 2022). However, it is widely accepted that together, they shape a person's identity and reputation, which in turn determine her roles in society.

This paper studies the importance of academic motivation, which falls under the motives domain. Unlike the literature on personality traits, psychological research on motivation has not converged on a common theoretical framework, system of measurement, or terminology (Murphy and Alexander, 2000; Roberts et al., 2006; Roberts and Yoon, 2022). Despite this heterogeneity, empirical studies have consistently found that motivation is predictive of success in life: for example, Steinmayr and Spinath (2009) document that motivation predicts school performance over and above intelligence, and Dunifon and Duncan (1998) find that having an orientation toward challenge predicts future earnings. In related work in economics, Segal (2012) shows that intrinsic motivation in adolescence and early adulthood, as measured by performance on a low-stakes coding speed test, predicts future earnings over and above cognitive skills.

The apparent importance of motivation for success in life has led psychologists to study potential ways to boost motivation among students. Results show that interventions that directly aim at increasing motivation, for example by helping students set learning goals or by instructing teachers to relate lesson content to students' experiences, can improve motivation and achievement (see Hulleman and Barron, 2015; Lazowski and Hulleman, 2016). In related research in economics, Heckman, Pinto, and Savelyev (2013) show that the Perry Preschool program boosted children's academic motivation, an effect that partly explains its positive impact on their longer-term educational success. In contrast, previous analyses of Project STAR did not find any evidence that class size affects motivation (Word et al., 1990; Schanzenbach, 2006).

3 Project STAR: background and data

3.1 Background on Project STAR

Project STAR was a randomized controlled trial designed to investigate the effect of class size on student achievement. The original experiment followed a single cohort of children at 79 schools in Tennessee from kindergarten through third grade. It started at the beginning of the 1985-86 school year, when 6,325 kindergarten students were randomly assigned to small classes (target size 13-17 students) or regular-sized classes (target size 22-25 students) within their school.[3] Because kindergarten was not mandatory at that time and due to normal residential mobility, 5,276 additional students joined this study cohort at participating schools during grades 1-3. These students were also randomized to classes within school upon entry, implying that the randomization pool for all participants was school-by-entry-grade. After the initial randomization, all students were supposed to stay in their assigned class type (small versus regular-sized) until the end of third grade, at which point the experiment ended.

Teachers were also randomly assigned to classes within school at the start of each grade. As is common in the United States, Project STAR teachers worked only in one specific grade (that is, they were not "looped"). As a consequence, students met a new, randomly assigned teacher in each and every grade.

As with any field experiment, the actual implementation of Project STAR deviated somewhat from the original plan. Thus, as children advanced from kindergarten to third grade, some students managed to move between small and regular-sized classes (for details, see Krueger, 1999). To account for this likely non-random sorting, I always define peer composition based on the initial random assignment when I estimate spillovers from motivated classmates below. Another deviation from the original study design was that a substantial number of students left the experiment either because they

[3]There was also a third type of class: regular-sized class with a full-time teacher's aide. Previous studies using data from Project STAR have not found any differences in treatment effects between regular-sized classes with and without a full-time teacher's aide. In the empirical analysis, I follow the convention in the literature and group these two types of classes together.

moved to other schools or because they were retained in grade. In Section 5 below, I provide evidence that this attrition is not driving my results.[4]

3.2 Data and variable definitions

An important feature of Project STAR is that researchers collected detailed data on participants both during the experiment and long after it ended. Most of these data are included in the Project STAR public use file, which I use in my analysis and which allows me to follow students from the time they entered the experiment until the end of high school. In what follows, I give an overview of the main variables that I draw from these data. Additional details can be found in Online Appendix A.

Academic motivation In the spring of each year from kindergarten through third grade, students' academic motivation was assessed using the Self-Concept and Motivation Inventory (SCAMIN; Milchus, Farrah, and Reitz, 1968). This psychological scale conceptualizes academic motivation as consisting of two facets. First, *achievement needs* is defined as the positive regard with which a student perceives the intrinsic and extrinsic rewards of learning and performing in school. In economic terms, this captures the utility that a child derives from learning and the associated social appreciation.[5] Second, *failure avoidance* is defined as the awareness and concern toward shunning the embarrassment and sanctions which are associated with failure in school. In economic terms, this captures the disutility from low school achievement and the associated embarrassment.

As is common in personality psychology, the SCAMIN measures academic motivation using a self-assessment questionnaire. The instrument is group-administered, which implies that all children complete the questionnaire in the classroom at the same time

[4]For additional details on the design and implementation of Project STAR, see Word et al. (1990), Krueger (1999), and Finn et al. (2007).

[5]I define this economic counterpart based on my analysis of the related SCAMIN questions. Unfortunately, due to copyright restrictions, not all SCAMIN question can be reproduced here, although two examples are given further below in the main text.

following instructions by their teacher. Specifically, students are first given an answer sheet that contains a number of faces ranging from sad to happy for each question. The teacher then reads out a series of questions starting with "What face would you wear..." and asks students to mark the appropriate face as a response. For example, students are asked "What face would you wear if you could read like a grown-up?" and "What face would you wear if you could make your teacher happy with your arithmetic?" The questions ask about both subject-specific achievement in reading and math and school achievement in general. Half of the questions measure achievement needs, and half measure failure avoidance.

The outcome of the assessment are individual-level academic motivation scores, which summarize each student's answers. In Project STAR, these scores were calculated centrally by the experimental staff following the SCAMIN scoring guidelines. These motivation scores, but not the answers to individual questions, are included in the Project STAR public use file and form the basis for the empirical analysis below.

Besides academic motivation, the SCAMIN also measures students' academic self-concept using a separate set of questions. Psychologists define self-concept as a person's perception of herself, which is formed through experience with her environment (Shavelson, Hubner, and Stanton, 1976). In the prototypical model of personality, self-concept forms part of a person's identity, which is shaped by the four core personality domains but which may itself also influence these domains via feedback processes (Roberts, 2006). In Section 5 below, I study how peer motivation affects academic self-concept.

As is usual for standardized tests for children, the SCAMIN has different test forms that are aimed at different grade levels: preschool/kindergarten, early elementary school, late elementary school, and secondary school. In Project STAR, the preschool/kindergarten form was administered at the end of kindergarten and the early elementary form was administered at the end of grades 1-3. These forms differ in the questions that are asked and the number of faces that are shown on the answer sheet, such that motivation scores are not directly comparable between them.[6]

Tests in personality psychology are often judged on various dimensions of quality, such as reliability and the ability to predict contemporaneous and future outcomes. As discussed in detail in Online Appendix A, the existing evidence points to a high quality of the SCAMIN early elementary form: for example, its test-retest reliability is similar to that found for tests measuring personality traits in children, and my results below show that its motivation score predicts a wide range of contemporaneous and future outcomes. Unfortunately, however, the preschool/kindergarten form does not meet this same high quality standard. In particular, there is some doubt about whether it really only captures motivation, and I found in separate analyses that its motivation score does not predict contemporaneous or future outcomes, including future motivation as captured by the early elementary form (see Online Appendix A). Given these serious problems, I decided not to use the kindergarten motivation scores and to focus only on motivation in grades 1-3 as measured by the SCAMIN early elementary form.

Achievement in reading and math At the end of each grade from kindergarten through third grade, participants in Project STAR wrote the grade-appropriate version of the Stanford Achievement Test. Moreover, in the spring of grades 5-8, all students who were enrolled in public schools in Tennessee wrote the Comprehensive Test of Basic Skills as part of a statewide testing program. Both tests are standardized assessments covering various subjects, and I use the reading and math scores included in the Project STAR public use file as my main measures of student achievement.

Classroom behavior When STAR participants were in fourth grade, their teachers rated a subset of them on their classroom behavior. Teacher ratings for 28 behaviors were recorded on a scale from 1-5 and then consolidated into four indices. The effort index measures behaviors such as showing persistence when confronted with difficult

[6] In particular, the preschool/kindergarten form has three different faces, whereas the early elementary form has five different faces. Both forms consist of twelve questions measuring motivation and another twelve questions measuring self-concept.

problems. The initiative index captures things such as actively participating in classroom discussions. The discipline index measures behaviors such as being quiet versus interfering with classmates' work. The value index captures to what extent a student appreciates the school learning environment. All indices are coded such that higher values reflect better behavior. In eighth grade, math and English teachers rated a different subset of STAR participants using a similar but shorter questionnaire, and the ratings were consolidated into the same four indices. In the analysis below, I measure classroom behavior using the total of eight fourth- and eighth-grade indices.

Educational attainment Most participants in Project STAR graduated from high school in 1998, and researchers collected information on the high school grade point average (GPA) and graduation status for participants attending selected high schools in 1999 and 2000. Besides this information, the public use file contains an indicator for whether a student had taken an ACT or SAT college-entrance test by 1998. This indicator is based on the administrative records of the two companies offering these tests and is the outcome of a data collection effort by Krueger and Whitmore (2001). It is available for the full sample of STAR participants and is a measure of college intent.

Student characteristics The data contain information on the following socio-demographic characteristics of students: age, gender, race, and an indicator for whether the student was ever eligible for free or reduced-price lunch during the experiment. Based on students' exact date of birth and the school entry cutoff date in Tennessee, I additionally construct an old-for-grade indicator, which identifies students who either entered school late or repeated a grade. In my previous research on Project STAR, I found that old-for-grade students perform substantially worse in school compared to their on-grade peers (Bietenbeck, 2020). I also construct a measure of predicted achievement, which combines the socio-demographic characteristics such that they optimally predict students' reading and math scores; specifically, I predict achievement from a regression of the averaged reading and math score at the end of students' first year in Project STAR on the five socio-demographic characteristics (including the old-for-grade indicator) and school-by-entry-grade fixed effects.

Class size Most of my regressions control for the original Project STAR treatment: assignment to a small class. I measure treatment assignment upon entry into the experiment in order to avoid issues of non-compliance in later grades (see Section 3.1).

3.3 Missing data

Like most other longitudinal data, the Project STAR data contain missing values in some variables, which could affect the results of my analysis. I distinguish between three cases of missing data. First, there are missing values in motivation scores. One main reason for this is a data matching problem: after teachers handed over the completed SCAMIN answer sheets to the experimental staff, many respondents could no longer be uniquely identified due to the lack of a consistently coded student identifier. If an answer sheet could not be uniquely matched, it was ignored, leading to missing motivation scores in the data (see Word et al., 1990, p.210). Another important reason for missing values in the motivation variable is that many students only entered Project STAR in one of the later grades, and thus did not participate in the SCAMIN assessment in the earlier grades. The missing data imply that I do not usually observe the motivation of all students in a class, with the consequence that peer motivation is measured with error. In Section 5, I discuss this problem in detail and also provide solutions.

Second, there is missing information on some outcome variables for some students. The reasons are manifold and include purposeful selective data collection in order to save money and time (like with the classroom behavior ratings and the high school outcomes), accidental selective data collection (for example, due to students being absent on the day of a test), and the loss of records (in particular, the lack of a unique student identifier meant that some test scores could not be matched to students, see Word et al., 1990, p.209). A consequence of these missing outcome data for the empirical analysis is that sample sizes differ between regressions with different dependent variables. Importantly, I show in Section 5 that peer motivation does not predict whether my main outcomes are observed for a given student, and that my results hold when the sample is restricted to students observed with all main outcomes.

Third, there are a few missing values in student socio-demographic variables, which I mostly use as controls in my regressions. In order not to reduce sample size unnecessarily, in all regressions in this paper I impute missing values in controls at the sample mean and include separate dummies for missing values on each control variable. Results are virtually identical if I instead exclude students with missing information on socio-demographic characteristics from the sample.

4 Academic motivation: correlates and predictive validity

4.1 Sample selection

In this section, I examine how academic motivation correlates with students' socio-demographic characteristics and measures of their own contemporaneous and future educational success. For this descriptive analysis, I focus on the 9,072 Project STAR participants for whom I observe a motivation score in at least one of grades 1, 2, and 3. I construct the average motivation of each student during these grades in three steps: (1) I standardize the motivation scores for each grade to have mean 0 and SD 1, (2) I average the available scores for each student across grades, and (3) I standardize the resulting average scores to have mean 0 and SD 1. I prefer this measure of motivation because averaging across grades reduces measurement error and increases statistical precision by maximizing sample size. Nevertheless, I also provide results for grade-specific measures of motivation, which are qualitatively similar.

4.2 Correlates of academic motivation

Table 1 shows estimates of regressions of average motivation in grades 1-3 on student socio-demographic characteristics, a small-class dummy, and school-by-entry-grade fixed effects. Column 1 shows that male students are substantially less motivated on average, with 0.29 SD lower motivation. In contrast, columns 2 and 3 show that there are no significant differences in motivation by race and free-lunch eligibility. Column 4 reveals that students who are old for grade are much less academically motivated, with 0.21 SD lower motivation, and that conditional on old-for-grade status, older students are slightly more motivated. Column 5 shows results from a regression in which all five student characteristics enter at the same time, which confirm the described patterns.

The final row in Table 1 shows the coefficients on the small-class dummy. Because assignment to small classes was random conditional on school-by-entry-grade fixed effects, these estimates capture the causal effect of class size on motivation. The results show that unlike targeted interventions that directly aim to improve students' motivation (see Hulleman and Barron, 2015; Lazowski and Hulleman, 2016), a non-targeted reduction in class size does not appear to boost students' motivation.

4.3 Predictive validity of academic motivation

I now examine the predictive validity of academic motivation. I estimate regressions of the following form:

$$y_{is} = \alpha + \beta \text{MOTIV}_i^{G1-G3} + X_i\gamma + \lambda_s + \varepsilon_{is}, \tag{1}$$

where i denotes students and s denotes school-by-entry-grade cells, that is, the Project STAR randomization blocks. y_{is} is a measure of classroom behavior or educational success. MOTIV_i^{G1-G3} is student i's average academic motivation across grades 1-3. X_i is a vector of socio-demographic controls. λ_i is a vector of school-by-entry-grade dummies, which account for differences between students entering the various schools participating in Project STAR in different grades. Finally, ε_{is} is the error term. In all regressions, I cluster standard errors at the level of school-by-entry-grade.

