Teacher Effects on Student Well-Being and Mental Health

Ana Brás-Monteiro Marius Opstrup Morthorst*

Job Market Paper

30 October 2026

Click here for the latest version

Abstract

This paper studies how teachers affect student well-being and subsequent mental health in adolescence. Using linked administrative and survey data from Denmark, we construct a value-added measure of teachers' effectiveness in improving student well-being during middle school and examine its longer-run effects. We find that teachers have meaningful impacts on students' contemporaneous well-being. These effects persist several years later: students exposed to teachers with higher well-being value-added during middle school report higher well-being in high school. Going beyond survey-based measures, we find persistent reductions in mental health care utilization and depression diagnoses in adolescence. We further show that our measure of well-being value-added is only weakly correlated with test score value-added, and that both shape distinct long-run outcomes: well-being value-added primarily reduces mental health care utilization and depression diagnoses, while test score value-added has a stronger positive impact on subsequent academic outcomes but also reduces anxiety diagnoses.

*Brás-Monteiro (corresponding author): University of Zürich (ana.brasmonteiro@econ.uzh.ch), Morthorst: Aarhus University and the Frisch Centre (mmorthorst@econ.au.dk). Ana Brás-Monteiro is very grateful to Ulf Zölitz, Ana Costa-Ramón, and Jonah Rockoff for their guidance and support. We thank seminar participants at Columbia University and the University of Zurich, and participants at EALE 2026, IWAEE 2026, the CESifo Junior Workshop on the Economics of Education 2026, and the LEER Conference 2026 for helpful comments.

1 Introduction

Mental health problems among children and adolescents impose large public health and economic costs. They are among the leading causes of illness and disability during adolescence worldwide (WHO, 2025).[1] Poor mental health has consequences that extend well beyond childhood, predicting lower educational attainment, employment, and earnings later in life (Currie, 2024). Understanding its determinants is therefore crucial: which features of children's social environments shape their well-being and mental health?

Teachers may play a key role. Children spend more than 30 hours a week in school, and outside their own families, teachers are the adults with whom they interact most throughout childhood. While a large literature documents teachers' influence on student test scores (Koedel, Mihaly and Rockoff, 2015; Chetty, Friedman and Rockoff, 2014*a*), it is increasingly recognized that teachers shape multiple dimensions of students' development (Jackson, 2018; Petek and Pope, 2023). Yet little is known about whether and how teachers influence students' well-being and mental health.

This paper provides the first evidence that teachers shape students' well-being and mental health. We combine rich, population-wide education and health registers with repeated student well-being surveys in Denmark to construct a measure of teachers' effectiveness in improving student well-being — teachers' well-being value-added — and examine its consequences for students' mental health outcomes during adolescence. We also relate this measure to test-score-based measures of teacher quality to study whether the two dimensions capture distinct aspects of teacher effectiveness and affect distinct long-run outcomes. Understanding this is crucial for designing teacher hiring and evaluation policies that identify teachers who can improve not only academic achievement but also students' well-being and mental health.

We start our analysis by documenting a strong relationship between student well-being in middle school and a broad set of later outcomes. Students with higher well-being scores go on to perform better academically in high school and are substantially less likely to use mental health care services or receive psychiatric diagnoses during adolescence, even after conditioning on test scores.

Motivated by this relationship, we construct a value-added measure of teachers' effectiveness in improving student well-being during middle school. Following Chetty, Friedman and Rockoff (2014*a*), we predict each teacher's effect in a given year from the well-being of the students she taught in all other years, controlling for lagged well-being, lagged test scores,

[1] Globally, one in seven children experiences a mental health condition. In Denmark, the setting of this paper, roughly 15 percent of children have been diagnosed with a mental illness by age 18 (Dalsgaard et al., 2020).

and student, classroom, and school characteristics, and we allow teacher quality to drift over time. We find substantial and meaningful variation in teachers' impacts on student well-being, comparable in magnitude to teacher effects on other non–test score outcomes documented in the literature. A one–standard deviation increase in teacher well-being value-added raises student well-being by 0.11 standard deviations.

Observable teacher characteristics, however, explain no more than 1 percent of the variation in well-being value-added. These results are broadly consistent with the test score value-added literature (Rockoff, 2004; Bau and Das, 2020). Nonetheless, we find that female and more experienced teachers are, on average, more effective at raising student well-being. Well-being value-added is 0.04 standard deviations higher for female than for male teachers. It also increases monotonically with teaching experience: teachers with more than 20 years of experience have a well-being value-added that is 0.09 standard deviations higher than those with five years or less.

We then examine how exposure to teachers with higher value-added affects students' outcomes several years later, including self-reported well-being in high school, mental health care utilization and psychiatric diagnoses between ages 16 and 18, and academic performance in high school. Importantly, while we use students' self-reports to construct teacher well-being VA, our mental health outcomes are based on administrative health records: visits to psychologists and psychiatrists, and psychiatric diagnoses. We find that students exposed to teachers with higher well-being value-added in middle school report significantly higher levels of well-being during high school. Moreover, they are substantially less likely to use mental health care services between ages 16 and 18, and they experience reductions in depression diagnoses. They also experience small improvements in academic outcomes in high school. Taken together, our results highlight that teachers' impacts on well-being extend beyond subjective survey responses to longer-run, objective mental health outcomes.

Our key identifying assumption is that students' assignment to teachers is unrelated to unobserved determinants of their outcomes, conditional on a rich set of lagged outcomes and student, classroom, and school characteristics. To assess the credibility of this assumption, we implement a series of empirical tests to detect signs of sorting of students to teachers. Following Chetty, Friedman and Rockoff (2014*a*,*b*), we examine forecast bias, test for selection on observable characteristics excluded from the value-added model, including parental and students' own prior psychiatric diagnoses, and exploit quasi-experimental variation arising from changes in teaching staff across cohorts. Across these exercises, we find little evidence of systematic sorting that would materially bias either the contemporaneous or long-run

estimates of teachers' impacts on student well-being and mental health.

To shed light on the mechanisms, we examine how well-being value-added affects students' perceptions of teacher practices, as well as their own agency and peer relations, during the year of exposure. We find that students taught by teachers with higher well-being value-added report better classroom management, instruction, and support, suggesting that students themselves recognize teachers who improve their well-being. They also report greater agency (e.g., being able to move on when a task is too difficult) and more positive peer relations, in particular working well with others. These findings suggest that teachers who improve well-being shape both the classroom environment and behaviors students may carry beyond it.

We next ask whether the effects of well-being value-added are evenly distributed across students, or whether certain groups benefit disproportionately. We find that effects on self-reported high school well-being are fairly homogeneous across subgroups. Effects on mental health outcomes, however, are concentrated among students with higher baseline achievement and well-being. These patterns suggest that while teachers improve the subjective well-being of all students, the downstream protection against mental health problems is concentrated among those who were already doing relatively well.

Beyond heterogeneity across students, we also find meaningful variation in effects across the grade of exposure to the teacher. Effects tend to be largest in the early grades of each school level, particularly grades 4 and 7, when students enter a new stage of schooling associated with changes in curriculum and instruction. This suggests that exposure to high well-being VA teachers may be especially consequential during educational transitions, when students might be facing new challenges in the school environment.

We then examine how teachers' well-being value-added relates to their test score value-added. We show that well-being value-added captures a distinct dimension of teacher quality: within teachers, the correlation between well-being and test score value-added is approximately 0.25. Consistent with the multidimensional view of teacher quality, we find that test score value-added has much stronger effects on long-run academic outcomes, whereas well-being value-added has the strongest effects on students' long-run well-being. In particular, when disaggregating high school well-being into its sub-dimensions, we find that being exposed to a high well-being VA teacher during middle school improves all dimensions of school life in high school, while test score VA improves academic well-being but has no positive effect on social or emotional dimensions.

The two dimensions of teacher quality also shape different mental health outcomes. Well-

being value-added reduces mental health care utilization, whereas test score value-added has little effect on it. When it comes to specific diagnoses, well-being value-added reduces depression while test score value-added reduces anxiety. Most of these patterns persist when both dimensions are included jointly in a regression, indicating that neither measure simply proxies for the other. The underlying mechanisms also differ: compared with well-being value-added, test score value-added has much smaller effects on teacher practices, student agency, and peer relations. These patterns suggest that the two dimensions of teacher quality shape students' long-run outcomes through distinct channels.

Finally, we examine how incorporating student well-being into teacher evaluations changes the predicted benefits of teacher selection policies. Following Hanushek (2011), we simulate replacing the bottom 5 percent of teachers with teachers of average quality. We identify the bottom 5 percent by ranking teachers on a weighted average of their test score and well-being value-added, and vary the weight placed on each measure. Ranking teachers solely on test score value-added produces large academic gains but much smaller improvements in student well-being, and no reduction in mental health care utilization. Giving well-being value-added a 25 percent weight nearly doubles the predicted well-being gain, from 0.032 to 0.060 standard deviations, while reducing the GPA gain only slightly, from 0.070 to 0.069 standard deviations. It also reduces mental health care utilization by 0.4 percentage points (3 percent of the mean). These results highlight that incorporating well-being into teacher evaluations can substantially improve non-academic outcomes at little academic cost. More generally, any ranking rule involves a trade-off between academic and non-academic outcomes, and choosing one ultimately depends on how a policymaker values mental health outcomes relative to academic ones.

Our paper contributes to two different strands of literature. First, we contribute to the growing literature on child and adolescent mental health by focusing on the role of teachers in shaping student well-being and, consequently, their mental health. A large body of work shows that mental health is shaped by early-life conditions, including family background and early shocks, as well as post-natal environments such as income shocks and parental job loss (Bütikofer et al., 2024; Adhvaryu, Fenske and Nyshadham, 2019; Persson and Rossin-Slater, 2018; Baird, De Hoop and Özler, 2013; Schaller and Zerpa, 2019; Currie et al., 2026). More recent studies emphasize other aspects of schools, including the role of school starting age, instructional intensity, peers and school violence (Dee and Sievertsen, 2018; Marcus et al., 2020; Rossin-Slater et al., 2020; Chen et al., 2026; Bütikofer et al., 2023). Relatively little is known about the role of individual teachers in shaping student well-being and mental health. Our

paper addresses this gap by documenting that teachers can have meaningful positive impacts on student well-being that lead to persistent improvements in mental health outcomes.

Our paper also contributes to the literature on teacher value-added and the multidimensional nature of teacher quality. Chetty, Friedman and Rockoff (2014*a*,*b*) document substantial heterogeneity in teachers' impacts on test scores and show that exposure to high–value-added teachers has lasting effects on educational attainment and earnings. A growing body of work further demonstrates that teacher quality extends beyond test scores, with teachers affecting absences, suspensions, grades, socio-emotional competencies, and even criminal justice contact (Jackson, 2018; Petek and Pope, 2023; Rose, Schellenberg and Shem-Tov, 2025; Backes et al., 2024; Lavy and Megalokonomou, 2024; Liu and Loeb, 2021). While this literature makes a compelling case that teachers shape non-cognitive skills, these skills are typically inferred from observed behaviors and school records rather than measured directly. We extend this literature by bringing direct measures of student well-being and population-wide mental health outcomes into the analysis of teacher effectiveness. Using the value-added framework, we show that teachers meaningfully influence student well-being, and that this dimension of teacher quality is largely distinct from test score value-added. We also document that this new measure of teacher quality has lasting effects on students' mental health outcomes. Our findings highlight the non-academic channels through which teachers shape students' long-run trajectories, and suggest that incorporating well-being into teacher selection policies could yield substantial mental health gains at little academic cost.