Table 2 reports the results. Panel A shows that motivation predicts good classroom behavior, as rated by teachers, in fourth and eighth grade. For example, a 1 SD higher motivation in grades 1-3 is associated with 0.10 SD higher effort and 0.09 SD higher discipline in fourth grade. More motivated students also show better initiative and appreciate the school learning environment more. The associations are also positive but slightly weaker for classroom behavior in eighth grade, which could reflect either fade-out or the fact that the questions on which teachers rated students were different in that grade.

Panel B shows that in line with previous research from psychology (e.g. Wong and Csikszentmihalyi, 1991; Steinmayr and Spinath, 2009), motivation predicts short- and long-term educational success. For example, a 1 SD higher motivation is associated with 0.05 SD higher standardized reading and math scores in both elementary school (grades 1-3) and middle school (grades 5-8). Motivation in early elementary school also predicts high school success and college intent: students with a 1 SD higher motivation have 0.3 points (0.04 SD) higher GPAs and are 1.5 percentage points more likely to take an ACT or SAT test around age 18, an increase that corresponds to about four percent of the sample mean.

How large are these associations? One way to gauge the size of the correlations between classroom behaviors and motivation is by comparing them to the gender gap, which has been widely documented in previous research (e.g. Bertrand and Pan, 2013). Across the eight measures of classroom behavior studied in Panel A of Table 2, the coefficients on motivation correspond to 22 percent of the gap between male and female students on average. Another salient reference point is the gap in educational outcomes between low- and high-socioeconomic-status students, as proxied by free-lunch eligibility. Panel B of Table 2 reveals that the estimated coefficients on motivation correspond to slightly more than 10 percent of the achievement gap in reading and math between these two groups. Taken together, the results in Table 2 show that the motivation score captures a dimension of personality that is reflected in students' actual behaviors and predictive of their educational success.[7]

5 Peer motivation and educational success

5.1 Sample selection and summary statistics

I now study how peer motivation affects educational success. Specifically, I estimate causal spillover effects on students who first entered Project STAR in second or third grade. The new classmates of these entrants had participated in the experiment and written the SCAMIN test in the previous (first or second) grade, which allows me to observe their academic motivation. As students in Project STAR were randomly assigned to classes within school upon entry, this means that there is random and observable variation in the motivation of second- and third-grade entrants' classmates, which I can use to estimate causal spillover effects.

A total of 2,962 students entered Project STAR in second or third grade. I construct peer motivation as the average motivation of these entrants' classmates as measured at the end the previous school year. This ensures that peer motivation is predetermined relative to the assignment of entrants to classes. For reasons described in Section 3, some classmates are not observed with a motivation score. In my main analysis, I ignore these missing values and compute peer motivation as the average of the available scores. Moreover, I drop 94 students from the sample for whom there is no information

[7] These findings are confirmed in regressions in which the main independent variable is grade-specific motivation, rather than average motivation across grades 1-3. Online Appendix Figure B.1 reports the corresponding estimates, which are based on a consistent sample of students observed with motivation in all three grades. While motivation in each grade is positively associated with almost all of the outcomes, the point estimates tend to be smaller than the ones for average (across grades) motivation. This supports the intuition that averaging reduces measurement error in the motivation variable.

on any of their classmates' motivation. In Subsection 5.6 below, I describe the problem of missing peer motivation scores in more detail and I show that as a consequence, my estimates are slightly biased toward zero.

For the remaining 2,868 students in the estimation sample, I construct a range of other peer variables, which I use as controls in some regressions. Specifically, I compute averages of classmates' socio-demographic characteristics and their reading and math achievement in the previous grade. To facilitate interpretation and comparison of results, I standardize both peer achievement and peer motivation to have mean 0 and SD 1.[8] I also construct a dummy for having a classmate who repeated kindergarten in the first year of Project STAR; this is the treatment I consider in Bietenbeck (2020), which captures exposure to a very low-achieving peer.

In line with the bulk of the previous research on peer effects, the main specifications focus on spillover effects on contemporaneous outcomes. Specifically, I estimate how exposure to motivated peers affects entrants' reading and math achievement at the end of their first year in Project STAR. In additional analyses, I also examine impacts on entrants' own academic motivation and self-concept at the end of their first year in Project STAR and their long-term educational success. For ease of interpretation, I standardize all achievement outcomes to have mean 0 and SD 1.

Table 3 shows summary statistics for the peer motivation sample. Due to the fact that Project STAR oversampled schools in poor neighborhoods, students are disproportionately likely to be black and eligible for free lunch. Fully 47 percent of students are old for grade, which implies that they either entered school late or repeated a grade. In terms of outcomes, only 73 percent of students graduated from high school and only 26 percent took an ACT or SAT test around the age of 18. Taken together, these statistics show that the sample mostly includes disadvantaged and low-achieving students.

[8] The standard deviation of average peer motivation before standardizing is 0.357. This implies that by multiplying regression coefficients by $1/0.357 = 2.8$, one obtains the estimated effect of raising average peer motivation by 1 SD on the individual-level motivation scale.

5.2 Regression specification

I estimate regressions of the following form:

$$y_{ics} = \theta\overline{\text{MOTIV}}_c^{G-1} + \phi\text{SMALL}_c + X_i\eta + \overline{Z}_c\rho + \omega_s + \mu_{ics}, \tag{2}$$

where i denotes students, c denotes classes, and s denotes school-by-entry grade cells. y_{ics} is the outcome of interest. $\overline{\text{MOTIV}}_c^{G-1}$ is the average motivation of students in class c who participated in Project STAR in the previous grade $(G-1)$; as described above, this average is computed based only on the non-missing motivation scores. SMALL_c is a dummy for assignment to a small class, the original treatment of interest in Project STAR. X_i is a vector of student socio-demographic characteristics and $\overline{Z}_c$ is a vector of predetermined peer characteristics shown in Table 3. Finally, ω_s is a vector of school-by-entry-grade dummies that accounts for fixed differences between randomization pools and μ_{ics} is the error term. For all regressions, I compute standard errors that allow for clustering at the level of school-by-entry-grade.

Equation 2 corresponds to a linear-in-means model, which is the most widely estimated model of peer effects (Sacerdote, 2011). The main coefficient of interest, θ, captures the causal impact of exposure to motivated peers under the assumption that variation in peer motivation is random within school-by-entry-grade cells, an assumption that I support with empirical evidence below. Since peer motivation is correlated with other peer characteristics, an obvious question is whether θ captures spillovers from motivation or from such other characteristics. I address this question by controlling for peers' previous achievement and socio-demographic characteristics, the main variables used to study peer effects in the previous literature (see Sacerdote, 2011; Paloyo, 2020). If the estimates are robust to the inclusion of these controls, this suggests that θ indeed captures spillovers from peer motivation, rather than from correlated observed and unobserved factors (Altonji, Elder, and Taber, 2005; Oster, 2019).[9]

5.3 Evidence on random assignment

Previous studies using data from Project STAR provide detailed evidence that students were randomly assigned to classes within school upon entry, see especially Chetty et al. (2011) and Sojourner (2013). Here, I complement this evidence by showing that peer motivation is unrelated to predetermined characteristics of students entering the experiment in second or third grade.

Table 4 reports results from regressions like in Equation 2 in which the dependent variables are students' predetermined socio-demographic characteristics (columns 1-5) and predicted achievement (column 6). Panel A shows estimates from separate regressions for peer motivation and, to further buttress the results, peers' past achievement in reading and math. Panel B shows estimates from specifications in which these three peer variables enter simultaneously instead. Across all regressions, most of the coefficients on the peer variables are close to zero and not statistically significant at conventional levels. In the regressions in Panel B, the coefficients are also jointly insignificant. This strongly suggests that second- and third-grade entrants in Project STAR were indeed randomized to classes within school upon entry.

In Online Appendix B, I present two further pieces of evidence in favor of random assignment. First, following Chetty et al. (2011), Online Appendix Table B.1 shows that class dummies do not jointly predict most predetermined characteristics of entrants, as should be the case if they were randomized into classes. Second, following Feld and Zölitz (2017), I ran separate regressions of these characteristics on class dummies for each school-by-entry-grade cell. After each regression, I conducted an F test for the joint

[9] Besides peer ability and peer socio-demographic characteristics, peer motivation might also be correlated with other aspects of peer personality. While not definitive, the results from the analysis based on the method developed by Oster (2019) below suggest that such correlated aspects of personality are not driving my results. At the very least, my estimates should be interpreted as capturing the effects of peer motivation and other correlated aspects of peer personality, which are distinct from spillovers from peer achievement and other observable peer characteristics.

significance of the class dummies and collected the p-value. Under random assignment, these p-values should be distributed roughly uniformly, and Online Appendix Figure B.2 shows that this is indeed the case. Moreover, the shares of p-values below certain confidence levels should be close to this level (for example, about five percent of p-values should be below 0.05), and Online Appendix Table B.2 confirms this. This evidence provides strong additional support for the assumption that second- and third-grade entrants were randomly assigned to classes within school in Project STAR.

5.4 Main results: effects on contemporaneous achievement

Table 5 reports my main estimates of the effect of exposure to motivated peers on reading and math achievement at the end of entrants' first year in Project STAR. Column 1 shows that having classmates with a 1 SD higher average motivation raises own reading achievement by 0.081 SD. Column 4 shows an effect on math achievement that is also positive but smaller at 0.036 SD and not statistically significant at conventional levels. Figure 1 visualizes these estimates and reveals that the effects are roughly linear in average peer motivation.

Columns 2 and 5 of Table 5 add three controls for peer ability to these regressions: classmates' average reading and math achievement in the previous school year and an indicator for whether the class includes a very low-achieving kindergarten repeater. If spillovers from motivated peers were mainly due to correlated peer ability, we would expect this to lead to a substantial reduction in the size of the coefficient on peer motivation.[10] However, the estimates are largely unchanged, suggesting that this is not the case. Columns 3 and 6 show that the results are also robust to controlling for classmates' socio-demographic characteristics. In this most demanding specification, a

[10] It might be surprising that the estimates do not change more when controlling for the presence of kindergarten repeaters, which I show to have very low achievement and non-cognitive skills in Bietenbeck (2020). However, I found in additional analyses that the correlation between classmates' average motivation and having a kindergarten repeater in class is low, with a correlation coefficient of –0.07. This explains why results hardly change with the inclusion of this control.

1 SD increase in peer motivation is estimated to raise own reading achievement by 0.071 SD and own math achievement by 0.027 SD, though the latter effect is not statistically significant at conventional levels.

The fact that the estimates in Table 5 change only little when controls for other peer variables are added to the regressions suggests that they capture a true personality spillover from classmates' motivation, rather than a spillover from correlated unobserved factors. I provide more formal evidence in support of this argument in Online Appendix Table B.3, where I use the method developed by Altonji, Elder, and Taber (2005) and refined by Oster (2019) to assess how large omitted variable bias would have to be in order to drive the estimated effect of peer motivation on reading achievement to zero. Under standard assumptions for this method, selection on unobservables would have to be more than 1.5 times as large as selection on observed peer achievement and socio-demographic characteristics to explain away the effect. This finding supports the interpretation of my estimates as capturing spillovers from peer motivation.

In additional analyses, I explore whether the effect of peer motivation differs by entrants' socio-demographic characteristics and two widely-studied educational inputs, class size and teacher experience. Figure 2 presents results from regressions in which the sample is split into corresponding subgroups. The effect appears to be larger for boys, black students, and on-grade students, although none of these differences is statistically significant at conventional levels.[11] It also appears to be larger (though not significantly so) in regular-sized classes as compared to small classes, and in classes with less experienced teachers. Overall, while these analyses point toward potential heterogeneities in the effect of peer motivation, the relatively small sample size means that I lack statistical power to draw definitive conclusions.

[11] I also tested whether there are distinct spillover effects from male versus female classmates' motivation. Online Appendix Table B.4 shows that this is not the case in general. However, the estimates do suggest that male students benefit disproportionately from having motivated male peers, and female students benefit disproportionately from having motivated female peers, although these differences are not statistically significant at conventional levels.

5.5 Further results: effects on own motivation and long-term educational success

In additional analyses, I examine the effects of peer motivation on further outcomes. First, an intriguing possibility is that peer personality affects own personality. In columns 1 and 2 of Table 6, I explore such spillovers by estimating the effect of peer motivation on entrants' own motivation and self-concept at the end of their first year in Project STAR. To the best of my knowledge, these are the very first estimates of spillovers from peer personality on own personality. As can be seen in the table, the estimated effect of peer motivation in both regressions is almost exactly zero, showing that peer motivation does not affect own motivation or self-concept.[12]

Second, given that peer motivation raises contemporaneous achievement, an obvious question is whether it also affects students' long-term educational success. I address this question by estimating effects on middle school test scores, high school outcomes, and college-test taking. When interpreting these estimates, it is important to realize that they capture the impacts of a relatively short exposure to more motivated peers during early elementary school. Specifically, when Project STAR ended after third grade, students were redistributed to ordinary classes. While I do not observe class composition beyond third grade, this re-shuffling likely meant that peer motivation in second or third grade was at most weakly related to peer motivation in later grades. Therefore, my estimates reflect the effects of differential exposure to more motivated peers for only one or two years during early elementary school.[13]

Columns 3 to 7 of Table 6 show the results from this long-term analysis. Across the five regressions, there is no indication that the short-term positive spillover from

[12] I also examined whether peer motivation affects classroom behavior. Unfortunately, teacher ratings of classroom behavior of second- and third-grade entrants were recorded only in eighth grade and only for a small subsample of 620 students. The regressions showed no statistically significant effects of peer motivation on classroom behavior in eighth grade in this subsample, but the effects were very imprecisely estimated due to the small sample size.

[13] In fact, even students who entered Project STAR in second grade did not experience the same peer environment for two years since some students left the experiment and others entered between grades 2 and 3. This further corroborates the interpretation of my estimates as reflecting short-term exposure to more motivated peers.

motivated peers on achievement translates into longer-term educational success. If anything, the estimates point toward a negative effect of peer motivation on later outcomes, although most coefficients are imprecisely estimated and I cannot exclude small positive effects. I discuss potential explanations for this apparent discrepancy between short- and long-term effects of peer motivation in Section 6 below.

5.6 Robustness

I now address various possible concerns about the validity of my findings, including bias due to missing data, selective attrition, and multiple-hypothesis testing.

Missing data on peer motivation As described in Section 3.3, motivation is not observed for all students, which implies that peer motivation is measured with error. In what follows, I characterize this measurement problem and show that it leads to a small bias toward zero in my estimates. I start by plotting the distribution of the share of classmates observed with motivation scores. Figure 3 shows that, on average, 67 percent of all classmates have motivation scores. Interestingly, the missing information for the other classmates is largely driven by new entrants: if several students enter a given class in the same grade, the co-entrants of any given entrant mechanically do not have motivation scores because they did not participate in Project STAR in the previous year. Figure 3 reveals that when such co-entrants are ignored, the share observed with motivation scores is much higher at 86 percent on average.

How do these missing values influence my results? As missing peer information is a common problem, it has been analyzed in detail in the previous literature. In the context of Project STAR, Sojourner (2013) shows that under random assignment to classes, missing information in the variable of interest leads estimates to be attenuated to zero if the peer average is constructed only from the available information.[14] The degree of attenuation is roughly proportional to the fraction of individuals with missing

[14] In an independent analysis on different data, Ammermueller and Pischke (2009) also show that missing data on peer characteristics bias estimates toward zero under random assignment.

data, and the bias is toward zero independently of whether the non-observed classmates have, on average, higher or lower values in the variable of interest than the observed classmates. An implication of this finding for my analysis is that I underestimate the effect of peer motivation above.

In order to assess the extent of this bias, I apply the correction developed by Sojourner (2013) in his paper. In particular, he shows that multiplying the peer variable with the fraction of classmates observed at the individual level removes the attenuation bias. Figure 4 presents results from regressions which apply this correction and reveals that this leads to small increases, in absolute value, of my estimates.[15] For example, the estimated effect on entry-grade reading achievement rises from 0.071 SD to 0.096 SD. Still, none of the estimates for the other outcomes reaches statistical significance at the 5 percent level. I conclude that the amount of bias in my main estimates due to missing data on peer motivation is relatively small.[16]

Measurement error in motivation The SCAMIN arguably measures academic motivation with substantial error. Although this measurement error is reduced when motivation is averaged across classmates, it could further attenuate my estimates.[17] In an attempt to correct for such measurement error, I ran regressions in which I instrumented peer motivation with peer self-concept. As mentioned in Section 3, psychologists view self-concept as being shaped by motivation, with potential feedback processes running in the opposite direction. Online Appendix Table B.7 shows that the instrumental variable estimate for the effect on entry-grade reading achievement is very similar to the main estimate at 0.085 SD, but that the estimate for math achievement is sub-

[15] In Online Appendix Table B.5, I apply the Sojourner (2013) correction to the regression-based balancing tests of Table 4. The results confirm that peer motivation does not predict entrants' predetermined characteristics.

[16] As an alternative way to correct for missing values in peer motivation, I restrict the sample to entrants for whom most classmates are observed with motivation. Online Appendix Table B.6 shows that the effect of peer motivation on reading scores in these regressions is similar to the main effect, although the estimate is less precise due to the lower number of observations.

[17] The reduction in measurement error is larger the more classmates there are, which might explain why the effects on entry-grade achievement appear to be larger in regular-sized classes, see Figure 2.

stantially larger at 0.156 SD. Most estimates for long-term outcomes remain close to zero. Unfortunately, however, all of these effects are very imprecisely estimated. While not conclusive, these results support the pattern of positive short-term effects of peer motivation on achievement but no longer-term impacts.[18]

Missing outcome data For reasons detailed in Section 3.3, not all outcomes are observed for all students in the sample, which opens up the possibility that my results are biased by selective attrition. To address this threat, Online Appendix Table B.9 shows estimates of the effect of peer motivation on indicators for being observed with each outcome. The coefficients from the regressions of contemporaneous achievement and most other outcomes are close to zero and not statistically significant at conventional levels, showing that the likelihood of being observed with these outcomes does not systematically vary with peer motivation. The one exception are middle school test scores, for which there is a marginally statistically significant negative effect on being observed; this could potentially explain the negative (but statistically not significant) point estimates of peer motivation on middle school achievement in Table 6.

An alternative way to address the concern that missing outcome data are biasing my results is to restrict the sample to students who are consistently observed with all outcomes. I do this in Online Appendix Figure B.3, where I restrict the sample to students observed with entry-grade reading and math achievement, entry-grade motivation, and middle-school reading and math achievement.[19] The figure reveals that estimates are very similar for this restricted and the main estimation sample. Taken together, these analyses suggest that missing data on outcomes do not bias my results much.

Multiple hypothesis testing I study effects on many different outcomes, which raises the possibility that the only statistically significant effect on contemporaneous reading

[18]Given that peer motivation and peer self-concept are correlated, one might be concerned that peer self-concept is driving my main results. Online Appendix Table B.8 shows that this is not the case, as the impact of peer motivation on entry-grade achievement is almost unchanged when peer self-concept is controlled for in the regression.

[19]In unreported regressions, I confirmed that peer motivation is balanced for this restricted sample.

achievement represents a chance finding. To mitigate this threat, Online Appendix Table B.10 reports estimates of the effect of peer motivation on word study skills, which are closely related to reading skills and which were also assessed by the Stanford Achievement Test.[20] The results show that a 1 SD increase in peer motivation raises word study skills scores by a highly statistically significant 0.081 SD, an effect that is almost identical in size to the impact on reading scores. I moreover confirmed that the effects of peer motivation on reading scores and word study skills scores remain statistically significant when I correct for multiple hypothesis testing using the method developed by Romano and Wolf (2005a,b), see Online Appendix Table B.11.

6 Discussion, mechanisms, and policy implications

6.1 Discussion of main results

I now discuss the evidence on spillovers from peer motivation, starting with the estimates on short-term outcomes. The results show that exposure to motivated peers in early elementary school increases achievement on standardized tests. What stands out is that the effect for reading is larger than the one for math in almost all regressions, despite the fact that the SCAMIN measures motivation related to both subjects. While I can only speculate on the reasons for this discrepancy, the instrumental variable estimates that attempt to correct for measurement error in peer motivation are larger for math than for reading (see Online Appendix Table B.7). Therefore, one possible explanation for the lower math estimate is that attenuation bias due to measurement error is larger in the corresponding regression.

More generally, how does the size of the short-term spillovers compare with that of

[20]The correlation coefficient between reading scores and word study skills scores is 0.88. For completeness, Online Appendix Table B.10 also shows the effect on listening skills, the fourth and final skills domain assessed by the Stanford Achievement Test in both second and third grade (the correlation coefficient between reading scores and listening scores is 0.64). I do not include word study skills and listening skills in the main analysis for conciseness and in order to keep in line with the previous literature on Project STAR, which has focused almost exclusively on reading and math.

other estimates of peer effects in education? Table 5 shows that the effect on reading achievement is about half as large as the effect of a 1 SD increase in peers' past reading achievement and about the same size, in absolute value, as the effect of being exposed to a kindergarten repeater in the same sample.[21] With respect to the few existing estimates of spillovers from peer personality, my estimates for reading are larger than those found in higher education settings by Golsteyn, Non, and Zölitz (2021) and Hancock and Hill (2022), whose main spillover estimates are 0.02 SD and 0.03 SD, respectively, but smaller than those found for 12-year-old students by Shure (2021), whose main spillover estimates range from 0.12 SD to 0.15 SD. While comparing estimates across different dimensions of personality and settings is difficult, these results appear broadly in line with the idea that skills are more malleable early in life (Kautz et al., 2014), and that therefore spillovers from peer personality are stronger at earlier ages.

Turning to long-term effects, throughout my analyses, I find no evidence that peer personality in early elementary school affects educational outcomes beyond the short term. This is somewhat surprising: given dynamic complementarities, one would expect some longer-term effects. However, it is important to note that the pattern of impacts is consistent with previous studies on childhood interventions, which have found that treatments are particularly successful at changing longer-term outcomes if they affect children's personality (e.g. Heckman, Pinto, and Savelyev, 2013), and with earlier papers on peer effects, which have argued that school peers influence children's long-term educational and labor market success mainly via their impact on non-cognitive skills (e.g. Carrell, Hoekstra, and Kuka, 2018; Bietenbeck, 2020). Thus, the absence of longer-term impacts of peer motivation might be due to the lack of an effect on own motivation. Perhaps the contemporaneous impact on reading scores by itself is simply not large

[21] In Bietenbeck (2020), I also document a negative effect of exposure to kindergarten repeaters on contemporaneous achievement. However, that paper focuses on first-time kindergarten students as the treated group, a sample that is significantly less disadvantaged than the second- and third-grade entrants considered here. In contrast to the results shown in Table 5, the effect of repeater exposure on kindergarten achievement is much stronger for math than for reading.

enough to generate measurable long-term effects.

In the end, however, I cannot provide definite evidence on the why the short-term and long-term impacts appear to differ. Factors that cannot be observed in the data, such as compensatory behavior by parents, might play a role. Moreover, the precision of my estimates does not let me rule out small positive effects of peer motivation also on long-term outcomes. Ultimately, the question whether peer personality matters also for long-term educational success will therefore have to be answered by future research.

6.2 Mechanism for short-term spillovers

I now discuss potential mechanisms behind the effect of peer motivation on contemporaneous achievement. First, the experimental setup lets me rule out the most obvious explanations that involve selection into peer groups, sorting to specific teachers, and selection into the sample. Second, another intuitive explanation is that peer motivation influences children's own personality, which in turn affects achievement. However, my results above provide no evidence of such personality change. A related possibility is that exposure to motivated peers changes students' norms about studying or doing homework. While I cannot observe such norms, this explanation is difficult to reconcile with the null effect on own motivation and with the apparent differences in effect size by class size and teacher experience, as it is unclear ex ante why studying norms should be influenced by these variables.

Third, yet another alternative mechanism is that motivated peers create a good learning environment in the classroom. As shown in Section 4, motivated students score higher on the discipline index, which measures the extent to which they (do not) interfere with their classmates' learning. Motivated students are also rated higher on other dimensions of good classroom behavior by their teachers. This implies that entrants whose peers are more motivated likely experience less distraction from them, which in turn could account for the documented increase in achievement. While I cannot provide direct evidence in favor of it, I consider this the most likely mechanism behind the positive spillover effects from motivated peers.

6.3 Policy implications

Focusing on the short-term impact of peer motivation on achievement, I now consider potential policy implications of my findings. The often implicit promise of peer effects is that one may be able to improve average student outcomes by optimally assigning students to classes. Importantly, any such improvement in average outcomes requires peer effects not to be linear in means (e.g. Hoxby and Weingarth, 2005). I test for such non-linearity by asking whether exposure to peers with particularly low or particularly high motivation has a disproportionate effect on achievement, in line with the "bad apple" and "shining light" models of peer effects suggested by Hoxby and Weingarth (2005). For this purpose, I replace the average peer motivation term in Equation 2 with the shares of classmates with top-tercile and bottom-tercile motivation scores. I estimate effects both for the full sample of students and separately for students with low and high predicted achievement.

Table 7 shows the results. Columns 1 and 4 reveal that the effect of peer motivation is driven by students with very low motivation, as exposure to such "bad apples" has a large negative effect on achievement. The results in the other columns further reveal that low-predicted-achievement entrants are disproportionately hurt by the presence of such students. One potential implication of these results is that average achievement might be improved by systematically assigning low-predicted-achievement students to more motivated peers. Alternatively, students with very low motivation might be placed into separate classes (although my data do not allow me to estimate how such students affect each other). However, Carrell, Sacerdote, and West (2013) provide a cautionary tale of actually implementing such "optimal" assignment policies: they show that endogenous peer group formation may offset the predicted gains from reassignment, something that

I cannot rule out would happen even in my setting.

Rather than providing a blueprint for optimally assigning students to classes, my results speak to the kind of targeted programs that previous research has shown can effectively change aspects of personality, including motivation, in children. In particular, my findings suggest that the benefits of such interventions may be underestimated, as the generated improvements in personality for treated children will positively affect the learning outcomes of their peers. Incorporating such spillover benefits in the evaluation of such interventions thus appears important.

7 Conclusion

Previous research in economics and psychology has documented the importance of personality for individuals' own life success. However, despite extensive evidence that peers matter for performance in school and in the workplace, only very few studies have examined spillovers of personality in the social environment. This paper helps fill this gap by showing that academic motivation, which is a key aspect of personality in the context of education, affects peers' educational success.

My empirical analysis exploits the random assignment of students to classes in elementary schools in Project STAR. I find that being assigned to more motivated classmates causally increases achievement on a standardized reading test at the end of the school year. This peer effect operates over and above spillovers of classmates' academic ability and socio-demographic composition, which suggests that it reflects a true personality spillover. Since peer motivation does not affect own motivation, I argue that the positive spillover on achievement is most likely due to an improved classroom learning environment: as I show, motivated students tend to distract their classmates less. The lack of an effect on own motivation also offers an explanation for the null effect of peer motivation on longer-term educational success.

My findings suggest that the benefits of interventions which positively affect children's personality may be underestimated, as the generated improvements for treated children will positively affect the learning outcomes of their peers. More generally, I show the effects of any educational input that has an impact on personality may extend beyond the students who are targeted, as personality affects other people in their social environment.

References

Alan, S., T. Boneva, and S. Ertac. 2019. "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit." *The Quarterly Journal of Economics* 134:1121–1162.

Alan, S., and S. Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126:1865–1911.

Almlund, M., A.L. Duckworth, J. Heckman, and T. Kautz. 2011. "Personality Psychology and Economics." In E. A. Hanushek, S. Machin, and L. Woessmann, eds. *Handbook of the Economics of Education*. Elsevier, vol. 4, pp. 1–181.