Closest to our work, a handful of recent studies use survey-based or standardized assessments to estimate teachers' effects on students' socio-emotional skills and personality traits (Kraft, 2019; Fleche, 2026; Montonen, 2025). In contrast, we focus on student well-being and, to our knowledge, are the first to link teacher value-added to population-wide administrative measures of mental health care utilization and psychiatric diagnoses.

The remainder of the paper is organized as follows. Section 2 describes the Danish education and health care systems and the data, and documents the relationship between student well-being and later outcomes. Section 3 outlines the empirical method for estimating teacher value-added and its effects on long-run student outcomes. Section 4 presents the main results on teachers' well-being value-added and its effects on students' mental health and academic outcomes. Section 5 tests the identifying assumptions. Section 6 examines the mechanisms behind the effects of well-being value-added, as well as heterogeneity across students and by the grade of exposure to the teacher. Section 7 compares well-being and test score value-added, their long-run effects, and the mechanisms through which they operate.

It also presents policy simulations that rank teachers on both measures. Section 8 concludes.

2 Setting and Data

This section describes the institutional setting and the data used in the empirical analysis. We first describe the Danish education and health care systems, which determine how students are assigned to teachers and how mental health outcomes are recorded in the administrative data. We then describe the data sources and measures. A key component is the national well-being survey, which schools administer annually to all students and which forms the basis of our measure of teacher quality. We link the survey to administrative registers that follow the same students through high school and record their mental health care utilization and psychiatric diagnoses. Finally, we describe the analysis sample and document that student well-being in middle school predicts later outcomes, even conditional on test scores.

2.1 Institutional Background

The Education System. Denmark offers a high-quality, publicly funded educational system. Compulsory schooling (*Grundskole*) spans ten years, beginning in the calendar year when children turn six, and is organized as an integrated school covering both primary and lower secondary education. It consists of one pre-school year (grade 0), nine years of primary and lower secondary education (grades 1–9), and an optional tenth grade.[2] While compulsory education is primarily publicly funded and free of charge (as is almost all post-compulsory education, including higher education institutions), approximately 18 percent of students attend private schools, where they pay tuition fees.

Upon entering grade 0, children are grouped into classes and typically remain in the same classroom, with an average class size of 21 pupils, until grade 9.[3] While schools have some discretion over class formation, national guidelines set out in the *Folkeskole Act* state that, as the Folkeskole is not an examination-oriented school, the main rule is that pupils should attend classes with peers of the same age rather than being grouped by subject-specific proficiency.

Classes are taught by multiple subject-specific teachers. However, each class also has a designated class teacher with primary responsibility for monitoring students' academic and social development, as well as ensuring coordination with other teachers and communication with families. Usually, the Danish teacher is the class teacher.

[2]Throughout this paper, we refer to grades 4–9 as "middle school", a term broadly equivalent to the upper-primary and lower-secondary years in other countries, including the United States.

[3]However, some schools may reshuffle children in grade 3 and/or grade 7 to try new group dynamics.

The schools administer several standardized tests and examinations during compulsory schooling. The national test program, introduced in 2010, includes ten standardized tests administered in grades 2 through 8, covering reading, Mathematics, and other subjects (Beuchert and Nandrup, 2018). Compulsory education then concludes with five mandatory exams in grade 9. The results of these exams contribute to the student's final GPA. Additionally, students receive year marks from their subject-specific teacher for their overall performance throughout the school year, which they use for applying to further education. Schools also administer yearly well-being surveys; further details are provided in the following subsection.

After grade 9, students choose between several post-compulsory paths. Approximately 36 percent enroll directly in academic high school, around 51 percent first complete an optional additional year of schooling known as 10th grade, and 6 percent enter vocational education and training (VET) programs directly.[4] The large majority of students who complete 10th grade subsequently enroll in academic high school. Academic high school programs prepare students primarily for higher education, whereas VET programs equip students with skills for a specific trade or industry, with limited opportunities for direct enrollment into higher education. Students may freely enroll in academic high school if their GPA is at least 5, or in the two-year preparatory program if their GPA is at least 4, on the Danish 7-point grading scale ranging from -3 to 12.[5] Within academic high school, students choose among several tracks that differ in curricular focus, including a general program (STX), a business-oriented program (HHX), and a technical program (HTX).

The Health Care System. Denmark operates a universal, publicly funded health care system in which general practitioners (GPs) serve as the primary point of contact and act as gatekeepers to specialist care. Health care is provided free of charge at the point of use, with specialist visits fully covered under the national health insurance scheme provided the patient holds a GP referral.

For children and adolescents with mental health concerns, the system follows a stepped model. GPs provide initial assessment and can prescribe psychotherapy, but guidelines issued by the Danish Health and Medicines Authority recommend that children at risk of a mental health disorder be referred to specialist care: mild cases to private practice child psychiatrists, and more severe cases to hospital-based psychiatric departments. A key feature of the system is that prescription drugs for mental health conditions for individuals below age 18 must

[4]Based on the 2018 cohort (Statistics Denmark, 2024).

[5]As of 2018, students must also pass all mandatory exams. Below these thresholds, students may still enroll in VET programs (which require a minimum grade of 02 in Danish and Mathematics) or apply to academic programs through individual assessment by the receiving school.

be initiated by a specialist in psychiatry rather than a GP. As a result, all children receiving pharmacological treatment for conditions such as depression or ADHD will have had formal contact with a specialist.

2.2 Data Sources and Measures

Our analysis draws on several Danish administrative registers that can be linked using unique individual identifiers. We begin with the Education Register, which contains enrollment records for all students attending public schools in Denmark. Our sample includes students enrolled in grades 4 through 9 during the 2015/2016–2019/2020 academic years. We link these data to the Teacher Register to assign students to their subject-specific teachers. The analysis uses teacher-student links in Danish and Mathematics, as these are the two subjects for which we have frequent information on test scores. We further link our sample to the Central Population Register to obtain demographic characteristics and parental identifiers. Below, we describe the key variables used in the analysis and the additional data sources from which they are drawn. We distinguish between measures observed in middle school, which we use to estimate teacher value-added and to study mechanisms, and long-run outcomes observed after students leave compulsory schooling, which we use to assess teachers' longer-term impacts.

2.2.1 Middle School Measures

Well-being. To measure well-being, we use the responses to the nationwide well-being survey, which has been conducted annually since 2015 for all students in compulsory education between January and April of every year. The survey is administered electronically, with a designated teacher in each classroom overseeing the process. The teacher further emphasizes the importance of honest responses and assures students that their individual answers will remain confidential and will not be shared with parents, teachers, or other school personnel. Even though it is mandatory for schools to administer the survey, it is not mandatory for students to complete it. Regardless, the response rate is very high, roughly 85 percent, and highest for earlier grades.

The questionnaire contains 40 questions addressing different aspects of well-being and school life in general. Among the full list of questions, we use the 18 questions that enter into the academic and social well-being subscales, as defined by Danish Ministry of Education (2016). The translated version of the questions can be found in Appendix B.1. The answers are coded to range from one to five, with higher values indicating more positive well-being.

We construct the final well-being measure by computing the within-individual average of the answers provided and standardizing it by cohort. The distribution of this score, before standardization, is depicted in Appendix Figure A.1.

School experience indices. We use other questions from the well-being survey to construct three indices of students' experience of school, which we use to study mechanisms.[6] We group the questions according to whose behavior they describe: the teacher's, the student's own, or the student's interactions with peers. This yields three indices, which we refer to as teacher quality, student agency, and peer relations.

The *teacher quality* index covers classroom management, instruction, and support. One caveat is that the survey asks students about their teachers in general rather than a specific teacher. Since students spend a large fraction of their school day with their Danish and Mathematics teachers, we use the teacher quality index as a proxy for students' perceptions of the teachers in our analysis.

The *student agency* index captures whether students take action on their own when they are bored, stuck, or treated unfairly. The *peer relations* index captures how students relate to their classmates through empathy, cooperation, and refraining from bullying.[7] Each index is the average of the corresponding answers, standardized by grade and year. The questions are listed in Appendix B.3.

Test scores. For grades 2 through 8, we use the results of the national tests in Danish and Mathematics. Specifically, we use reading test scores in grades 2, 4, 6, and 8, and Mathematics test scores in grades 3, 6, and 8 (for 2018 only). All test scores are standardized by subject and cohort. For grade 9, we use results from the exit exams, which we obtain from administrative records. We focus on final scores in Danish and Mathematics, which we standardize by cohort.

2.2.2 Long-Run Outcomes

Education outcomes. We study two post-compulsory education outcomes: enrollment in academic high school and final high school GPA. For the former, we define the variable to take value 1 if the student has ever enrolled in an academic high school program. For the latter, we use administrative records on high school graduation, restricted to academic high school graduates, which we standardize by graduation year and academic track.[8]

[6] None of these questions enters our well-being measure.

[7] Questions on how students are treated by their peers, such as being accepted or bullied, enter the well-being measure and are therefore excluded.

[8] Academic high school programs differ in curriculum and do not all include the same coursework, making raw GPA scores not directly comparable across tracks.

High school well-being. To measure well-being in high school, we draw on data from the Danish Well-Being Survey administered in academic high school programs. Although the questionnaire differs from that used during compulsory schooling, Danish Ministry of Education (2023) constructs the same two domains. *Academic well-being* includes questions on engagement, motivation, and academic self-efficacy, while *social well-being* reflects students' sense of belonging and their relationships with classmates. We construct the well-being measure analogously and focus on the first year of high school. Participation in this survey is voluntary, resulting in a smaller sample. Note that, unlike the mental health outcomes, which cover all students, this measure is by construction limited to the subsample who enrolled in academic high school.

As supplementary analysis, we also examine the three additional sub-indices constructed by the Ministry of Education. *Learning environment* reflects students' views of teaching and teacher support. *Pressure and concerns* includes questions on how often students feel pressured by school, homework, grades, and their own expectations. *Bullying* includes questions on exposure to teasing, threats, unwanted attention, and digital bullying. All sub-indices are coded so that higher values indicate more positive experiences: higher values of *pressure and concerns* and *bullying* correspond to less pressure and less bullying.[9]

Mental health outcomes. The mental health outcomes capture mental health care utilization and psychiatric diagnoses, which we measure from two separate sources. We measure mental health care utilization using the *Health Insurance Register*, which records reimbursements to private clinics for services covered by national health insurance. We define utilization as any contact with a private psychologist, a private practicing psychiatrist or child and adolescent psychiatrist, or a GP providing mental health counseling.

We measure psychiatric diagnoses using hospital records, combining the *Psychiatric Central Research Register* with the *National Patient Register*. Together, these cover all inpatient, outpatient and emergency contacts with public and private hospitals. We define an indicator for any psychiatric diagnosis using ICD-10 codes starting with "F", and further define indicators for depression (F32–F33) and anxiety (F40–F48).[10] We measure both outcomes between ages 16 and 18, restricting the sample to students observed at all ages in this window.

[9]See Appendix B.4 for the translated version of the questions entering each sub-index.

[10]See Appendix B.2 for details on the exact codes used to construct the mental health outcomes.

2.3 Analysis Sample

We treat the student-subject-year as the relevant unit of observation. Therefore, each observation includes a student's subject-specific test score for a given year, their lagged test score, their contemporaneous well-being measure, lagged well-being, and the long-run outcomes.

To ensure sufficient information to estimate teacher effects, we drop teacher–subject–year cells with fewer than ten students, corresponding approximately to the bottom one percent of the student count distribution. We further impose restrictions based on the availability of well-being data, as this is the basis of our teacher quality measure. Specifically, we drop classrooms in which fewer than 20 percent of students have non-missing values for both current and lagged well-being. Finally, we restrict the analysis to student–subject–year observations with complete information on current and lagged well-being.

Table 1 reports summary statistics for the full sample and the analysis sample. The two samples have a similar demographic composition, with the analysis sample, if anything, slightly positively selected. Since our sample restrictions depend on well-being survey response, we examine selection into responding to the survey. Among students in grades 4 through 9, 74 percent respond to the survey. Within school-grade-year cells, demographic characteristics are only weakly related to response: girls and students whose mother or father holds a university degree are about 1 percentage point more likely to respond, and students with a migration background 0.5 percentage points less likely. Together, these characteristics explain 0.1 percent of the variation in response within cells (Appendix Table A.1).

2.4 Student Well-Being and Later Outcomes

We begin by documenting the relationship between student well-being in middle school and a range of later outcomes. Table 2 reports regressions of high school academic outcomes, adolescent well-being, mental health care utilization, and psychiatric diagnoses during ages 16 to 18, on standardized measures of student well-being measured in middle school, conditional on the same set of controls we use to compute the value-added measure. We find that higher student well-being in middle school is strongly associated with better subsequent well-being in high school and substantially lower mental health care utilization and psychiatric diagnoses during adolescence, even when controlling for test scores. Student well-being is also predictive of later academic outcomes, though test scores are more strongly related to academic performance in high school.

While these relationships are descriptive and may reflect selection on unobservables, they highlight student well-being as a distinct and consequential dimension of student develop-

ment that is not captured by test scores alone. We view these patterns as a validation of the well-being survey measure, which is particularly important given that it forms the basis of the teacher value-added measure we construct next.

3 Empirical Strategy

3.1 Estimating Teacher Value-Added

We estimate teacher value-added (VA) following the framework developed by Chetty, Friedman and Rockoff (2014*a*), adapting it to measure teachers' impacts on student well-being. The goal is to isolate teachers' persistent contributions to student outcomes by comparing realized outcomes to those predicted from prior outcomes, background characteristics, and classroom context.

Let Y_{ijt} denote the outcome of interest—either test scores or well-being—for student i taught by teacher j in year t. We construct value-added measures by first residualizing the outcome of interest with respect to a rich set of controls X_{ijt} and teacher fixed effects. We flexibly control for prior well-being and test scores using lags of a cubic polynomial of the student's well-being, Danish and Mathematics test scores, and we interact these cubics with the student's grade level.[11] We also control for students' ethnicity, gender, age and maternal education. We also include the following class- and school-level controls: (i) cubics in class and school-grade means of prior-year well-being and test scores, each interacted with grade; (ii) class and school-grade means of all the other individual covariates; (iii) class size; and (iv) grade-by-year fixed effects.

Let v_{ijt} be the residualized outcome. We estimate $\hat{\beta}$ via OLS including teacher fixed effects α_j to prevent teacher quality from being absorbed into the coefficients on student characteristics:

$$Y_{ijt} = \alpha_j + \beta X_{ijt} + \varepsilon_{ijt} \tag{1}$$

Following Chetty, Friedman and Rockoff (2014*a*), we do not include school fixed effects.

[11] We exclude observations with missing current or prior well-being in all specifications, as well-being is the primary outcome of interest. In the estimation of well-being value-added, we allow prior test scores in either subject to be missing; when missing, prior test scores are set to zero and flagged with missing indicators interacted with controls for prior well-being. In the estimation of test score value-added, we exclude observations with missing current or prior test scores in the subject for which value-added is being estimated. When prior test scores in the other subject are missing, we set them to zero and include an indicator for missing data interacted with the controls for prior own-subject test scores.

Since β is estimated within teacher, school fixed effects would be identified only from teachers who switch schools, and would force mean teacher quality to be equal across schools. Our value-added measure therefore captures differences in teacher quality both within and across schools. One concern is that it partly reflects school-level factors, such as school climate, rather than the teacher. The school-grade means in X_{ijt} account for differences in student composition, and the quasi-experimental design in Section 5, which exploits staffing changes within school-grade cells, holds such factors fixed.

We then form the residual by removing only the estimated effect of observables:

$$\nu_{ijt} = Y_{ijt} - \hat{\beta} X_{ijt} = \mu_{jt} + \epsilon_{it} \tag{2}$$

where μ_{jt} represents the teacher's contribution to student outcomes in year t and ϵ_{it} captures idiosyncratic student-year shocks. For each teacher-year, we compute the mean residual across students:

$$\bar{\nu}_{jt} = \frac{1}{N_{jt}} \sum_{i=1}^{N_{jt}} \nu_{ijt} \tag{3}$$

which provides a noisy estimate of the teacher's effectiveness in year t. Because these estimates are subject to sampling error and transitory shocks, we follow Chetty, Friedman and Rockoff (2014*a*) and construct a noise-adjusted measure of persistent teacher quality using a leave-year-out empirical Bayes procedure. Specifically, we predict a teacher's value-added in year t using information from the teacher's residuals in other years:

$$\widehat{\mu}_{jt} = \sum_{\substack{s=t-a \\ s \neq t}}^{t+a} \psi_s \bar{\nu}_{js} \tag{4}$$

where we allow the weights ψ_s to vary by the number of years before and after year t, and a is the maximum number of years available in the data for teacher j. Intuitively, a teacher's performance one year ago might be more predictive than their performance four years ago, because teacher quality can evolve over time, so ψ_{t-1} and ψ_{t+1} might be larger than ψ_{t-4}. By estimating separate weights for each relative year position, the model learns how much each neighboring year's data should contribute. We estimate these weights by minimizing the mean squared error of the forecasts:

$$\boldsymbol{\psi} = \underset{\{\psi_{t-a},\ldots,\psi_{t+a}\}}{\arg\min} \sum_{j} \left(\bar{\nu}_{jt} - \sum_{\substack{s=t-a \\ s \neq t}}^{t+a} \psi_s \bar{\nu}_{js} \right)^2 \tag{5}$$

This procedure produces leave-year-out jackknife value-added predictions that allow teacher quality to change over time, and it shrinks the value-added predictions to the mean through Bayesian shrinkage.

We implement this methodology separately for test scores and for well-being, yielding two distinct value-added measures. Throughout the paper, all value-added estimates are constructed using leave-year-out predictions, ensuring that contemporaneous outcomes are never mechanically used to construct the teacher quality measures applied to those outcomes. In Section 5, we discuss the identifying assumptions underlying this approach and present a series of tests for forecast bias and sorting, as well as a quasi-experimental analysis exploiting teachers who switch grades and schools.

3.2 Estimating the Long-Term Effects of Teacher Value-Added

Having constructed measures of teacher quality, we examine whether exposure to higher- or lower-quality teachers affects students' medium- and long-run outcomes. We estimate the following specification:

$$y_i = \sum_{k=1}^{K} \delta_k \widehat{\mu}_{jkt} + \beta X_{ijt} + \eta_{ijt} \tag{6}$$

where y_i denotes the outcome of interest (e.g., high school GPA or an indicator for mental health care utilization between ages 16 and 18), and $\widehat{\mu}_{jkt}$ represents teacher value-added in dimension k (e.g., test scores or well-being). We normalize each value-added measure by the standard deviation of the corresponding teacher effects, so that δ_k can be interpreted as the effect of a one standard deviation increase in teacher value-added. We estimate models that include each dimension separately as well as specifications that include both jointly. The coefficients δ_k therefore capture how different dimensions of teacher quality predict students' longer-run outcomes. We cluster standard errors at both the student and the classroom level. This accounts for common shocks among students in the same class, as well as for the fact that each student contributes multiple observations to our data, one for each subject and grade.

In order to be able to interpret these estimates as causal, one would need to make sure that unobserved determinants of students' long-term outcomes are uncorrelated with teacher quality, conditional on observables. Given the finding that conditioning on prior scores and other observables is adequate to obtain unbiased estimates of teachers' causal impacts on test scores, one may expect that comparing the long-term outcomes of students assigned to different teachers conditional on the same control vector will yield unbiased estimates of

teachers' long-term impacts. However, students may be sorted to teachers on unobserved characteristics that predict long-run outcomes but are not captured by prior outcomes. To address this concern, we perform a quasi-experimental analysis that uses teachers switching grades and schools in Section 5 (Chetty, Friedman and Rockoff, 2014*b*).

4 Main Results

4.1 Teacher Value-Added Estimates

We first characterize the magnitude of teacher value-added on student well-being. Table 3 reports the square roots of the estimated variance components, estimated separately for Danish and Mathematics teachers. Row 4 reports the standard deviation of teacher effects, computed from the within-year between-class covariance of mean residuals.

We estimate substantial heterogeneity in teacher well-being effects. The SD of teacher effects on well-being is 0.120 for Danish teachers and 0.101 for Mathematics teachers, with a pooled SD of 0.109.[12] This means, for example, that having a Danish teacher with well-being value-added at the 85th versus 50th percentile would increase student well-being by roughly 0.120 standard deviations. These magnitudes are comparable to benchmark estimates of teacher effects on test scores in prior work (e.g., Chetty, Friedman and Rockoff, 2014*a*; Jackson, 2018; Rose, Schellenberg and Shem-Tov, 2025), suggesting that teachers meaningfully influence students' well-being.

The within-year between-class covariance is identified only for teachers who teach multiple classes in the same year. In smaller schools where a teacher has a single class per year, this moment is unavailable and that teacher does not contribute to the estimation of the within-year between-class covariance. For comparison, Row 5 of Table 3 reports the square root of the lag-1 cross-year covariance, which is available for all teachers with at least two years of data.

The empirical distribution of our jackknife well-being VA estimates is plotted in Figure 1. The standard deviation of these estimates is 0.067, smaller than our estimate of the standard deviation of teacher effects in Table 3. This is expected: the best linear predictor shrinks noisy estimates toward the mean to minimize mean-squared error, so the dispersion of the shrunken estimates understates the dispersion of teacher effects. For reference, Table 3 and Figure 1 also report the corresponding estimates for test score value-added, which we discuss

[12]Estimating the variance of teacher effects with the unbiased estimator of Kline, Saggio and Sølvsten (2020) and Rose, Schellenberg and Shem-Tov (2025) yields SDs in a similar range (Appendix Table A.3).

in more detail in Section 7.

4.2 Teacher Characteristics

Using our well-being VA estimates, we estimate the association between value-added and observed teacher characteristics. Specifically, we relate well-being VA to two sets of characteristics: professional factors, including further education beyond a teaching degree and years of experience, and personal factors, including gender, having children, coming from a teacher family, and having any psychiatric diagnosis. Figure 2 reports the results. Observable characteristics account for less than 1 percent of the variation in well-being VA ($R^2 = 0.005$), consistent with prior evidence that effective teachers are difficult to identify from observable characteristics (Rockoff, 2004).

Among the characteristics we examine, years of experience is the strongest predictor: relative to teachers with fewer than six years of experience, teachers with more than 20 years of experience have well-being VA that is approximately 0.09 standard deviations higher, with a monotonically increasing pattern across experience categories. Female teachers are also associated with modestly higher well-being VA, while further education and having a teacher family background show no significant association. Teachers with a psychiatric diagnosis have somewhat lower well-being VA, though the estimate is imprecise.