Altonji, J.G., T.E. Elder, and C.R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113:151–184.

Ammermueller, A., and J.S. Pischke. 2009. "Peer effects in European primary schools: Evidence from the progress in international reading literacy study." *Journal of Labor Economics* 27:315–348.

Ballis, B. 2023. "Dreamers and Beyond: Examining the Broader Educational Effects of DACA." *Journal of Human Resources* forthcoming.

Bertrand, M., and J. Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5:32–64.

Bietenbeck, J. 2020. "The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR." *Journal of the European Economic Association* 18:392–426.

Booij, A.S., E. Leuven, and H. Oosterbeek. 2017. "Ability Peer Effects in University: Evidence from a Randomized Experiment." *The Review of Economic Studies* 84:547–578.

Borghans, L., A.L. Duckworth, J.J. Heckman, and B. Ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43:972–1059.

Brenoe, A.A., and U. Zölitz. 2019. "Exposure to More Female Peers Widens the Gender Gap in STEM Participation." *Journal of Labor Economics* 38:1009–1054.

Cadena, B.C., and B.J. Keys. 2015. "Human Capital and the Lifetime Costs of Impatience." *American Economic Journal: Economic Policy* 7:126–153.

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

Carrell, S.E., and M.L. Hoekstra. 2010. "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids." *American Economic Journal: Applied Economics* 2:211–228.

Carrell, S.E., B.I. Sacerdote, and J.E. West. 2013. "From natural variation to optimal policy? The importance of endogenous peer group formation." *Econometrica* 81:855–882.

Chetty, R., J.N. Friedman, N. Hilger, E. Saez, D.W. Schanzenbach, and D. Yagan.

2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *The Quarterly Journal of Economics* 126:1593–1660.

Clarke, D., J.P. Romano, and M. Wolf. 2020. "The Romano-Wolf Multiple-Hypothesis Correction in Stata." *The Stata Journal* 20:812–843.

Davis, T.M., and J.M. Johnston. 1987. "On the Stability and Internal Consistency of the Self-Concept and Motivation Inventory: Preschool/Kindergarten Form." *Psychological Reports* 61:871–874.

Davis, T.M., P.A. Sellers, and J.M. Johnston. 1988. "The Factor Structure and Internal Consistency of the Self-Concept and Motivation Inventory: What Face Would You Wear? Preschool/Kindergarten Form." *Educational and Psychological Measurement* 48:237–246.

Drummond, R.J., and W.G. McIntire. 1975. "Note on Test-Retest Reliability of the Self-Concept and Motivation Inventory." *Psychological Reports* 36:563–566.

Duckworth, A.L., C. Peterson, M.D. Matthews, and D.R. Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology* 92:1087–1101.

Dunifon, R., and G.J. Duncan. 1998. "Long-run effects of motivation on labor-market success." *Social Psychology Quarterly*, pp. 33–48.

Feld, J., and U. Zölitz. 2017. "Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects." *Journal of Labor Economics* 35:387–428.

Figlio, D.N. 2007. "Boys Named Sue: Disruptive Children and Their Peers." *Education Finance and Policy* 2:376–394.

Finn, J.D., and C.M. Achilles. 1990. "Answers and questions about class size: A statewide experiment." *American Educational Research Journal* 27:557–577.

Finn, J.D., J. Boyd-Zaharias, R.M. Fish, and S.B. Gerber. 2007. "Project STAR and Beyond: Database User's Guide." Report, HEROS Incorporated.

Finn, J.D., and D. Cox. 1992. "Participation and Withdrawal among Fourth-Grade Pupils." *American Educational Research Journal* 29:141–162.

Gensowski, M. 2018. "Personality, IQ, and Lifetime Earnings." *Labour Economics* 51:170–183.

Golsteyn, B.H., H. Grönqvist, and L. Lindahl. 2014. "Adolescent Time Preferences Predict Lifetime Outcomes." *The Economic Journal* 124:F739–F761.

Golsteyn, B.H., A. Non, and U. Zölitz. 2021. "The Impact of Peer Personality on Academic Achievement." *Journal of Political Economy* 129:1052–1099.

Guryan, J., K. Kroft, and M.J. Notowidigdo. 2009. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments." *American Economic Journal: Applied Economics* 1:34–68.

Hancock, S.A., and A.J. Hill. 2022. "The Effect of Teammate Personality on Team Production." *Labour Economics* 78:102248.

Heckman, J., R. Pinto, and P. Savelyev. 2013. "Understanding the Mechanisms through which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103:2052–2086.

Heckman, J.J., T. Jagelka, and T.D. Kautz. 2019. "Some Contributions of Economics to the Study of Personality." NBER Working Paper No. 26459.

Hoxby, C. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper No. 7867.

Hoxby, C.M., and G. Weingarth. 2005. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." Working paper, Harvard University.

Hulleman, C.S., and K.E. Barron. 2015. "Motivation Interventions in Education: Bridging Theory, Research, and Practice." In *Handbook of Educational Psychology*. Routledge, pp. 174–185.

Kautz, T., J.J. Heckman, R. Diris, B.t. Weel, and L. Borghans. 2014. "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success." OECD Education Working Paper No. 110.

Krueger, A.B. 1999. "Experimental Estimates of Education Production Functions." *The Quarterly Journal of Economics* 114:497–532.

Krueger, A.B., and D.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." *The Economic Journal* 111:1–28.

Lavy, V., M.D. Paserman, and A. Schlosser. 2012. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom." *The Economic Journal* 122:208–237.

Lavy, V., and A. Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics* 3:1–33.

Lazowski, R.A., and C.S. Hulleman. 2016. "Motivation Interventions in Education: A Meta-Analytic Review." *Review of Educational Research* 86:602–640.

Mas, A., and E. Moretti. 2009. "Peers at Work." *American Economic Review* 99:112–145.

McIntire, W.G., and R.J. Drummond. 1976. "The Structure of Self-Concept in Second and Fourth Grade Children." *Educational and Psychological Measurement* 36:529–536.

Measelle, J.R., O.P. John, J.C. Ablow, P.A. Cowan, and C.P. Cowan. 2005. "Can Children Provide Coherent, Stable, and Valid Self-Reports on the Big Five Dimensions? A Longitudinal Study From Ages 5 to 7." *Journal of Personality and Social Psychology* 89:90–106.

Milchus, N.J., G.A. Farrah, and W. Reitz. 1968. *The Self-concept and Motivation Inventory: What Face Would You Wear?*. Dearborn Heights, MI: Person-O-Metrics.

Murphy, P.K., and P.A. Alexander. 2000. "A Motivated Exploration of Motivation Terminology." *Contemporary Educational Psychology* 25:3–53.

Oster, E. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37:187–204.

Paloyo, A.R. 2020. "Peer Effects in Education: Recent Empirical Evidence." In S. Bradley and C. Green, eds. *The Economics of Education (Second Edition)*. Academic Press, pp. 291–305.

Piatek, R., and P. Pinger. 2016. "Maintaining (Locus of) Control? Data Combination for the Identification and Inference of Factor Structure Models." *Journal of Applied Econometrics* 31:734–755.

Poropat, A.E. 2009. "A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance." *Psychological Bulletin* 135:322–338.

Roberts, B.W. 2006. "Personality Development and Organizational Behavior." *Research in Organizational Behavior* 27:1–40.

Roberts, B.W., and W.F. DelVecchio. 2000. "The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies." *Psychological Bulletin* 126:3–25.

Roberts, B.W., P. Harms, J.L. Smith, D. Wood, and M. Webb. 2006. "Using Multiple Methods in Personality Psychology." In M. Eid and E. Diener, eds. *Handbook of Multimethod Measurement in Psychology*. American Psychological Association, pp. 321–335.

Roberts, B.W., and H.J. Yoon. 2022. "Personality psychology." *Annual review of psychology* 73:489–516.

Romano, J.P., and M. Wolf. 2005a. "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing." *Journal of the American Statistical Association* 100:94–108.

—. 2005b. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73:1237–1282.

Sacerdote, B. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In E. A. Hanushek, S. Machin, and L. Woessmann, eds. *Handbook of the Economics of Education*. Elsevier, vol. 3, pp. 249–277.

Schanzenbach, D.W. 2006. "What Have Researchers Learned from Project STAR?" *Brookings Papers on Education Policy* 9:205–228.

Segal, C. 2012. "Working when no one is watching: Motivation, test scores, and economic success." *Management Science* 58:1438–1457.

Shavelson, R.J., J.J. Hubner, and G.C. Stanton. 1976. "Self-Concept: Validation of Construct Interpretations." *Review of Educational Research* 46:407–441.

Shure, N. 2021. "Non-cognitive peer effects in secondary education." *Labour Economics* 73:102074.

Sojourner, A. 2013. “Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR.” *The Economic Journal* 123:574–605.

Sorrenti, G., U. Zölitz, D. Ribeaud, and M. Eisner. 2020. “The Causal Impact of Socio-Emotional Skills Training on Educational Success.” CEPR Discussion Paper No. 14523.

Steinmayr, R., and B. Spinath. 2009. “The Importance of Motivation as a Predictor of School Achievement.” *Learning and Individual Differences* 19:80–90.

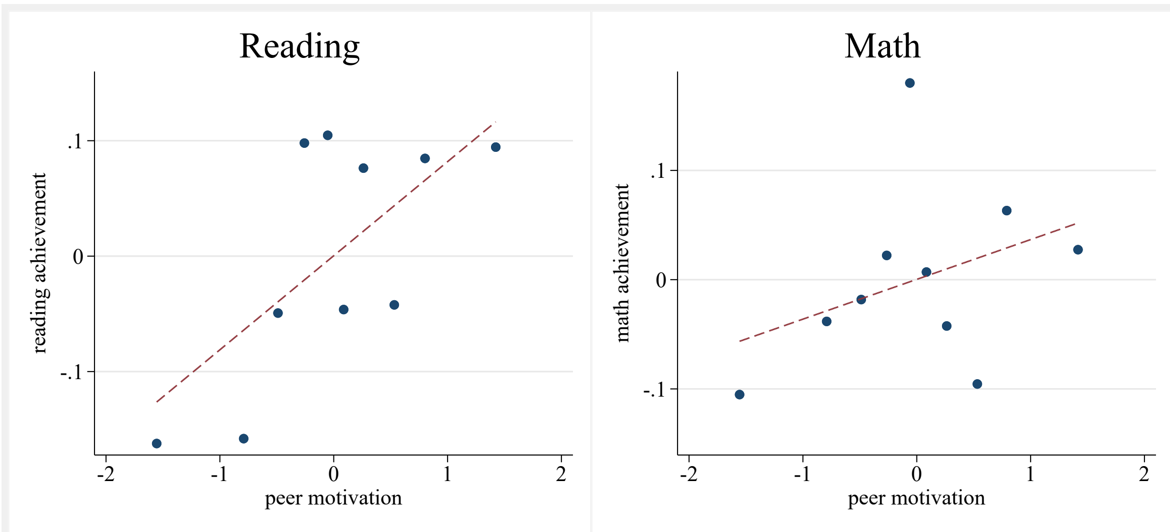
Whitmore, D. 2005. “Resource and Peer Impacts on Girls’ Academic Achievement: Evidence from a Randomized Experiment.” *American Economic Review* 95:199–204.

Wong, M.M., and M. Csikszentmihalyi. 1991. “Motivation and Academic Achievement: The Effects of Personality Traits and the Duality of Experience.” *Journal of Personality* 59:539–574.

Word, E., J. Johnston, H.P. Bain, D.B. Fulton, C.M. Achilles, M.N. Lintz, J. Folger, and C. Breda. 1990. “The State of Tennessee’s Student/Teacher Achievement Ratio (STAR) Project: Technical Report 1985-1990.” Report, Tennessee State Department of Education.