4.3 Long-Run Effects of Well-Being Value-Added

4.3.1 Well-Being and Mental Health Outcomes

This section tests whether teachers who improve well-being cause improved well-being and mental health outcomes for students in later years, after exposure to the teacher. We estimate Equation 6 with well-being VA, in which the main outcomes are well-being during high school, whether the student utilized any mental health care services between ages 16 and 18, and whether the student had any psychiatric diagnosis between ages 16 and 18. We present the results using binscatters (see Figure 3) of the relationship between teachers' well-being value-added and each of the outcomes, conditional on the set of controls used to compute value-added measures. The plotted points show the relationship between the mean residualized outcome and the mean residualized value-added variable (with the unconditional mean of the outcome and value-added variable added back in) for 20 equally sized bins of teacher value-added measures. The coefficients and standard errors reported in the figures are from an OLS (ordinary least squares) regression, using the micro data, of the outcome variable on

the value-added variable, conditional on the same set of controls.

Figure 3a shows that students with higher well-being VA teachers in middle school report higher levels of well-being during high school. Specifically, a one standard deviation improvement in teacher well-being VA in a single grade increases high school well-being by 0.044 standard deviations. This compares with an effect of 0.109 standard deviations on contemporaneous well-being (Table 3), indicating that part of the initial effect fades out but a substantial share persists into high school. We examine which components of high school well-being drive this effect in Section 7.

We next examine whether the improvements in self-reported well-being translate into effects on more objective mental health outcomes, including mental health care utilization and psychiatric diagnoses. Figure 3b shows that assignment to a teacher with higher well-being value-added in a single grade significantly reduces the probability that a student visits a psychologist or psychiatrist between ages 16 and 18. The estimated effect corresponds to a 5.4 percent reduction relative to the mean rate of utilization, indicating that teachers who improve students' well-being also reduce subsequent demand for mental health care. We also find reductions in the likelihood of receiving a psychiatric diagnosis over the same age range, though these effects are smaller and less precisely estimated (Figure 3c).

The presence of consistent effects across both utilization and diagnosis supports the idea that teachers' impacts on well-being extend beyond subjective reports to clinically relevant mental health outcomes. Figure 4 presents estimates separately for different diagnoses. When we disaggregate psychiatric diagnoses by type, the effects are concentrated in depression, with no detectable effects on anxiety.

A natural concern in interpreting mental health care utilization is that it can capture different underlying mechanisms. Higher utilization may reflect worsening mental health, as students seek care in response to distress, but it may also indicate improved access to care or a greater willingness to seek help. While we cannot fully distinguish between these interpretations, the reductions in utilization we observe occur alongside sustained increases in self-reported well-being. Taken together, this pattern is more consistent with improvements in underlying mental health than with reduced help-seeking.

4.3.2 Academic Outcomes

While the previous results show that teachers' well-being value-added has meaningful effects on students' mental health and well-being, it is important to assess whether these effects extend to outcomes such as high school enrollment in academic tracks and grade point

average, which are central to educational policy and long-run economic outcomes.

Figure 5 shows that teachers' well-being value-added also has positive, though more modest, effects on academic outcomes. Assignment to a teacher with one standard deviation higher well-being value-added increases the probability of enrolling in an academic high school track by 0.007, relative to a mean enrollment rate of 0.73. We also find a positive effect on high school GPA: a one–standard deviation increase in well-being value-added raises students' standardized GPA by 0.007 standard deviations. This effect could be seen as a lower bound on the true effect. If we believe students induced into academic high school by being exposed to higher well-being VA teachers are marginal enrollees, they face a more demanding curriculum than they would have otherwise. This could mechanically bias the GPA estimate downward.

Although modest in size, these effects are informative about what well-being value-added captures. A natural concern is that teachers make students feel better by being less demanding, in which case gains in well-being would come at the expense of learning. We find no such trade-off. The same teachers who raise students' well-being and reduce their later use of mental health care also modestly improve their academic attainment. We contrast these effects with those of teachers' test score value-added in Section 7, where we show that test score value-added has substantially larger impacts on academic outcomes but a distinct pattern of effects on mental health.

5 Testing the Identifying Assumptions

5.1 Bias in the VA Estimates

We assess the potential for bias in the teacher value-added estimates using three complementary approaches, following Chetty, Friedman and Rockoff (2014*a*): tests of forecast unbiasedness, analyses of selection on observable characteristics, and a quasi-experimental design exploiting changes in teaching staff across cohorts.

5.1.1 Forecast Bias

First, we test whether the well-being value-added measure is forecast unbiased. Under the stationarity assumption, regressions of contemporaneous residualized outcomes on leave-year-out value-added should yield a coefficient of one by construction. Panel A of Figure 6 shows that this is the case: the estimated relationship is close to one. However, this relationship could be driven by the causal impact of teachers on well-being, or by persistent differences in

student characteristics across teachers. Therefore, we focus now on whether there is systematic sorting of students to teachers.

5.1.2 Selection on Observables

Although we cannot directly test for sorting on unobserved characteristics, we can test for sorting on observable characteristics that are excluded from the value-added model. We use two sets of predictors. The first follows Chetty, Friedman and Rockoff (2014*a*) and uses maternal income and twice-lagged well-being. Since our main long-run outcomes concern mental health, the second uses parental psychiatric diagnosis and the student's own prior psychiatric diagnosis.

We show that, after conditioning on our rich control vector, students from higher income families are largely not sorting to higher well-being VA teachers (Panel B of Figure 6). Using twice-lagged well-being, we estimate a small coefficient of −0.03 (s.e. 0.003) (Panel C of Figure 6), which is close to zero and opposite in sign to prior estimates in the literature.[13] The negative coefficient implies that, if anything, students with higher twice-lagged predicted achievement are weakly sorted toward lower value-added teachers, which would bias our estimates downward rather than upward. In all cases, the magnitude of this relationship is small relative to the relationship between teacher value-added and contemporaneous outcome residuals.

Turning to baseline mental health risk, Figure 6 shows that, after conditioning on our control vector, neither parental nor the student's own prior psychiatric diagnosis is meaningfully related to well-being VA. The coefficients are close to zero in both cases, indicating that students at higher baseline mental health risk are not sorting to systematically better or worse well-being VA teachers. Overall, these results mirror the evidence for test score value-added, where models with rich controls, in particular prior scores, yield approximately unbiased estimates of teacher effects (e.g. Chetty, Friedman and Rockoff, 2014*a*; Bacher-Hicks et al., 2019).

A related concern is that teachers themselves affect who responds to the well-being survey. We find that this effect is negligible: a one standard deviation increase in well-being value-added raises the probability of responding by 0.3 percentage points, relative to a mean response rate of 91 percent among students in classes included in our value-added sample (Appendix Table A.2). Because marginal respondents tend to be disadvantaged (Section 2),

[13]For comparison, Chetty, Friedman and Rockoff (2014*a*) report a coefficient of 0.022 (s.e. 0.002), and Gilraine and Pope (2023) report a coefficient of 0.028.

any such selection would, if anything, bias our estimates toward zero.

5.1.3 Quasi-Experimental Variation from Staffing Changes

The evidence in Section 5.1.2 does not rule out the possibility that students are sorted to teachers based on unobservable characteristics that are orthogonal to maternal income and twice-lagged outcomes. Therefore, we use quasi-experimental variation that leverages staffing changes among the full distribution of teachers to obtain an estimate of forecast bias. To do so, we construct a leave-two-year-out estimate of VA for each teacher using data from all years except $t-1$ and t. We then calculate the change in mean teacher VA for each school-grade-subject-year cell.

Panel A of Figure 7 plots changes in cohort-level outcomes against changes in mean well-being VA. We find a strong positive relationship: increases in the quality of the teaching staff predict increases in mean student well-being across cohorts. The baseline estimate exceeds one, but this relationship is sensitive to extreme observations. When trimming the top and bottom five percent of changes, the estimated coefficient is close to one, consistent with forecast unbiasedness. This sensitivity likely reflects the smaller number of cohorts and greater noise in cohort-level means relative to other work that uses longer panels.

As a further placebo test, we repeat the staffing-change analysis using predicted outcomes based on maternal income. In this case, changes in well-being VA are uncorrelated with changes in predicted outcomes, supporting the assumption that staffing changes are not systematically related to changes in student composition (Panel B of Figure 7).

Overall, while the quasi-experimental staffing-change design is necessarily less precise in our setting, due to a shorter panel than other studies and less frequent standardized testing, the full set of bias checks yields a consistent picture. Across forecast bias tests and placebo outcomes, we find little evidence of systematic sorting that would materially bias our value-added estimates.

5.1.4 Repeated Teacher Assignments

In Denmark, teachers often follow the same class for several years, a practice commonly referred to as looping. Among students observed in the previous school year, 56 percent have the same Danish or Mathematics teacher as in that year, and among students observed two years earlier, 31 percent still have the same teacher. This stands in contrast with the usual US setting used for teacher value-added estimation.

This looping phenomenon might have implications for our estimates. The value-added

model treats the lagged outcome as a baseline determined before the current teacher's influence. When the student had the same teacher in the previous year, the lagged outcome already partly reflects that teacher's contribution, and conditioning on it removes part of the effect we aim to measure. Value-added estimates for teachers who frequently follow their class are therefore attenuated toward zero, and the dispersion of teacher effects reported in Section 4 is, if anything, understated. The same logic applies to the long-run regressions, which also condition on the lagged outcome. If anything, looping therefore leads us to understate teachers' effects.

Two exercises suggest that this bias is likely small in practice. First, we re-estimate well-being value-added using only class-years in which the teacher is new to the students, so that the lagged outcome was produced under a different teacher. This removes the contaminated baseline by construction but comes at a cost. Each teacher is observed in fewer class-years, so the estimates are noisier and more heavily shrunk, and the sample is inherently more selected, as this practice might be unevenly distributed across grades and schools. Despite these differences, the resulting measure is highly correlated with our main estimate. The correlation between the two, computed at the teacher-year level, is 0.86, suggesting that the two measures capture largely the same underlying teacher quality.

Second, rather than dropping looping observations, we change the baseline. We replace the lagged outcome in the control vector, together with its class and school-grade means, by its twice-lagged counterpart, so that the baseline precedes the current teacher except for the 31 percent of observations in which the same teacher also taught the student two years earlier. This requires outcomes measured two years earlier, which we do not observe for the first grades of our window, so the measure is estimated on a smaller sample. It correlates 0.65 with our main estimate at the teacher-year level. The lower correlation is expected, since the twice-lagged baseline is a weaker proxy for students' starting points and both measures are estimated with error. Nevertheless, we interpret this as evidence that our main estimates are not largely driven by looping.

The two exercises point the same way. Whether the contaminated baseline is removed by restricting the sample or by moving the control further back, the resulting measures rank teachers similarly to our main estimate. Looping therefore does not seem to materially drive our estimates of teacher effects on well-being. To the extent that it matters, it should bias our main estimates toward zero.

5.2 Bias in the Long-Term Effects

Our approach to estimating teachers' long-term impacts builds on the evidence above that conditioning on prior outcomes and a rich set of observables yields approximately unbiased estimates of teachers' causal effects on contemporaneous outcomes. Under this assumption, differences in longer-run outcomes across students assigned to different teachers—conditional on the same control vector—can be interpreted as reflecting teachers' long-term impacts. We assess the validity of this approach by exploiting quasi-experimental variation in teacher value-added arising from staffing changes across grades and schools. Following Chetty, Friedman and Rockoff (2014*b*), we regress changes in school-grade-year long-term outcomes on changes in the mean teacher value-added weighted by the number of students.

Table 4 reports these quasi-experimental estimates for well-being value-added. The signs of the estimated coefficients are generally consistent with the results presented in Section 4. Despite the loss of statistical power, we continue to find a statistically significant effect of well-being value-added on high school well-being and mental health care utilization, as well as on the remaining outcomes in the expected direction.