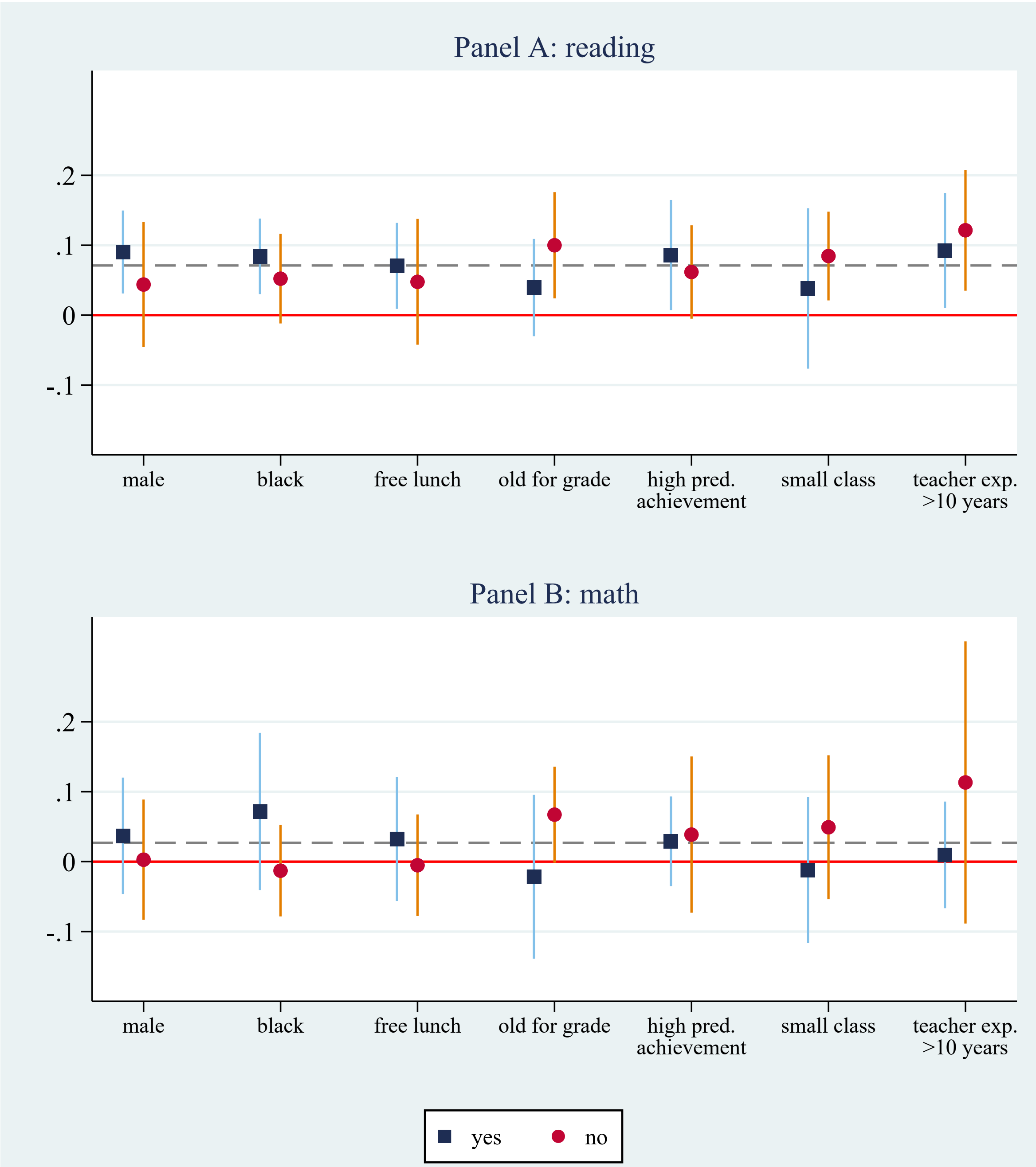
Figures and Tables

Figure 1: Peer motivation and entry-grade achievement



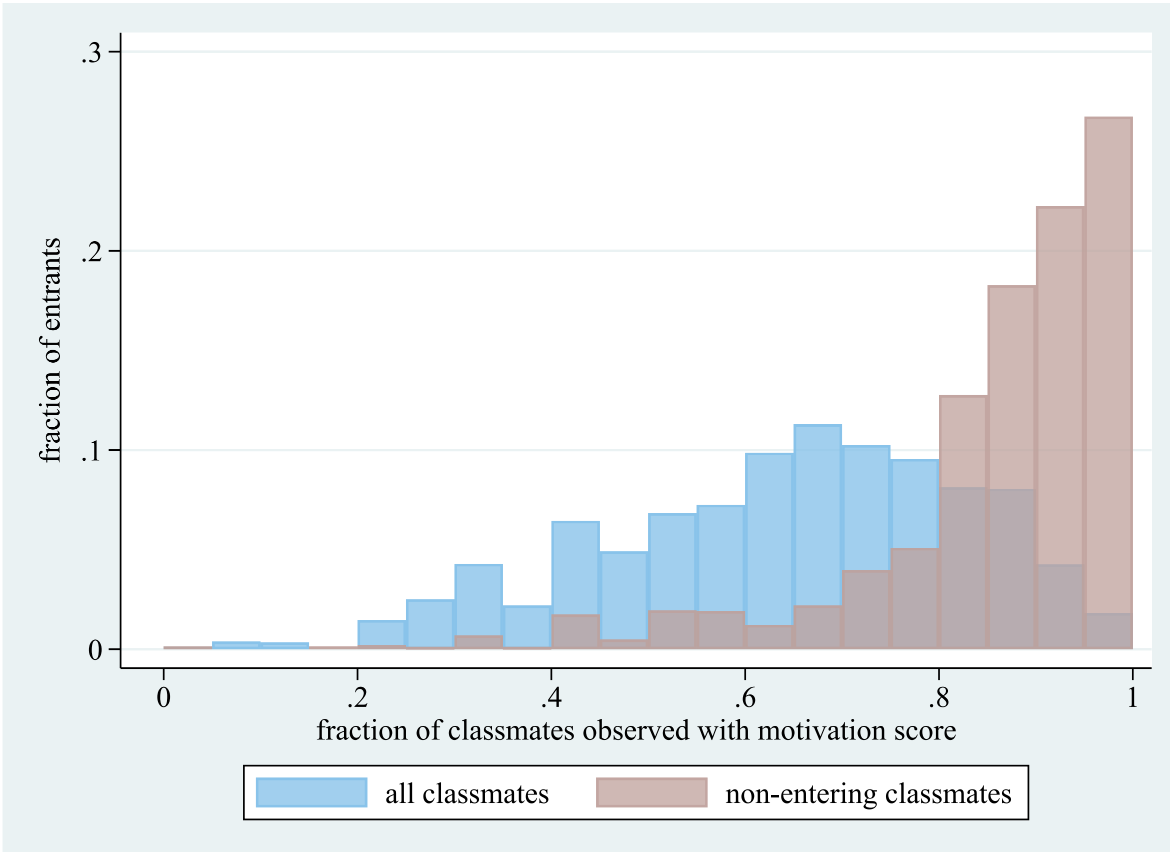
Notes: The figure shows estimates of the effect of peer motivation on reading and math achievement at the end of entrants' first year in Project STAR. To construct these plots, I first residualize achievement scores and peer motivation on the controls included in the specifications in columns 1 and 4 of Table 5. I then group residualized peer motivation into ten equal-sized bins and plot the mean of the residualized achievement scores for each bin. The regression line in each plot is based on the underlying individual-level data and thus visualizes the corresponding regression in Table 5.

Figure 2: Peer motivation and entry-grade achievement, heterogeneity by student characteristics, class size, and teacher experience



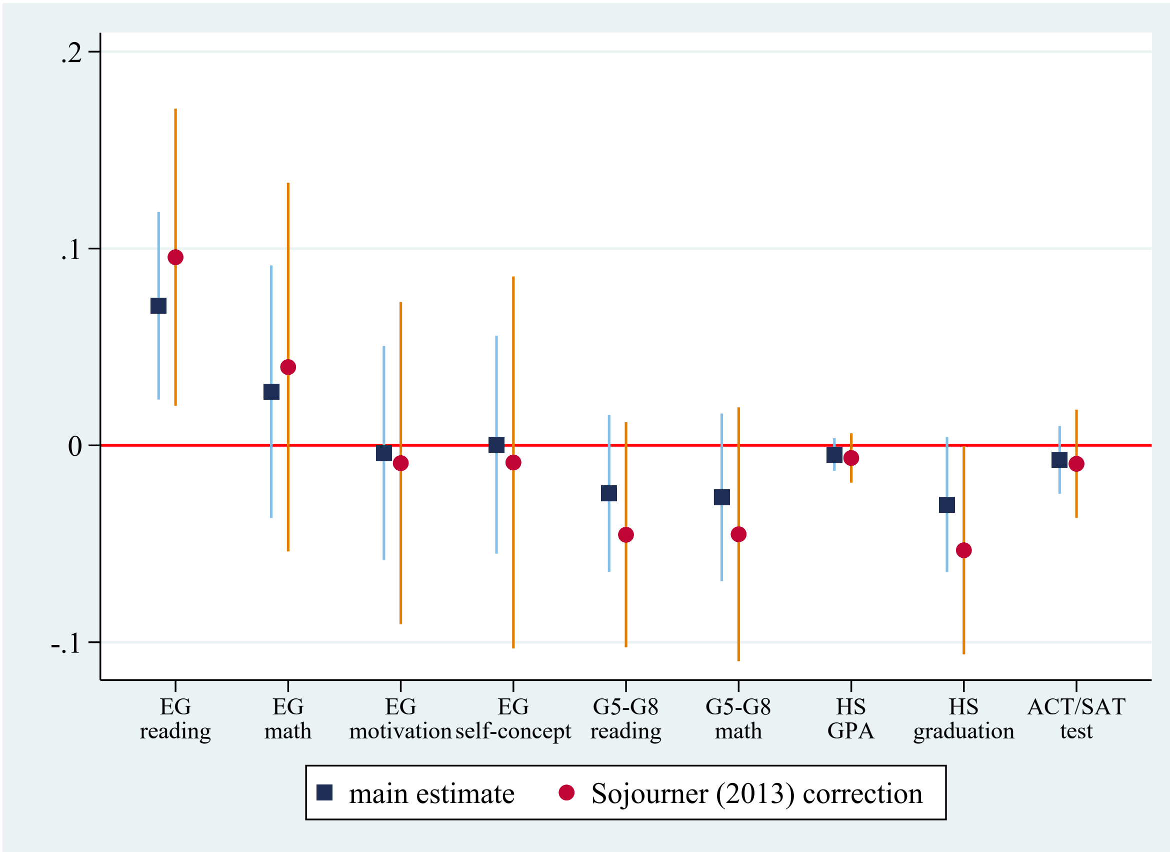
Notes: The figure shows point estimates and 95 percent confidence intervals of the effect of peer motivation on achievement in reading (panel A) and math (panel B), separately for different groups of students. The specifications correspond to the ones in columns 3 and 6 of Table 5 but focus on subsamples of students as indicated on the horizontal axes: squares indicate point estimates for students with the respective characteristic, and circles indicate point estimates for students without this characteristic. High predicted achievement is an indicator for whether predicted achievement is above average. The 10-year cutoff for teacher experience is chosen for consistency with the analysis of the role of teacher experience in Project STAR in Chetty et al. (2011). The dashed line in each panel shows the main estimate for all students from columns 3 and 6 of Table 5 .

Figure 3: Shares of classmates observed with motivation scores



Notes: The histograms show the fractions of classmates observed with motivation scores. The light blue distribution refers to all classmates, and the brown distribution refers to the subsample of classmates who had participated in Project STAR during the previous schools year (that is, classmates who did not enter the experiment in the same year as the second- or third-year entrant for whom the share is calculated). Statistics for the distribution for all classmates: mean = .67, median = .64, p25 = .5, p75 = .78. Statistics for the distribution for non-entering classmates: mean = .86, median = .9, p25 = .82, p75 = .95.

Figure 4: Correcting bias due to missing information on classmates' motivation



Notes: The figure shows point estimates and 95 percent confidence intervals from regressions of the outcome variables indicated on the horizontal axis on peer motivation. The blue squares correspond to the main estimates shown in Table 5 and Table 6. The red circles show estimates based on the correction proposed by Sojourner (2013). The only difference in these corrected estimates is that peer motivation is multiplied with the individual-level share of classmates observed with motivation. EG = entry-grade, G5 = grade 5, G8 = grade 8, HS = high school. For this figure only, high school GPA is re-scaled to range from 0-1.

Table 1: Correlates of motivation

	Grades 1-3 motivation				
	(1)	(2)	(3)	(4)	(5)
Male	–0.292***				–0.285***
	(0.023)				(0.023)
Black		–0.026			–0.023
		(0.046)			(0.050)
Free lunch			–0.002		0.011
			(0.026)		(0.027)
Age in years				0.065*	0.078**
				(0.034)	(0.033)
Old for grade				–0.214***	–0.190***
				(0.047)	(0.047)
Small class	–0.000	–0.001	–0.001	–0.004	–0.003
	(0.027)	(0.028)	(0.028)	(0.028)	(0.027)
Observations	9,072	9,072	9,072	9,072	9,072

Notes: The table shows estimates of regressions of students' average motivation in grades 1-3 on student socio-demographic characteristics and a dummy for assignment to small class upon entry into Project STAR. The sample includes the 9,072 students for whom a motivation score is observed in at least one of grades 1, 2, and 3. The dependent variable is standardized to have mean 0 and SD 1. All regressions control for school-by-entry-grade fixed effects (regression that omit these fixed effects yield very similar results). Standard errors in parentheses are clustered by school-by-entry-grade. * p<0.10, ** p<0.05, *** p<0.01.

Table 2: Own motivation, classroom behavior, and educational success

Panel A: classroom behavior

	Grade 4				Grade 8			
	effort (1)	initiative (2)	discipline (3)	value (4)	effort (5)	initiative (6)	discipline (7)	value (8)
Grades 1-3 motivation	0.102***	0.081***	0.090***	0.110***	0.060***	0.038*	0.065***	0.078***
	(0.027)	(0.026)	(0.028)	(0.031)	(0.022)	(0.023)	(0.024)	(0.024)
Male	–0.357***	–0.218***	–0.490***	–0.350***	–0.440***	–0.264***	–0.520***	–0.395***
	(0.041)	(0.044)	(0.049)	(0.042)	(0.039)	(0.046)	(0.041)	(0.043)
Free lunch	–0.423***	–0.472***	–0.228***	–0.248***	–0.304***	–0.266***	–0.242***	–0.198***
	(0.053)	(0.057)	(0.049)	(0.054)	(0.046)	(0.044)	(0.045)	(0.052)
Observations	2,212	2,212	2,212	2,212	2,693	2,693	2,693	2,693

Panel B: educational success

	Grades 1-3		Grades 5-8		High school		
	reading (1)	math (2)	reading (3)	math (4)	GPA (5)	grad. (6)	ACT/SAT (7)
Grades 1-3 motivation	0.045***	0.051***	0.049***	0.052***	0.285*	0.007	0.015***
	(0.011)	(0.012)	(0.014)	(0.013)	(0.149)	(0.007)	(0.005)
Male	–0.184***	0.008	–0.100***	–0.138***	–3.080***	–0.070***	–0.133***
	(0.021)	(0.020)	(0.022)	(0.023)	(0.274)	(0.012)	(0.010)
Free lunch	–0.424***	–0.408***	–0.460***	–0.428***	–3.437***	–0.140***	–0.269***
	(0.029)	(0.028)	(0.028)	(0.028)	(0.374)	(0.014)	(0.015)
Observations	8,530	8,678	7,497	7,493	3,360	4,368	9,072

Notes: The table shows estimates from regressions of the outcome variables indicated in the column headers on students' academic motivation, averaged across grades 1-3. Measures of classroom behavior in panel A are standardized to have mean 0 and SD 1. In columns 1-4 of panel B, test scores are averaged across the grades indicated in the column headers and are then standardized to have mean 0 and SD 1. The sample means of the high school outcomes used in columns 5-7 of panel B are: 83.5 (GPA), 0.82 (graduation), 0.38 (ACT/SAT test-taking). Sample sizes differ across outcomes because of different data collection procedures and sample attrition, see Online Appendix A for details. All regressions in Panels A and B control for school-by-entry-grade fixed effects, dummies for male, black, eligibility for free or reduced-price lunch, old for grade, and age. Standard errors in parentheses are clustered by school-by-entry-grade.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3: Summary statistics for the peer motivation sample

	Mean	SD	N
Socio-demographic characteristics			
Male	0.55	0.50	2,861
Black	0.42	0.49	2,766
Free lunch	0.66	0.47	2,730
Age in 1985	6.01	0.70	2,845
Old for grade	0.47	0.50	2,845
Peer motivation and other peer characteristics			
Peer motivation	0.00	1.00	2,868
Peer reading achievement	0.00	1.00	2,841
Peer math achievement	0.00	1.00	2,850
KG repeater peer in class	0.21	0.40	2,868
Peer share male	0.51	0.11	2,868
Peer share black	0.42	0.43	2,868
Peer share free lunch	0.61	0.30	2,868
Entry-grade achievement			
Reading score	0.00	1.00	2,185
Math score	0.00	1.00	2,196
Entry-grade own personality			
Own motivation	0.00	1.00	2,276
Own self-concept	0.00	1.00	2,276
Long-term educational outcomes			
Reading scores in grades 5-8	0.00	1.00	2,118
Math scores in grades 5-8	0.00	1.00	2,119
High school GPA (0-100)	81.50	7.46	665
High school graduation	0.73	0.44	1,018
Took ACT/SAT	0.26	0.44	2,868

Notes: The table shows means and standard deviations and the number of students observed with each variable for the 2,868 students included in the peer motivation sample. KG repeater refers to a child who repeated kindergarten in the first year of Project STAR.