6 Further Analysis

Section 4 showed that assignment to a teacher with higher well-being value-added raises students' well-being in high school by 0.044 standard deviations, reduces their use of mental health care between ages 16 and 18 by 5.4 percent relative to the mean, and reduces depression diagnoses. In this section, we ask three questions to better understand these effects. First, what do these teachers change in students' school experience? Second, which students benefit most from a teacher with high well-being value-added? Third, does the timing of exposure matter?

6.1 Mechanisms

There are several ways in which teachers could have lasting effects on students' well-being and mental health. They may help students develop socio-emotional skills that persist beyond their classroom, shape peer relations in the class, prevent early symptoms from becoming persistent problems, or affect students' engagement with school and their academic trajectories. Although we cannot precisely identify these mechanisms, we now ask what these teachers do and change during the year in which students are exposed to them. To do so, we use other questions from the well-being survey that do not enter our well-being measure.

In particular, we use the three school experience indices described in Section 2, which capture students' perceptions of teacher quality, their own agency, and their relationship with peers. We estimate Equation 6 by regressing each index, measured in the year of exposure, on the leave-year-out measure of teacher well-being value-added, conditional on the baseline control vector used in the value-added estimation.[14] To see which behaviors drive each index, we also estimate the same specification separately for each of the survey items that make up the indices. All outcomes are standardized by grade and year, so each coefficient measures the effect of a one standard deviation increase in well-being value-added in standard deviations of the outcome. Figure 8 reports the results.

First, we find that students seem to be able to recognize teachers who improve their well-being. A one standard deviation increase in well-being value-added raises the teacher quality index by 0.103 standard deviations.[15] The largest effect is on whether teachers can quickly create calm when the class is noisy (0.098 standard deviations). This suggests that classroom management is a central part of what well-being value-added captures.

Second, we also find that these teachers impact skills that students may carry beyond the classroom and the interaction with that teacher. Well-being value-added raises student agency by 0.041 standard deviations and peer relations by 0.031 standard deviations. These effects are smaller than the effect on perceived teacher quality, but they concern students' own behavior rather than their view of the teacher.

The effects on agency and peer relations are concentrated in how students handle school-work. The largest effects are on whether students can do something to move on when a task is too difficult (0.042 standard deviations), whether they can make things better when they are bored (0.036), and whether they work well with others (0.033). Remaining items concerning empathy or bullying are less affected, although all of these effects remain statistically significant.

Taken together, these results suggest that much of what well-being value-added captures is what teachers do in the classroom, in particular keeping order. The effects on students' agency and peer relations are smaller, but they may matter more for the long run. Unlike the classroom environment, the abilities to get past difficulties and to work with others are skills that students carry with them beyond a given teacher's classroom, which offers a plausible

[14] Because well-being value-added is a leave-year-out measure, based only on the students a teacher taught in other years, these relationships are not mechanically driven by the responses used to construct it.

[15] Although the survey asks about teachers in general rather than a specific teacher, students spend a large share of their school day with either their Danish or Mathematics teachers, so their answers are likely to reflect these teachers to a large extent.

channel for the persistence of the effects documented in Section 4.

6.2 Heterogeneity by Student Characteristics

We examine heterogeneity in these effects by re-estimating our main specifications separately for subgroups defined by four baseline characteristics: academic achievement (below/above median prior test scores), well-being (below/above median prior well-being), gender, and mental health risk (below/above median predicted probability of mental health care utilization, based on observable characteristics at baseline). For each split, we estimate Equation 6 separately and report the effect of a one standard deviation increase in well-being VA for each subgroup. Figure 9 shows the results across the three main outcomes, together with the p-value from a test of equality of effects across each pair of subgroups.

Effects on self-reported high school well-being are broadly homogeneous, ranging from roughly 0.042 to 0.049 standard deviations across subgroups, with differences that are small, though statistically significant by gender ($p = 0.02$) and by predicted mental health risk ($p = 0.008$). The effects on mental health care utilization, by contrast, differ across groups and are concentrated among students with stronger baseline outcomes. The reduction is larger for students with high baseline test scores than low (−6.3 vs. −4.2 percent relative to the group mean, $p < 0.001$) and for students with high prior well-being than low (−7.4 vs. −4.1 percent, $p = 0.002$). Boys show larger reductions than girls ($p < 0.001$), consistent with evidence that teachers' effectiveness can differ by student gender (Barrios-Fernández and Riudavets-Barcons, 2024). Differences by predicted mental health risk point the same way, with larger reductions for lower-risk students, though they are not statistically significant ($p = 0.21$).

Because baseline utilization differs across groups, relative and absolute effects do not necessarily coincide. For prior test scores and prior well-being, they do: in percentage points, the reduction in care utilization is 0.57 versus 0.81 for students with low and high prior test scores, and 0.67 versus 0.74 for students with low and high prior well-being, so that students who start out ahead benefit at least as much in absolute terms and more relative to their lower baseline. For predicted risk and gender, the ranking reverses. High-risk students and girls experience larger absolute reductions (0.81 versus 0.55 and 0.88 versus 0.54 percentage points), but because their baseline utilization is two to three times higher, the relative effects are smaller.

The pattern is even starker for psychiatric diagnoses. Among students with low baseline test scores, the estimated effect is essentially zero, whereas among high-achieving students it is a reduction of 4.2 percent relative to the mean ($p < 0.001$). The same holds for baseline well-

being: low-WB students show near-zero effects, while high-WB students show reductions of 5.2 percent ($p < 0.001$). There is no significant difference by gender ($p = 0.68$), and lower-risk students show somewhat larger reductions than higher-risk students, although the difference is only marginally significant ($p = 0.09$).

Taken together, these results reveal a consistent pattern: the effects of well-being value-added on subjective well-being are broadly distributed across students, but its effects on objective mental health outcomes are more concentrated among students with higher prior test scores and well-being. These differences may reflect different experiences of the student with the teacher. To investigate this, we re-estimate the effects on the school experience indices from Section 6.1 separately for students with below- and above-median prior well-being.

Students with different baseline characteristics may experience the same teacher differently. This does not appear to be the case: students with below- and above-median prior well-being report nearly identical improvements in perceived teacher quality (0.103 standard deviations in both groups; Appendix Figure A.2). We do, however, find statistically significant differences in how students themselves change. Students with higher prior well-being gain more in agency (0.046 versus 0.035 standard deviations, $p < 0.001$), while students with lower prior well-being gain slightly more in peer relations (0.034 versus 0.027), although this difference is not statistically significant.

These differences are modest, and they are unlikely to fully account for the pattern above. One interpretation is that there are limits to what a classroom teacher can achieve on their own. Clinically relevant mental health problems are relatively rare, and for students who already struggle academically or emotionally, preventing them may require more targeted support than a better classroom experience can provide.

6.3 Dynamic Effects by Grade

We next examine whether the effects of well-being value-added vary across grades. Danish compulsory schooling is formally subdivided into three levels: *indskoling* (grades 1–3), *mellemtrin* (grades 4–6), and *udskoling* (grades 7–9), each associated with changes in curriculum and instruction. Grades 4 and 7 therefore mark natural transition points within compulsory schooling, motivating us to examine whether exposure to high well-being VA teachers is especially consequential at these moments. For each grade from 4 to 9, we estimate the effect of a one standard deviation increase in well-being VA on each outcome. Figure 10 reports the results, alongside the sum of coefficients across all grades as an upper bound on the cumulative effect. The sum equals the cumulative effect only if there is no tracking of

students, which is plausible in our setting, and no diminishing returns to having high-quality teachers in consecutive years (Kinsler, 2016). If returns to consecutive high-quality teachers diminish, the sum overstates the cumulative effect.

Consistent with transitions being particularly consequential, effects on high school well-being are positive in every grade but substantially larger in grades 4 and 7, at approximately 0.08 standard deviations each, than in the other grades. The sum of coefficients is 0.26 standard deviations. The effects on mental health care utilization broadly mirror this pattern: the reduction is largest in grade 4, falls to near zero in grade 6, and increases again in grade 7, with a cumulative sum of −0.051. Taken at face value, having a teacher with one standard deviation higher well-being value-added in every grade from 4 to 9 would reduce the probability of mental health care utilization between ages 16 and 18 by up to 5.1 percentage points. The effects on psychiatric diagnoses are largest in grade 4 and decline steadily across grades, turning slightly positive by grade 9.

Academic outcomes display a similar pattern. The effects of well-being value-added on high school GPA are concentrated in grades 4 and 7, and effects on academic high school enrollment are smallest in grade 6 and rise sharply in grade 7 (Appendix Figure A.3).

This pattern contrasts with Petek and Pope (2023), who find that teachers in later grades have larger effects on high school outcomes. In our setting, the largest effects on mental health outcomes come from grade 4, the grade furthest removed from the outcomes we measure, which suggests that exposure early in the educational life cycle, or at the start of a new school stage, is particularly consequential.

7 Multidimensional Teacher Quality

Recent work emphasizes that teacher quality is multidimensional, with teachers affecting students' long-run outcomes through distinct channels (Petek and Pope, 2023). While the focus of this paper is on measuring teachers' impacts on student well-being and their longer-run mental health consequences, it is natural to ask how these effects relate to more traditional measures of teacher quality. In particular, we examine two questions: (i) whether teachers who improve students' well-being also raise test scores, and (ii) whether teachers who raise test scores have spillover effects on longer-run mental health outcomes.

As reported in Table 3, the SD of teacher effects on test scores is 0.151 for Danish teachers and 0.202 for Mathematics teachers — somewhat larger than the corresponding well-being VA estimates, and consistent with prior estimates in the literature (e.g., Chetty, Friedman and

Rockoff, 2014*a*; Rose, Schellenberg and Shem-Tov, 2025). The jackknife test score VA estimates have a standard deviation of 0.079 (Figure 1).[16] We now ask how these two dimensions of teacher quality relate to each other and to students' long-run outcomes.

7.1 Within-Teacher Correlation Between VA Measures

Figure 11 plots test score and well-being value-added and shows that, although both dimensions of teacher quality are positively correlated, the correlation is relatively weak. Some teachers who perform poorly as measured by test score value-added perform well on the well-being value-added dimension, and vice versa. For example, only 27 percent of teachers in the top quintile of well-being value-added are also in the top quintile of test-score value-added.

However, because each value-added measure is measured with error, the observed correlation between the two may differ from the correlation between teachers' true effects. Random estimation error can attenuate correlations toward zero, while estimation errors that are correlated across outcomes within the same year may bias correlations upward.

To address these concerns, we implement a split-sample approach to estimate an attenuation-corrected correlation between the two VA measures, following Denning et al. (2026). We randomly split students within each teacher-year cell into two independent samples (A and B). We then estimate each teacher's VA on outcome 0 (well-being) and outcome 1 (test scores) separately in each sample, yielding four VA estimates per teacher: μ_0^A, μ_0^B, μ_1^A, μ_1^B. We then estimate the correlation as:

$$\hat{r}_{01} = \frac{\hat{corr}(\mu_0^A, \mu_1^B)}{\sqrt{\hat{corr}(\mu_0^A, \mu_0^B)\hat{corr}(\mu_1^A, \mu_1^B)}} \tag{7}$$

The split-sample correlation in the numerator eliminates spurious correlation arising from shared estimation error across outcomes, while the denominator corrects for attenuation due to measurement error in each value-added measure. We bootstrap this estimate with 100 random sample splits, each yielding two cross-sample estimates, and report the average of the 200 resulting estimates. Applying this procedure yields an attenuation-corrected correlation of approximately 0.25 between teachers' well-being and test score value-added. This estimate is close to the raw correlation (0.20), indicating that the weak relationship between the two measures is not driven by attenuation from estimation error.