Table 4: Balancing tests for peer motivation and peer achievement

	Male	Black	Free lunch	Age	Old for grade	Pred. achievement
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: separate regressions for each peer variable						
Peer motivation	0.002	–0.007	–0.005	–0.023	–0.004	0.014
	(0.012)	(0.006)	(0.009)	(0.017)	(0.011)	(0.018)
Peer reading achievement	0.017	–0.008	–0.014	–0.024	–0.005	0.024
	(0.015)	(0.009)	(0.021)	(0.023)	(0.015)	(0.029)
Peer math achievement	0.024	–0.012	–0.028*	–0.020	–0.010	0.038
	(0.015)	(0.010)	(0.016)	(0.028)	(0.019)	(0.031)
Panel B: joint regressions for all peer variables						
Peer motivation	0.002	–0.007	–0.004	–0.022	–0.004	0.014
	(0.012)	(0.006)	(0.010)	(0.016)	(0.011)	(0.018)
Peer reading achievement	–0.000	0.002	0.009	–0.015	0.003	–0.006
	(0.021)	(0.010)	(0.029)	(0.029)	(0.019)	(0.038)
Peer math achievement	0.024	–0.013	–0.033	–0.009	–0.012	0.042
	(0.021)	(0.012)	(0.020)	(0.036)	(0.025)	(0.039)
p-value (joint significance)	0.44	0.37	0.22	0.42	0.95	0.59
Observations (both panels)	2,861	2,766	2,730	2,845	2,845	2,868

Notes: The table shows estimates of regressions of students' socio-demographic characteristics and predicted achievement on the characteristics of their classmates. Estimates are based on the peer motivation sample. In Panel A, each coefficient comes from a separate regression of the outcome indicated in the column header on the peer variable indicated in the row. In Panel B, coefficients are instead based on a single regression in which all peer variables enter jointly. The p-value reported in Panel B comes from an F test for the joint significance of the three peer variables. All regressions in both panels control for school-by-entry-grade fixed effects. Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 5: Peer motivation and entry-grade achievement

	Reading			Math		
	(1)	(2)	(3)	(4)	(5)	(6)
Peer motivation	0.081***	0.074***	0.071***	0.036	0.032	0.027
	(0.023)	(0.023)	(0.024)	(0.032)	(0.031)	(0.032)
Peer reading achievement		0.154**	0.152**		0.150**	0.134**
		(0.064)	(0.066)		(0.067)	(0.067)
Peer math achievement		0.038	0.042		0.051	0.062
		(0.058)	(0.059)		(0.057)	(0.058)
KG repeater peer in class		–0.069	–0.077		0.004	–0.003
		(0.073)	(0.073)		(0.086)	(0.088)
Peer share male			–0.194			–0.421*
			(0.271)			(0.233)
Peer share free lunch			0.146			0.006
			(0.252)			(0.282)
Peer share black			0.158			0.036
			(0.307)			(0.333)
Observations	2,185	2,185	2,185	2,196	2,196	2,196

Notes: The table shows estimates of the effect of peer motivation on achievement in reading (columns 1-3) and math (columns 4-6) at the end of students' first year in Project STAR. Estimates are based on the peer motivation sample. All regressions control for own socio-demographic characteristics, a dummy for small class, and school-by-entry-grade fixed effects. Regressions in columns 2, 3, 5, and 6 additionally control for averages of classmates' reading and math achievement in the previous school year and an indicator for whether the class includes a kindergarten repeater, and regressions in column 3 and 6 additionally control for averages of classmates' socio-demographic characteristics. Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 6: Peer motivation, entry-grade own motivation and self-concept, and long-term educational success

	Entry grade		Grades 5-8		High school		College
	motivation (1)	self-concept (2)	reading (3)	math (4)	GPA (5)	grad. (6)	ACT/SAT (7)
Peer motivation	–0.004	0.000	–0.024	–0.026	–0.467	–0.030*	–0.007
	(0.028)	(0.028)	(0.020)	(0.022)	(0.419)	(0.017)	(0.009)
Peer achievement controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,276	2,276	2,118	2,119	665	1,018	2,868

Notes: The table shows estimates of the effect of peer motivation on the outcome variables indicated in the column headers. Regressions control for own socio-demographic characteristics, averages of classmates' reading and math achievement in the previous school year, an indicator for whether the class includes a kindergarten repeater, averages of classmates' socio-demographic characteristics, a dummy for small class, and school-by-entry-grade fixed effects. Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 7: Peer motivation and entry-grade achievement, bad apples and shining lights

	Reading			Math		
	All students	By pred. achievement		All students	By pred. achievement	
		low	high		low	high
	(1)	(2)	(3)	(4)	(5)	(6)
Share peers with top 33% motivation	0.136	0.054	0.432	0.074	0.250	0.066
	(0.187)	(0.285)	(0.289)	(0.295)	(0.475)	(0.315)
Share peers with bottom 33% motivation	–0.429***	–0.459**	–0.310	–0.222	–0.290	–0.109
	(0.157)	(0.218)	(0.278)	(0.174)	(0.285)	(0.244)
Peer achievement controls	Yes	Yes	Yes	Yes	Yes	Yes
Peer demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,185	1,143	1,042	2,196	1,142	1,054

Notes: The table shows estimates of the effect of peer motivation on achievement in reading and math. Peer motivation is measured as the shares of classmates with top 33% and bottom 33% motivation scores. Regressions control for own socio-demographic characteristics, averages of classmates' reading and math achievement in the previous school year, an indicator for whether the class includes a kindergarten repeater, averages of classmates' socio-demographic characteristics, a dummy for small class, and school-by-entry-grade fixed effects. Standard errors in parentheses are clustered by school-by-entry-grade. * p<0.10, ** p<0.05, *** p<0.01.

– ONLINE APPENDIX –

A Data appendix

In this appendix, I provide additional details about the Project STAR data. The appendix is very similar, and in parts identical, to the data appendix prepared for a previous paper, in which I use data from the same experiment (Bietenbeck, 2020).

Project STAR was planned and implemented by a consortium of researchers from four universities and various state institutions in Tennessee. The experiment ran from the beginning of kindergarten until the end of third grade, but some researchers continued to collect data on participating students in the years afterwards, see Finn et al. (2007) for details. The Project STAR public use file, which is the basis for the empirical analysis in this paper, combines these data such that students can be followed throughout their scholastic careers until the end of high school. In what follows, I present the main independent and dependent variables that I draw from this dataset.

Academic motivation. As described in the main text, students participating in Project STAR were assessed on their academic motivation and self-concept using the Self-Concept and Motivation Inventory (SCAMIN; Milchus, Farrah, and Reitz, 1968) in the spring of each year from kindergarten through third grade. The group-administered, standardized psychological test asks students to indicate pictorially their response to different situations. Based on the answers, a motivation score and a self-concept score are calculated for each student. These scores are included in the public use file.

Tests in personality psychology are often judged by their levels of content-related, construct-related, and criterion validity (Borghans et al., 2008). Content-related validity concerns qualitative judgments by experts about whether a test adequately represents the psychological construct of interest. Construct-related validity refers to the degree to which a test actually measures what it claims to measure and is often assessed using factor analysis. Criterion validity concerns the ability of a test to predict contemporaneous and future outcomes. Finally, another important measure of test quality is reliability, as captured for example by test-retest correlations.

Several previous studies and my own analysis of data from Project STAR indicate a high quality of the SCAMIN early elementary form, which was administered in grades 1-3. Thus, Finn and Cox (1992) point out its strong content validity due to the careful and structured approach taken when creating questions. McIntire and Drummond (1976) show that the motivation score based on the early elementary form correlates with a conceptually related score from the more widely used Coopersmith Self-Esteem Inventory Scales, providing some evidence of construct validity. My results in Section 4 establish criterion validity, as they show that motivation scores predict a wide range of contemporaneous and future outcomes.

Regarding the reliability of the early elementary form, Drummond and McIntire (1975) calculate five-months test-retest correlations of motivation scores of 0.37 and 0.51 in samples of first and second grade students, respectively. Using data from Project STAR, I find a one-year test-retest correlation of 0.31 for both first-grade and second-grade motivation scores. These values are broadly similar to test-retest correlations found for personality traits in children: for example, Measelle et al. (2005) document one-year test-retest correlations for Big Five traits ranging from 0.33 to 0.59 in children aged six to seven, and a meta study by Roberts and DelVecchio (2000) finds an average test-retest correlation of 0.43 for Big Five Traits in children aged six to eleven.

The available evidence paints a different picture of the quality of the SCAMIN preschool/kindergarten form, which was administered in the spring of kindergarten. Thus, Davis, Sellers, and Johnston (1988) analyzed the form's questions using factor analysis and found that they could recover the motivation and self-concept subscales only after disregarding some of the questions, which casts doubt on its construct validity. Moreover, Online Appendix Table A.1 shows that kindergarten motivation scores do not predict any of the measures of educational success studied in the paper, indicating that it has very low (or indeed no) criterion validity.

As for reliability, Davis and Johnston (1987) found three-week test-retest correlations for kindergarten motivation scores of 0.45-0.58 in a sample of 167 kindergarten students. However, Online Appendix Table A.2 shows that kindergarten motivation scores are slightly *negatively* correlated with motivation scores in later grades in the larger sample of Project STAR. As the later scores based on the early elementary form are supposed to measure the same underlying construct (academic motivation), this casts serious doubt on the reliability of the motivation scores based on the preschool/kindergarten form. Given the breadth and severity of these problems, I decided not to use the kindergarten motivation scores in my analysis.[22]

Test scores. At the end of each school year from kindergarten through third grade, students in Project STAR wrote the grade-specific version of the Stanford Achievement Test. From fifth grade through eighth grade, students who were still residing in Tennessee took the Comprehensive Test of Basic Skills (CTBS) as part of a statewide testing program.[23] Both tests are standardized multiple-choice assessments with components in reading and math. The second- and third-grade versions of the Stanford Achievement Test further include tests of word study skills and listening skills.

The public use file contains Stanford Achievement Test scores for all students who took these tests. However, it contains CTBS scores only for students who were on grade level, i.e. students who attended grade 5/6/7/8 in 1991/1992/1993/1994, respectively. This implies that test scores are not observed for a number of students who had been

[22]Schanzenbach (2006) describes the reliability of the SCAMIN scale as "only moderate," citing work by Finn and Achilles (1990). As it turns out, this conclusion by Finn and Achilles (1990) is based on the analyses by Davis and Johnston (1987) and Davis, Sellers, and Johnston (1988), which only consider the preschool/kindergarten form. As described in the main text, the results from the preschool/kindergarten form, which is indeed problematic, and the early elementary form, which I show to have predictive validity and adequate test-retest reliability, are not comparable.

[23]An unrepresentative subsample of students took the CTBS also in fourth grade, see Finn et al. (2007). Due to the selective nature of this subsample, I chose not to analyze fourth-grade test scores.

retained in grade by those years.[24] Diane Schanzenbach generously provided me with a different version of the Project STAR data, which contains CTBS scores for students who attended grades 5-8 in Tennessee in any year between 1990 and 1997. Test scores are provided as scale scores, which are comparable across grade levels (Finn et al., 2007). In order to increase sample size, I define test scores for a given grade level as scores obtained in the school year in which participating students were supposed to be in that grade (e.g., eighth-grade scores are defined as scores obtained in 1994, even though some students were attending seventh grade in that year).

Classroom behavior. In November 1989, fourth-grade teachers of a subset of former participants in Project STAR were asked to rate their students on their behavior. Specifically, teachers completed a questionnaire that asked them how often each student had engaged in 31 different behaviors over the last two to three months. Ratings were recorded on a scale from 1 ("never") to 5 ("always"), and ratings of 28 of these behaviors were consolidated into four indices. The effort index includes items such as whether a student is persistent when confronted with difficult problems, whether she completes her homework, and whether she gets discouraged easily when encountering an obstacle in schoolwork. The initiative index is based on such items as whether a student participates actively in classroom discussions, whether she does more than just the assigned work, and whether she often asks questions. The discipline index captures such characteristics as whether a student often acts restless, whether she needs reprimanding, and whether she interferes with peers' work. The value index measures how much a student appreciates the school learning environment.[25]

During the 1993-94 school year, eighth-grade math and English teachers of a different

[24] Note that students who were retained in grade at any point between kindergarten and third grade dropped out of the STAR cohort and therefore did not write the subsequent Stanford Achievement Tests. However, these students did write the CTBS in later grades as long as they stayed in Tennessee.