[16] Test score value-added passes the same forecast bias and staffing-change tests as well-being value-added (Appendix Figures A.4 and A.5).

Because the degree of shrinkage differs across outcomes, correlating two sets of shrunk estimates may not recover the correlation between teachers' true effects (Rose, Schellenberg and Shem-Tov, 2025). The split-sample procedure corrects for attenuation from estimation error, but as a further check we also apply the unbiased variance and covariance estimator of Kline, Saggio and Sølvsten (2020) and Rose, Schellenberg and Shem-Tov (2025). This estimator characterizes the distribution of teacher effects, but not individual teacher effects. It yields a correlation of 0.35. Although somewhat higher than our main estimate, it still indicates that well-being and test score value-added capture largely distinct dimensions of teacher quality.

7.2 Long-Run Effects of Test Score Value-Added

The weak correlation between well-being and test score value-added suggests that these measures capture distinct dimensions of teacher quality. We next examine whether teachers' test score value-added has independent long-run effects on students' outcomes, and how these effects compare to those associated with well-being value-added. To do so, we estimate models that include both test score and well-being value-added jointly, allowing us to assess the extent to which each dimension independently affects students' long-term outcomes.

Figure 12a shows that, for high school well-being, the effect of well-being value-added remains large and statistically significant: a one–standard deviation increase in well-being VA raises high school well-being by 0.049 standard deviations, while the coefficient on test score value-added is small (0.006) and statistically insignificant.

To understand in more detail which dimensions of students' well-being each type of teacher affects, Figure 13 reports effects on the two components of our high school well-being measure, academic and social well-being, as well as on three additional sub-indices constructed by the Ministry of Education: learning environment, pressure and concerns, and bullying (see Section 2). We find that well-being value-added improves all five dimensions, with the largest effects on academic well-being, social well-being, and learning environment, and effects about half as large on pressure and concerns and on bullying. Test score value-added, by contrast, has a positive and significant effect only on academic well-being, which captures students' engagement, motivation, self-efficacy, and perceived academic performance. This suggests that teachers who raise test scores in middle school not only improve academic achievement in high school, but also improve students' academic well-being, reflecting a coherent academic dimension of teacher quality that shows up both in performance and in how students experience school. At the same time, test score value-added has null or even slightly negative effects on the remaining dimensions — social well-being, learning environment,

pressure and concerns, and bullying — suggesting that the gains associated with high test score VA teachers may come partly at the expense of students' social and emotional experience of school. This illustrates why the two measures capture distinct aspects of teacher quality.

For mental health care utilization, Figure 12b shows that well-being value-added continues to reduce the probability of a visit (−0.008), whereas test score value-added is associated with a small and marginally significant *increase* in utilization (0.002). Turning to psychiatric diagnoses, Figure 12c shows that both dimensions remain relevant: test score value-added is associated with a marginally significant reduction in diagnoses, driven primarily by anxiety, while the aggregate effect of well-being value-added is smaller and imprecisely estimated, though diagnosis-specific patterns persist—well-being value-added continues to predict reductions in depression, while test score value-added predicts reductions in anxiety (Figure 14). Results for academic outcomes in the joint specifications are reported in Appendix Figure A.6.[17]

Taken together, these results highlight the multidimensional nature of teacher quality. Test score and well-being value-added are only weakly correlated, and each predicts different long-run outcomes. Well-being value-added is the primary predictor of students' long-run well-being, mental health care utilization and depression diagnosis, while test score value-added strongly predicts academic outcomes and is associated with reductions in psychiatric diagnosis of anxiety. These patterns indicate that neither measure simply proxies for the other and that evaluations based solely on test score value-added miss important dimensions of teachers' impacts on students' long-run outcomes.

The mechanisms underlying the two dimensions of teacher quality also differ. To examine this, we regress the school experience indices and their items from Section 6.1 on both value-added measures jointly (Appendix Figure A.7). The effects of well-being value-added are essentially unchanged when test score value-added is included. Test score value-added, by contrast, has much smaller effects on all three indices and barely affects peer relations. The few items it affects relate to learning, such as whether teachers can quickly create calm and whether students can move on when stuck. These patterns further suggest that the two measures capture distinct dimensions of teacher quality.

7.3 Implications for Teacher Selection Policies

Hanushek (2011) estimates that replacing the bottom 5 to 8 percent of teachers with teachers of average quality would move the United States near the top of international achievement rankings, and Chetty, Friedman and Rockoff (2014*b*) find that replacing a teacher in the

[17]Quasi-experimental estimates for test score value-added on its own are reported in Appendix Table A.5.

bottom 5 percent of the value-added distribution with an average teacher would raise the present value of a classroom's lifetime earnings by about $250,000. In both cases, as in most evaluation systems that use value-added, teachers are ranked by their contribution to students' test scores.

The weak correlation documented in Section 7.1 means that the two measures identify substantially different teachers as least effective: only 13 percent of teachers in the bottom 5 percent of test score value-added are also in the bottom 5 percent of well-being value-added. In an exercise similar to Petek and Pope (2023) and Rose, Schellenberg and Shem-Tov (2025), we therefore ask how much of the potential gain in student well-being and mental health is forgone when teachers are selected using test scores alone, and how much can be recovered by combining the two measures.

Consider a policy that replaces the bottom 5 percent of teachers with teachers of average quality, where teachers are ranked by a weighted combination of the two value-added measures:

$$I_j = w\,\mu_j^{WB} + (1-w)\,\mu_j^{TS} \tag{8}$$

We start by focusing on three rules: selection on test score value-added alone ($w = 0$), on well-being value-added alone ($w = 1$), and on an equally weighted average of the two ($w = 1/2$). For each rule, we compute the effect on a student's outcomes of being assigned an average teacher instead of a teacher in the bottom 5 percent. The calculation uses the joint estimates of each outcome on both value-added measures from Section 7.2 and the correlation between the two measures, and assumes that value-added is normally distributed.[18]

Table 5 reports the results. On academic outcomes, selecting on test score value-added alone raises high school GPA by 0.070 standard deviations and high school enrollment by 3.1 percentage points, against 0.020 standard deviations and 1.4 percentage points when teachers are selected on well-being value-added. When focusing on well-being and mental health outcomes, the pattern is reversed: selecting on well-being value-added raises high school well-being by 0.104 standard deviations and reduces the probability of mental health care utilization by 1.5 percentage points (11 percent of the mean), against 0.032 standard deviations and no reduction in utilization under selection on test scores. For a class of 21 students, a 1.5 percentage-point reduction corresponds to roughly one fewer student using

[18]We simulate ten million teachers from a joint normal distribution with the estimated correlation between the two measures, and compute the average well-being and test score value-added of the teachers selected by each rule. The calculation considers a single year of exposure and ignores general equilibrium effects.

mental health care every three classes. As for psychiatric diagnoses, they would fall by 0.4 percentage points under selection on test scores only, but by only 0.2 percentage points, and not significantly, under selection on well-being. This is consistent with the results documented in Section 7.2, where the two types of value-added reduce different diagnoses: test score value-added mainly reduces anxiety, which drives its larger effect on overall diagnoses, while well-being value-added mainly reduces depression.

Equal weighting, however, captures most of both. Relative to selecting on test scores alone, it raises the well-being gain from 0.032 to 0.087 standard deviations, and reduces care utilization by 0.9 percentage points, where selection on test scores alone produced no reduction. The trade-off is modest: 17 percent of the GPA gain (0.058 against 0.070) and 6 percent of the enrollment gain (2.9 against 3.1 percentage points). Selection on test scores alone delivers only 31 percent of the well-being gain obtained under selection on well-being, while selection on well-being alone delivers 29 percent of the GPA gain obtained under selection on test scores. This is a direct consequence of the weak correlation between the measures. Because teachers who rank poorly on one dimension are mostly not the same teachers who rank poorly on the other, a combined index removes teachers who are moderately below average on both rather than extremely low on one, and this costs little on either dimension.

The three rules considered so far are three points in a continuum of possible selection rules. Which of them is appropriate depends on how a policymaker values mental health outcomes relative to academic ones. Figure 15 therefore traces the trade-off between the two dimensions as the weight on well-being varies from zero to one, focusing on high school GPA, high school well-being and mental health care utilization. The markers denote $w = 0$, 0.25, 0.5, 0.75 and 1, where the first, third and fifth correspond to the three rows of Table 5.

Panel A plots the gain in high school well-being against the gain in high school GPA. The curve is close to vertical at the test score end, which means that the first departures from selection on test scores alone are nearly costless. Moving from $w = 0$ to $w = 0.25$ reduces the GPA gain from 0.070 to 0.069 standard deviations while raising the well-being gain from 0.032 to 0.060. Because the two measures are weakly correlated, teachers close to the test score cutoff vary widely in well-being value-added. A small weight on well-being is enough to remove the weakest among them, at almost no cost in test scores. Panel B repeats the exercise for mental health care utilization and shows a steeper trade-off. Selection on test scores alone does not reduce utilization. A weight of 0.25 delivers a reduction of 0.4 percentage points, a quarter of what is attainable, and equal weighting delivers 0.9 percentage points, or 60 percent of it, at a cost of 17 percent of the GPA gain. The full reduction of 1.5 percentage points requires $w = 1$,

which lowers the GPA gain to 0.020.

These gains are upper bounds. The simulation ranks teachers on their true value-added, which a policymaker cannot observe. Survey-based measures may also be easier to manipulate than test scores if they carried stakes. Since Danish teachers are not dismissed on the basis of value-added, we interpret this exercise as quantifying the stakes attached to each dimension of teacher quality rather than as evaluating an existing policy.

Taken together, these results give the multidimensionality documented in Sections 7.1 and 7.2 a direct policy interpretation. Evaluating teachers on test scores alone secures nearly all of the attainable academic gains but under a third of the attainable gain in student well-being, and none of the reduction in mental health care utilization. Over a wide range of weights, adding well-being value-added to the selection rule recovers most of these gains at a small academic cost.

8 Conclusions

This paper shows that teachers play a meaningful role in shaping students' well-being and mental health. Importantly, ours is the first paper to examine teachers' impacts on mental health outcomes using administrative records. Using data from Denmark, we document substantial heterogeneity in teachers' impacts on student well-being during middle school. These effects persist: students assigned to a teacher with higher well-being value-added report higher well-being in high school several years later. These students are also less likely to use mental health care and to be diagnosed with depression between ages 16 and 18.