[25] Note that what the paper refers to as the "discipline index" is the inverse of the "index of nonparticipatory behavior" in the original data. See Finn et al. (2007) for a complete listing of the behaviors included in each of the indices.

subset of participants were asked about student behaviors on a similar but shorter questionnaire. Thirteen of these behaviors were again consolidated into four indices measuring each student's effort, initiative, discipline, and value. For my analysis, I averaged the eighth-grade indices across math and English for each student.

High school GPA and graduation. Most students in Project STAR graduated from high school in 1998, and transcripts were gathered from selected high schools in 1999 and 2000. High schools were chosen for data collection based on the likelihood that participants would attend them given the locations of students' last known middle schools. Course grades from transcripts were transferred to a scale from 0-100 if necessary, and separate GPAs for math, science, and foreign languages were computed and are available in the public use file. The empirical analysis in this paper uses overall GPA, defined as the average of the these three subject-specific GPAs, as an outcome variable.

Information on high school graduation was also derived from the transcripts and cross-checked with data from the Tennessee State Department of Education in ambiguous cases. Nevertheless, graduation status could not be determined with certainty for all students. In these cases, the data collectors made a best guess whether a student "probably graduated" or "probably dropped out" based on the available course grades, information on attendance, and additional information from the Tennessee State Department of Education. The variable used in the empirical analysis codes students who graduated, students who probably graduated, and students who received a General Educational Development certificate as graduates, and students who dropped out and students who probably dropped out as dropouts.

College-test taking. ACT/SAT-test taking was recorded by Krueger and Whitmore (2001), who matched all students in Project STAR to the administrative records of the two companies responsible for these tests in 1998. The outcome variable used in the empirical analysis is an indicator that takes value 1 if a student took either of these college entrance exams in 1998 and 0 otherwise.

Online Appendix Table A.1: Kindergarten motivation and educational success

	Kindergarten		Grades 1-3		Grades 5-8		High school		College
	reading score	math score	reading scores	math scores	reading scores	math scores	GPA	gradua-tion	took ACT/SAT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Motivation in KG	0.006	0.002	–0.016	–0.017	–0.017	0.001	0.025	–0.007	0.006
	(0.015)	(0.014)	(0.017)	(0.017)	(0.018)	(0.017)	(0.174)	(0.009)	(0.007)
Observations	5,038	5,038	3,716	3,774	4,051	4,049	2,015	2,456	5,038

Notes: The table shows estimates from regressions of the outcome variables indicated in the column headers on students' academic motivation in kindergarten. All regressions control for school-by-entry-grade fixed effects, dummies for male, black, and eligibility for free or reduced-price lunch, and age. Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

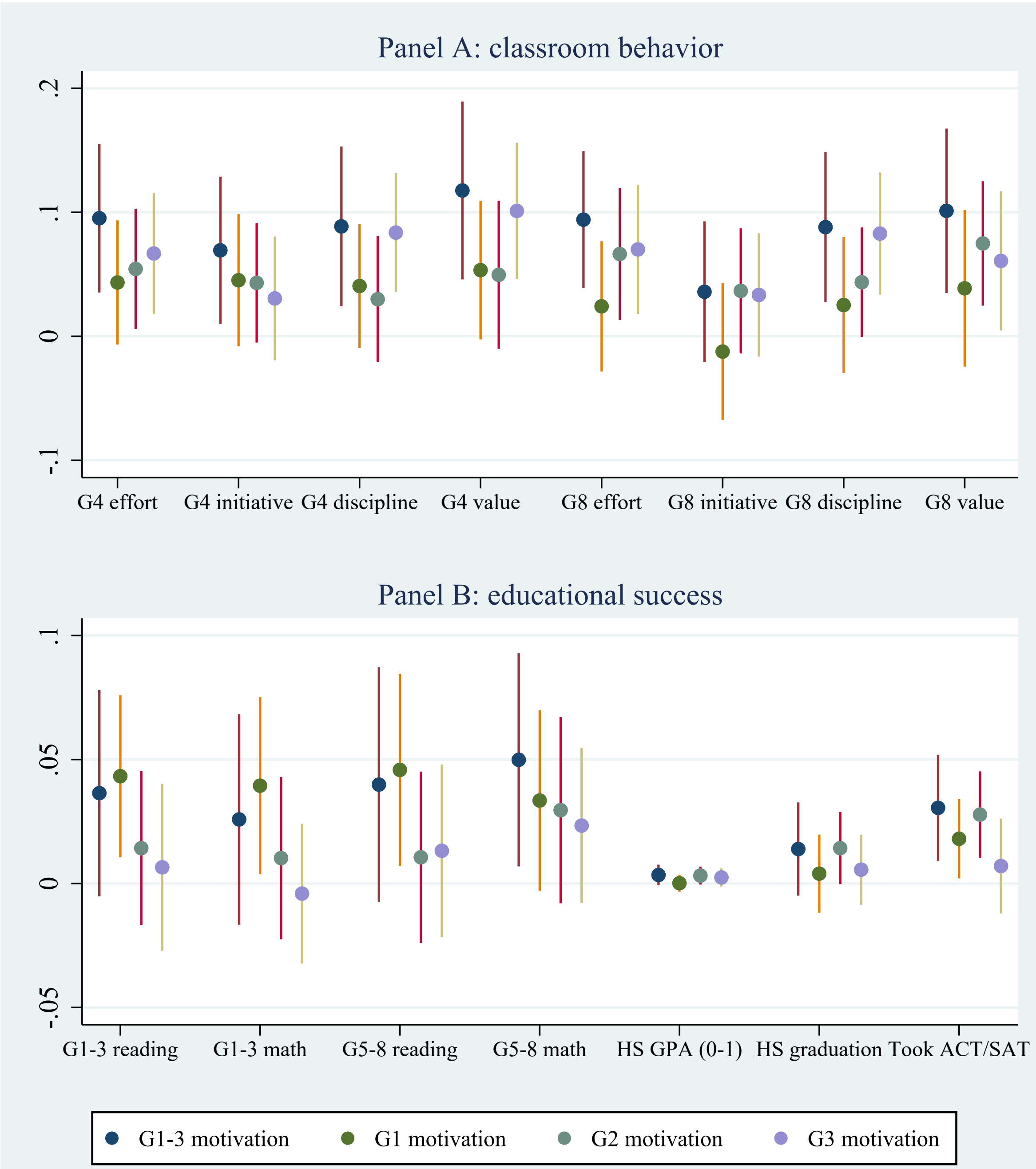
Online Appendix Table A.2: Correlations between motivation scores in different grades

Motivation	Kindergarten	Grade 1	Grade 2	Grade 3
Kindergarten	1.000			
Grade 1	-0.042	1.000		
Grade 2	-0.056	0.309	1.000	
Grade 3	-0.047	0.220	0.313	1.000

Notes: The table shows correlations between motivation scores in different grades.

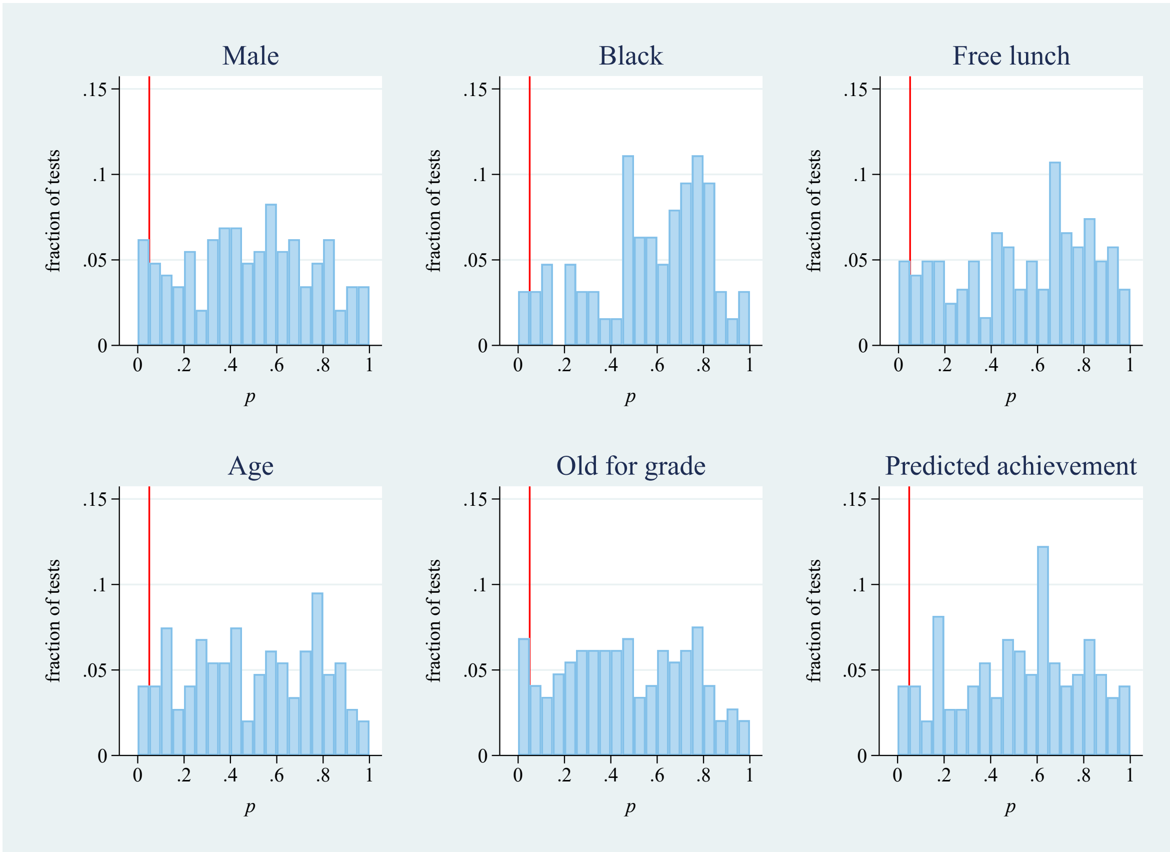
B Results from additional analyses

Online Appendix Figure B.1: Own motivation, classroom behavior, and educational success, by grade in which motivation is measured



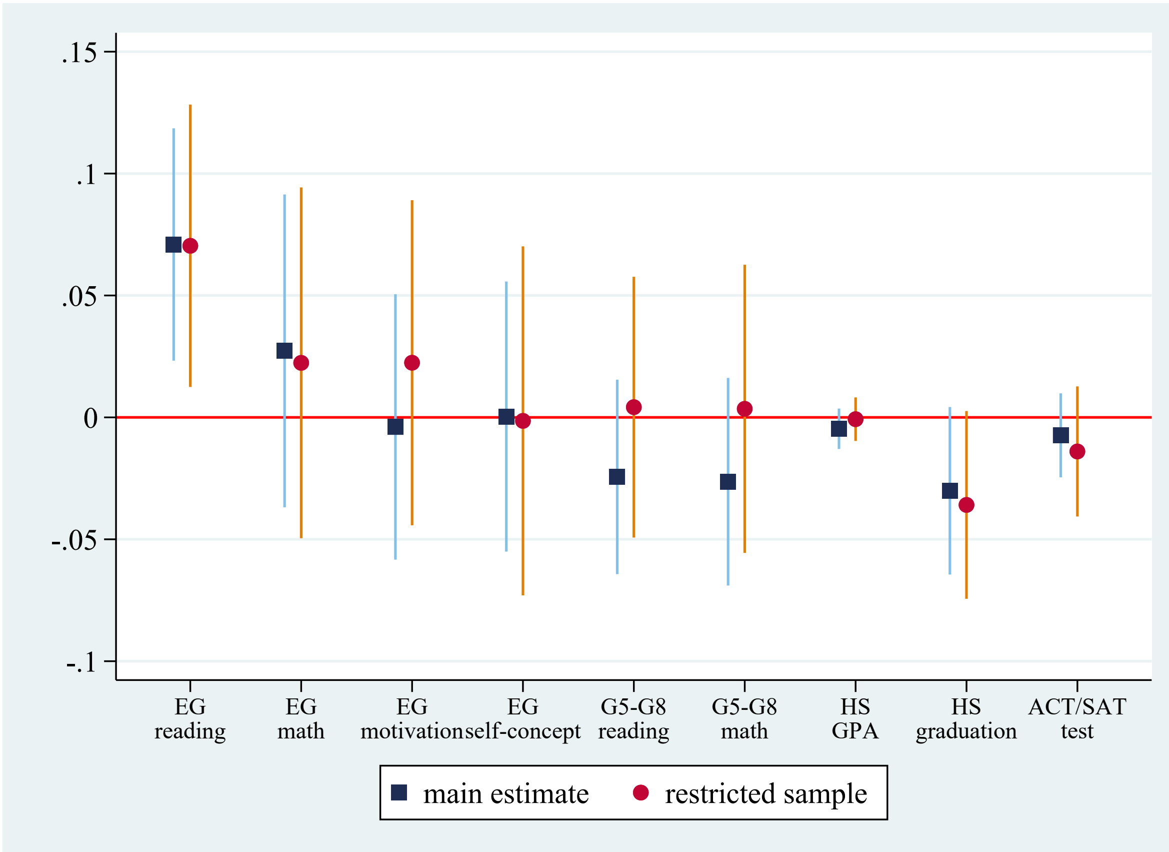
Notes: The figures shows point estimates and 95 percent confidence intervals from regressions of the outcome variables indicated on the horizontal axes on students' motivation in grades 1-3. G1 refers to grade 1, G2 to grade 2, etc. The analysis is based on the subsample of students who are observed with motivation scores in all three grades (N=3,313). The regressions use standardized grade-specific motivation as the main independent variables, but are otherwise identical to the ones in Table 2. Regressions using average motivation across grades 1-3 for the same subsample of students are also shown for reference. For this figure only, high school GPA is re-scaled to range from 0-1.

Online Appendix Figure B.2: Randomization check like in Feld and Zoelitz (2017), distribution of p-values



Notes: The figure reports results from a test for random assignment of students to classes similar to the one conducted in Feld and Zölitz (2017). For this test, I ran separate regressions of the variables indicated above the six plots on class dummies for each school-by-entry-grade cell. After each regression, I conducted an F test for the joint significance of the class dummies and collected the p-value. Under random assignment, these p-values should be distributed roughly uniformly. The plots in this figure show the distributions of these p-values for each variable. The red vertical line indicates the p-value of 0.05.

Online Appendix Figure B.3: Results for a sample of students observed with most outcomes



Notes: The figure shows point estimates and 95 percent confidence intervals from regressions of the outcome variables indicated on the horizontal axis on peer motivation. The blue squares correspond to the main estimates shown in Table 5 and Table 6. The red circles show estimates based on a sample that is restricted to students observed with entry-grade reading and math scores, entry-grade motivation, and middle-school reading and math scores. This restricted sample includes 1,510 students. Because of remaining missing information, sample sizes for the regressions on HS GPA and HS graduation are 474 and 719, respectively. EG = entry-grade, G5 = grade 5, G8 = grade 8, HS = high school. For this figure only, high school GPA is re-scaled to range from 0-1.

Online Appendix Table B.1: Randomization check like in Chetty et al. (2011)

	Male	Black	Free lunch	Age	Old for grade	Pred. achievement
	(1)	(2)	(3)	(4)	(5)	(6)
p-value	.18	.95	.54	.18	.04	.66
Observations	2,861	2,766	2,730	2,845	2,845	2,868

Notes: The table reports results from a test for random assignment of students to classes similar to the one conducted in Chetty et al. (2011). The intuition of this test is that if students were indeed randomly assigned to classes, then class dummies should not predict their predetermined characteristics. For this table, I regressed each of the variables indicated in the column headers on school-by-entry-grade fixed effects and class dummies (leaving out one dummy per school-by-entry-grade cell to avoid collinearity). I then conducted an F test for the joint significance of all class dummies. The table reports the corresponding p-values.

Online Appendix Table B.2: Randomization check like in Feld and Zoelitz (2017), number of p-values below certain thresholds

	No. of tests	No. of p-values below			Share of p-values below		
		10%	5%	1%	10%	5%	1%
Male	145	16	9	3	11.03%	6.21%	2.07%
Black	63	4	2	1	6.35%	3.17%	1.59%
Free lunch	121	11	6	2	9.10%	4.96%	1.65%
Age	147	12	6	3	8.16%	4.08%	2.04%
Old for grade	146	16	10	2	10.95%	6.85%	1.37%
Pred. achievement	147	11	4	1	7.48%	2.72%	0.68%

Notes: The table reports results from a test for random assignment of students to classes similar to the one conducted in Feld and Zölitz (2017). For this test, I ran separate regressions of the variables indicated in rows on class dummies for each school-by-entry-grade cell. After each regression, I conducted an F test for the joint significance of the class dummies and collected the p-value. Under random assignment, the shares of p-values below certain confidence levels should be close to this this level (for example, about five percent of p-values should be below 0.05). The table shows the number of tests conducted for each variable and the number and share of p-values below the thresholds of 10%, 5% and 1%. The number of tests conducted is lower than the number of school-by-entry-grade cells, 147, for some variables due to missing data or due to collinearity (for example, if all students entering a certain school in a certain grade were black).

Online Appendix Table B.3: Peer motivation and entry-grade achievement, analysis of omitted variable bias

	Reading		Math	
	(1)	(2)	(3)	(4)
Peer motivation	0.081***	0.071***	0.036	0.027
	(0.023)	(0.024)	(0.032)	(0.032)
Peer achievement controls	No	Yes	No	Yes
Peer demographic controls	No	Yes	No	Yes
Observations	2,185	2,185	2,196	2,196
R2 (within)	0.122	0.134	0.275	0.286
$\delta(Rmax = 1.3 \times R^2)$		1.551		0.129

Notes: The table quantifies the amount of omitted variable bias that would be needed to drive the coefficient on peer motivation in the regressions in Table 5 to zero. The analysis is based on the method developed by Altonji, Elder, and Taber (2005) and refined by Oster (2019) and compares the coefficient estimates and R^2 values from baseline regressions (columns 1 and 3) to those from regressions which additionally control for averages of classmates' reading and math achievement in the previous school year, an indicator for whether the class includes a kindergarten repeater, and averages of classmates' socio-demographic characteristics (columns 2 and 4). For further details on controls included in the specifications, see Table 5. The last row in the table shows estimates of δ, which is the ratio of the impact of unobservables to the impact of the controls for peer achievement and socio-demographic characteristics that would drive the coefficient on peer motivation to zero. To compute δ, one needs to make an assumption about the hypothetical maximum R^2 achievable if all relevant controls were observed, the $Rmax$. Oster (2019) suggests setting $Rmax$ equal to 1.3 times the R^2 from the controlled regression, and the table uses this value. Calculations of δ are made using the Stata package `-psacalc-` and treat school-by-entry-grade fixed effects as nuisance parameters (that is, the R^2 is calculated within school-by-entry-grade cells). Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Online Appendix Table B.4: Motivation of male peers and female peers and entry-grade achievement

	Reading			Math		
	All students	By gender		All students	By gender	
		female	male		female	male
	(1)	(2)	(3)	(4)	(5)	(6)
Motivation of male peers	0.075**	0.020	0.119***	0.023	0.004	0.045
	(0.038)	(0.066)	(0.045)	(0.047)	(0.069)	(0.060)
Motivation of female peers	0.066	0.085	0.058	0.033	0.014	0.027
	(0.044)	(0.075)	(0.058)	(0.049)	(0.064)	(0.071)
Peer achievement controls	Yes	Yes	Yes	Yes	Yes	Yes
Peer demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,185	976	1,207	2,196	974	1,220

Notes: The table shows estimates of regressions in which peer motivation is measured separately for male and female peers. In columns 2 and 5 (3 and 6), the sample is restricted to female (male) entrants. All regressions control for own socio-demographic characteristics, averages of classmates' reading and math achievement in the previous school year, an indicator for whether the class includes a kindergarten repeater, averages of classmates' socio-demographic characteristics, a dummy for small class, and school-by-entry-grade fixed effects. Standard errors in parentheses are clustered by school-by-entry-grade. * p<0.10, ** p<0.05, *** p<0.01.

Online Appendix Table B.5: Balancing tests for peer motivation, with correction of bias due to missing information on classmates' motivation

	Male	Black	Free lunch	Age	Old for grade	Pred. achievement
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: peer motivation only						
Peer motivation	0.001	–0.009	–0.000	–0.043*	–0.005	0.018
	(0.019)	(0.009)	(0.015)	(0.026)	(0.018)	(0.028)
Panel B: joint regressions with peer achievement						
Peer motivation	–0.000	–0.008	0.000	–0.042	–0.005	0.016
	(0.019)	(0.009)	(0.015)	(0.026)	(0.018)	(0.029)
Peer reading achievement	0.000	0.002	0.009	–0.013	0.003	–0.006
	(0.021)	(0.010)	(0.029)	(0.028)	(0.019)	(0.038)
Peer math achievement	0.024	–0.013	–0.033	–0.009	–0.012	0.042
	(0.021)	(0.012)	(0.020)	(0.036)	(0.025)	(0.040)
p-value (joint significance)	0.45	0.51	0.28	0.32	0.95	0.63
Observations (both panels)	2,861	2,766	2,730	2,845	2,845	2,868

Notes: The table shows estimates of regressions of students' socio-demographic characteristics and predicted achievement on the characteristics of their classmates. Estimates follow the specifications in Table 4, with the difference that peer motivation is corrected for missing data using the method proposed by Sojourner (2013). Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Online Appendix Table B.6: Peer motivation and entry-grade achievement, results for subsamples with information on motivation for a high share of peers

	Reading (1)	Math (2)
Panel A: more than 50% of classmates observed with motivation scores		
Peer motivation	0.063*	0.023
	(0.032)	(0.034)
Peer achievement controls	Yes	Yes
Peer demographic controls	Yes	Yes
Observations	1,590	1,602
Panel B: more than 66% of classmates observed with motivation scores		
Peer motivation	0.062*	0.027
	(0.035)	(0.038)
Peer achievement controls	Yes	Yes
Peer demographic controls	Yes	Yes
Observations	1,094	1,104
Panel C: more than 75% of classmates observed with motivation scores		
Peer motivation	0.077	0.055
	(0.056)	(0.051)
Peer achievement controls	Yes	Yes
Peer demographic controls	Yes	Yes
Observations	643	647

Notes: The table shows estimates of the effect of peer motivation on achievement in reading and math. In Panel A/B/C, the sample is restricted to students for whom more than 50/66/75 percent of their classmates are observed with motivation scores from the previous school year. Regressions control for own socio-demographic characteristics, averages of classmates' reading and math achievement in the previous school year, an indicator for whether the class includes a kindergarten repeater, averages of classmates' socio-demographic characteristics, a dummy for small class, and school-by-entry-grade fixed effects. Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Online Appendix Table B.7: Instrumental variable estimates

	Entry grade				Grades 5-8		High school		College
	reading	math	motivation	self-concept	reading	math	GPA	grad.	ACT/SAT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Peer motivation	0.085	0.156	–0.090	0.067	–0.012	0.030	–1.004	0.012	–0.022
	(0.085)	(0.097)	(0.091)	(0.098)	(0.074)	(0.087)	(1.304)	(0.076)	(0.031)
Peer achievement controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F	16.24	16.75	16.68	16.68	15.08	15.05	7.27	11.04	20.29
Observations	2,184	2,195	2,274	2,274	2,117	2,118	650	1,012	2,868

Notes: The table shows instrumental variable estimates of the effect of peer motivation on the outcomes listed in the column headers. Peer self-concept is used as an instrument for peer motivation. Regressions control for own socio-demographic characteristics, averages of classmates' reading and math achievement in the previous school year, an indicator for whether the class includes a kindergarten repeater, averages of classmates' socio-demographic characteristics, a dummy for small class, and school-by-entry-grade fixed effects. Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Online Appendix Table B.8: Peer motivation and entry-grade achievement, controlling for peer self-concept

	Reading (1)	Math (2)
Peer motivation	0.069***	0.013
	(0.025)	(0.031)
Peer achievement controls	Yes	Yes
Peer demographic controls	Yes	Yes
Observations	2,185	2,196

Notes: The table shows estimates of the effect of peer motivation on achievement in reading and math. Regressions control for own socio-demographic characteristics, averages of classmates' socio-demographic characteristics and their math and reading scores in the previous school year, an indicator for whether the class includes a kindergarten repeater, a dummy for small class, and school-by-entry-grade fixed effects. Regressions also control for the average of classmates' self-concept score in the previous school year. Standard errors in parentheses are clustered by school-by-entry-grade. * p<0.10, ** p<0.05, *** p<0.01.

Online Appendix Table B.9: Effects of peer motivation on being observed with different outcomes

	Outcome is an indicator for being observed with								
	entry grade				grades 5-8		high school		college
	reading	math	motivation	self-concept	reading	math	GPA	grad.	ACT/SAT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Peer motivation	0.008	0.005	–0.003	–0.003	–0.018*	–0.018*	–0.012	–0.009	–
	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.010)	–
Peer achievement controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–
Peer demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–
Observations	2,868	2,868	2,868	2,868	2,868	2,868	2,868	2,868	–

Notes: The table shows estimates from regressions of dummies for being observed with the outcomes indicated in the column headers on peer motivation. Column 9 is empty because ACT/SAT test-taking is observed for all students. Regressions control for own socio-demographic characteristics, averages of classmates' reading and math achievement in the previous school year, an indicator for whether the class includes a kindergarten repeater, averages of classmates' socio-demographic characteristics, a dummy for small class, and school-by-entry-grade fixed effects. Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Online Appendix Table B.10: Peer motivation and entry-grade achievement in other subjects

	word study (1)	listening (2)
Peer motivation	0.081***	0.027
	(0.024)	(0.028)
Peer achievement controls	Yes	Yes
Peer demographic controls	Yes	Yes
Observations	2,507	2,187

Notes: The table shows estimates of the effect of peer motivation on achievement in word study skills and listening, which were assessed by the Stanford Achievement Test next to reading and math. Achievement scores are standardized to have mean 0 and SD 1 in each subject. Regressions control for own socio-demographic characteristics, averages of classmates' reading and math achievement in the previous school year, an indicator for whether the class includes a kindergarten repeater, averages of classmates' socio-demographic characteristics, a dummy for small class, and school-by-entry-grade fixed effects. Standard errors in parentheses are clustered by school-by-entry-grade. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Online Appendix Table B.11: Peer motivation and educational success, correction for multiple hypothesis testing

	Entry grade						Grades 5-8		High school		College
	reading	math	word st. skills	listening	motivation	self-concept	reading	math	GPA	gradua-tion	took ACT/SAT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Peer motivation	0.071	0.027	–0.004	0.000	0.081	0.027	–0.024	–0.026	–0.467	–0.030	–0.007
	[0.004]	[0.402]	[0.887]	[0.990]	[0.001]	[0.335]	[0.228]	[0.223]	[0.267]	[0.085]	[0.399]
	[0.032]	*[0.813]*	*[0.012]*	*[0.813]*	*[0.976]*	*[0.976]*	*[0.813]*	*[0.813]*	*[0.793]*	*[0.570]*	*[0.829]*
Peer ach. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer dem. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,185	2,196	2,276	2,276	2,507	2,187	2,118	2,119	665	1,018	2,868

Notes: The table shows estimates of the effect of peer motivation on the outcome variables indicated in the column headers along with two different sets of p-values. The p-values in brackets shown directly below the coefficient estimates are based on the main estimates in Tables 5 and 6 and Online Appendix Table B.10. The p-values in italics and brackets in the next row are corrected for multiple hypothesis testing using the procedure by Romano and Wolf (2005a,b). To implement this procedure, I use the Stata `rwolf` command described in Clarke, Romano, and Wolf (2020).