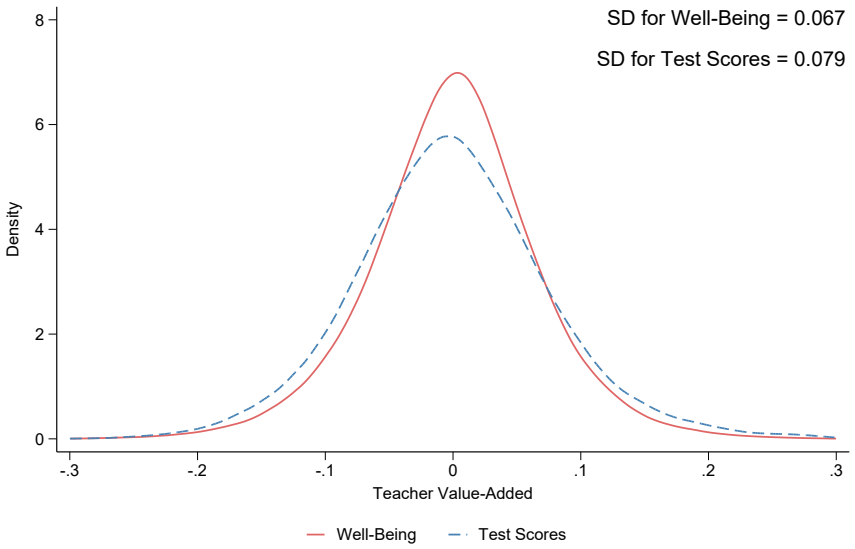
Our results highlight the multidimensional nature of teacher quality. Teachers who are effective at improving student well-being are often not the same teachers who raise test scores, and these two dimensions of teacher quality predict different long-run outcomes. Well-being value-added is the primary predictor of students' subsequent well-being and mental health care utilization, while test score value-added strongly predicts later academic outcomes. Evaluations of teacher effectiveness that rely solely on test scores therefore miss important channels through which teachers shape students' long-run trajectories. For example, a policy that replaces the bottom 5 percent of teachers ranked on both measures nearly triples the gain in high school well-being relative to one based on test scores alone, at the cost of 17 percent of the gain in high school GPA. It also reduces the probability of mental health care utilization by 0.9 percentage points, whereas selection on test scores alone produces no reduction.

More broadly, our findings suggest that schools play an important role in child and

adolescent mental health, a growing concern with key implications for adult outcomes in the labor market. Our results suggest that incorporating student well-being and mental health into the analysis of education policies provides a richer understanding of the production of human capital.

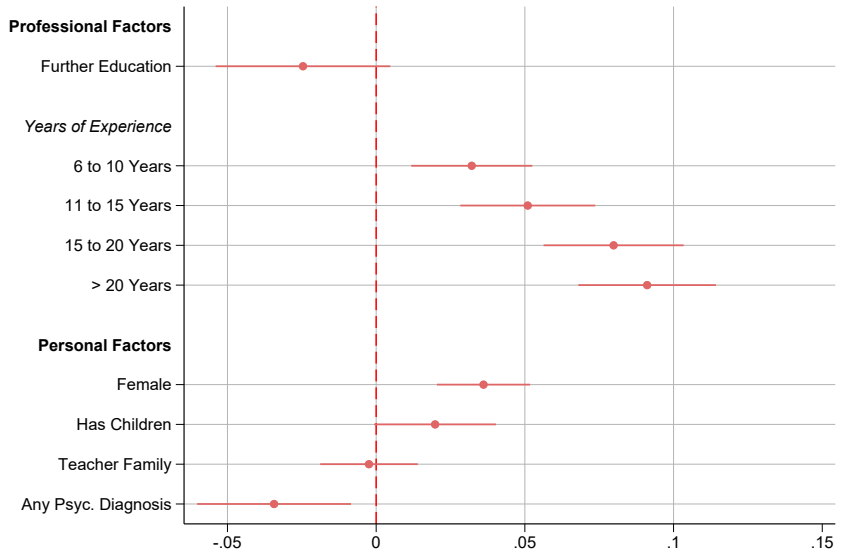
Figures

Figure 1: Empirical Distributions of Teacher VA Estimates



Notes: This figure plots kernel densities of the empirical distribution of teacher VA estimates on well-being and test scores, pooled across subjects. We also report the standard deviations of these empirical distributions of VA estimates. Note that these standard deviations are smaller than the standard deviation of true teacher effects reported in Table 3 because VA estimates are shrunk toward the mean to account for noise and obtain unbiased forecasts.

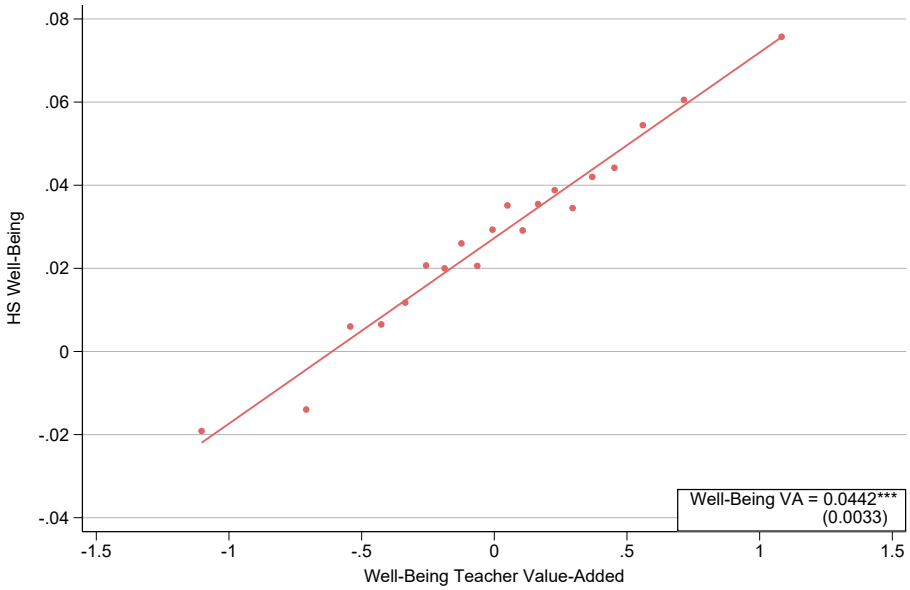
Figure 2: Teacher Characteristics and Well-Being Value-Added



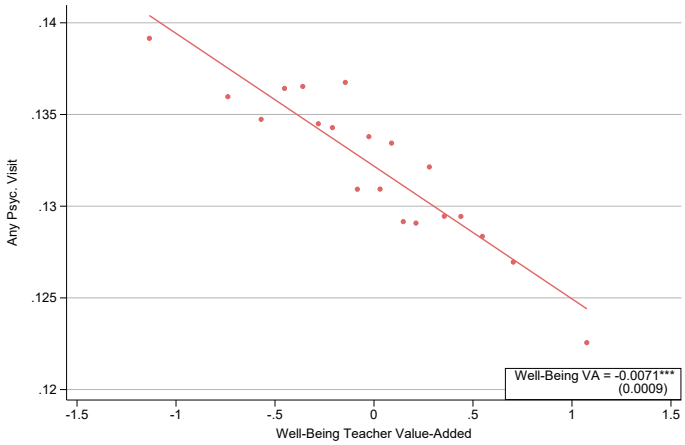
Notes: This figure reports estimates of the association between teacher well-being value-added and a set of observable professional and personal teacher characteristics. Each coefficient is from a regression of the teacher's well-being VA estimate on the indicated characteristic, estimated at the teacher-year level with standard errors clustered at the teacher level. Professional factors include an indicator for further education beyond a teaching degree and indicators for years of experience relative to fewer than six years (the omitted category). Personal factors include indicators for being female, having children, coming from a teacher family, and having any psychiatric diagnosis.

Figure 3: Effects of Teacher Well-Being Value-Added on Well-Being and Mental Health

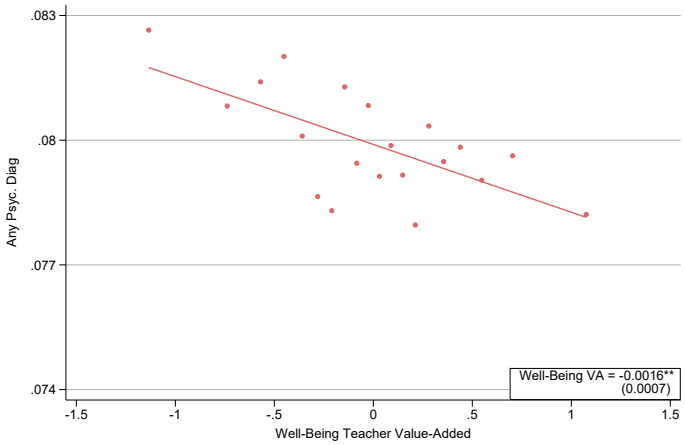
(a) High School Well-Being



(b) Mental Health Care Utilization (Ages 16–18)

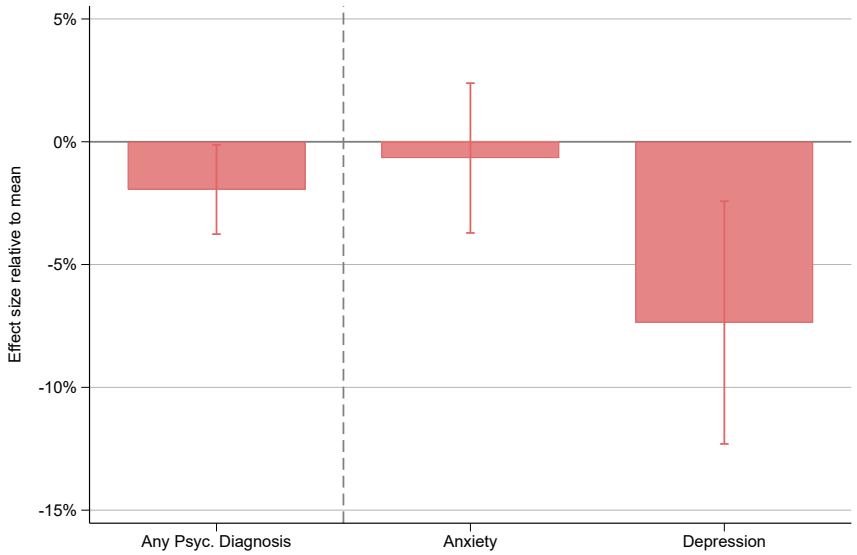


(c) Psychiatric Diagnosis (Ages 16–18)



Notes: This figure plots the relationship between teacher well-being value-added in middle school and students' well-being and mental health outcomes. In particular, we focus on three outcomes: (i) self-reported well-being during the first year of high school, (ii) indicator that takes value one if the student visited a psychiatrist or psychologist between ages 16 and 18, and (iii) indicator that takes value one if the student was diagnosed with any psychiatric condition between ages 16 and 18. The unit of observation is a student–subject–year. Both the outcome and the teacher value-added measure are residualized with respect to the baseline control vector used in the value-added estimation. Teacher value-added estimates are divided into 20 equal-sized bins, and the figure plots the mean residualized outcome within each bin against the mean value-added estimate, with the unconditional means added back. We also plot the best linear fit estimated before binning the data and report its slope coefficient and standard error, clustered at the student and classroom level.

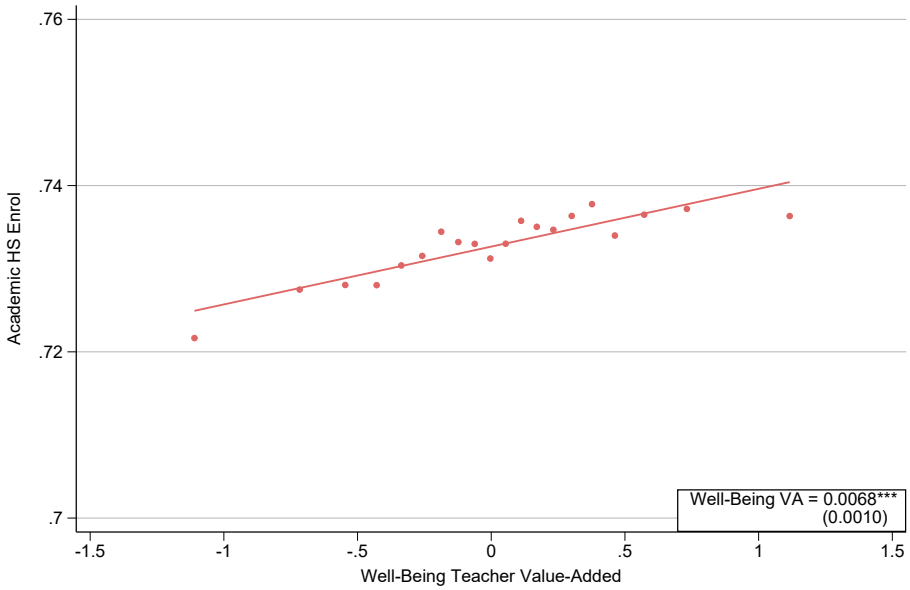
Figure 4: Effects of Well-Being Value-Added on Psychiatric Diagnoses



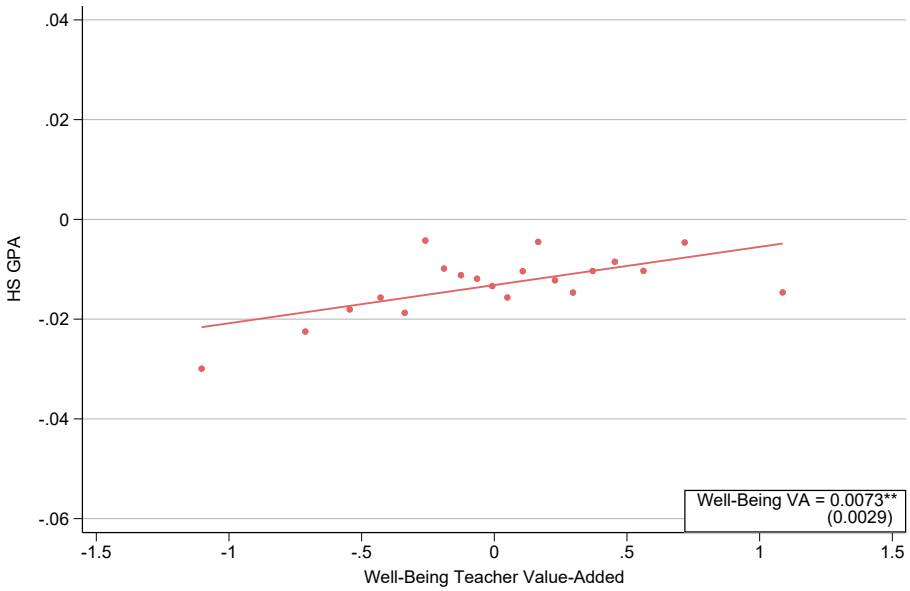
Notes: This figure reports estimates of the effects of well-being value-added on the probability of receiving specific psychiatric diagnoses between ages 16 and 18. The value-added measure is standardized and constructed using leave-year-out estimates. All regressions include the baseline control vector used in the value-added estimation. Standard errors are clustered at the student and classroom level. Point estimates and standard errors are reported in Appendix Table A.4.

Figure 5: Effects of Teacher Well-Being Value-Added on Academic Outcomes

(a) Academic HS Enrollment



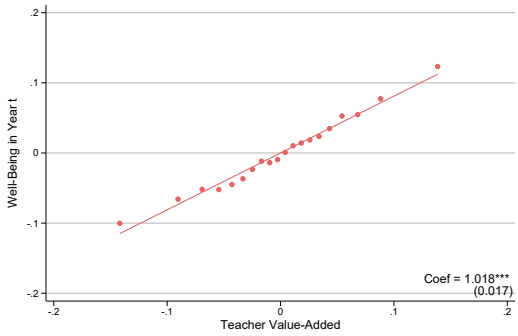
(b) High School GPA



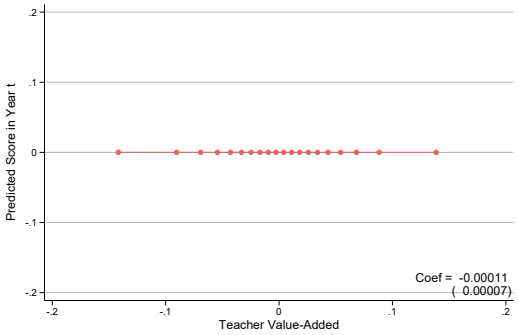
Notes: This figure plots the relationship between teachers' well-being value-added in middle school and students' academic outcomes. In particular, we focus on two outcomes: (i) indicator that takes value one if the student enrolled in academic high school and (ii) high school GPA, standardized by graduation year and academic track. The unit of observation is a student–subject–year. Both the outcome and the teacher value-added measure are residualized with respect to the baseline control vector used in the value-added estimation. Teacher value-added estimates are divided into 20 equal-sized bins, and the figure plots the mean residualized outcome within each bin against the mean value-added estimate, with the unconditional means added back. We also plot the best linear fit estimated before binning the data and report its slope coefficient and standard error, clustered at the student and classroom level.

Figure 6: Effects of Well-Being Teacher Value-Added on Actual and Predicted Scores

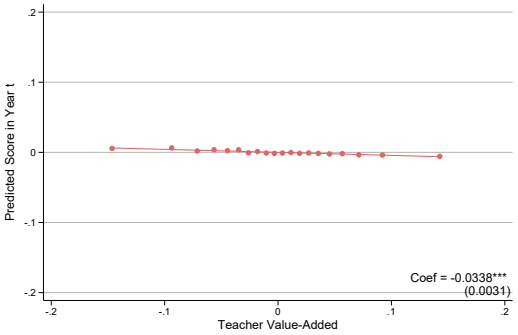
(a) Actual Score



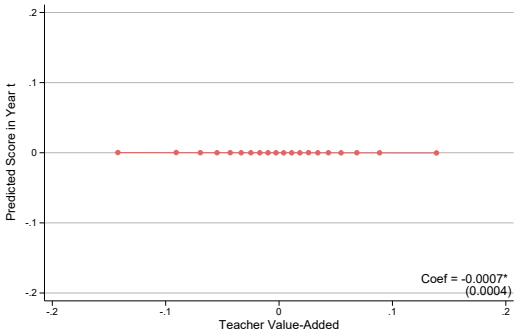
(b) Predicted Score Using Maternal Income



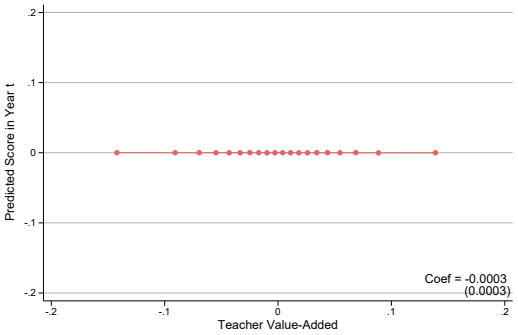
(c) Predicted Score Using Twice-Lagged Well-Being



(d) Predicted Score Using Parental Psychiatric Diagnosis

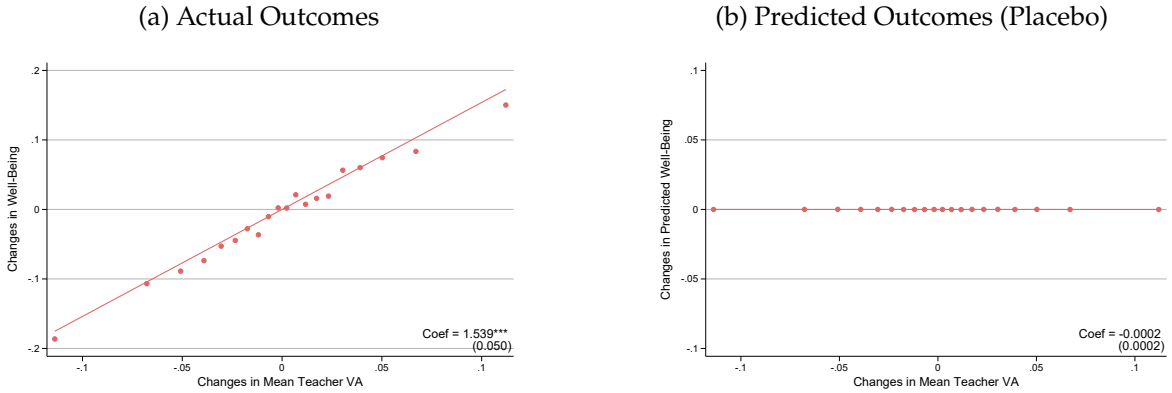


(e) Predicted Score Using Prior Student Psychiatric Diagnosis



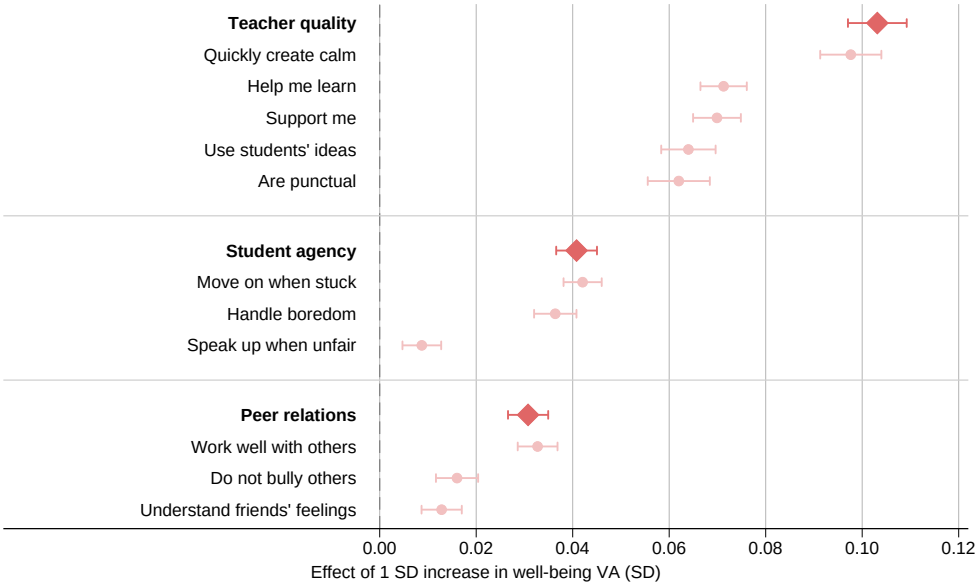
Notes: This figure assesses the validity of the well-being value-added estimates. Panel A plots contemporaneous residualized well-being against teacher well-being value-added. Under the stationarity assumption, the coefficient should equal one. Panels B to E assess whether students sort to teachers based on observable characteristics excluded from the value-added model. Panel B plots predicted outcomes based on maternal income, and Panel C plots predicted outcomes based on twice-lagged well-being. Because our main long-run outcomes concern mental health, Panels D and E test for sorting on baseline mental health risk, using predicted outcomes based on parental psychiatric diagnosis and on the student's own prior psychiatric diagnosis, respectively. Predicted outcomes are obtained by residualizing the predictor with respect to the baseline control vector and regressing residualized outcomes on this residualized predictor. In all panels, teacher value-added estimates are divided into 20 equal-sized bins and the figures plot the mean outcome within each bin against the mean value-added estimate. Solid lines show best linear fits estimated on the underlying microdata using OLS. Standard errors are clustered at the school-cohort level.

Figure 7: Effects of Changes in Teaching Staff on Student Outcomes Across Cohorts: Well-Being Value-Added



Notes: This figure plots changes in cohort-level mean outcomes against changes in mean well-being value-added within school–grade–subject cells, exploiting staffing changes across cohorts. Panel A uses actual outcomes, while Panel B uses predicted outcomes based on maternal income as a placebo check. Teacher value-added is constructed using leave-two-year-out estimates. Lines show best linear fits estimated on the underlying microdata.

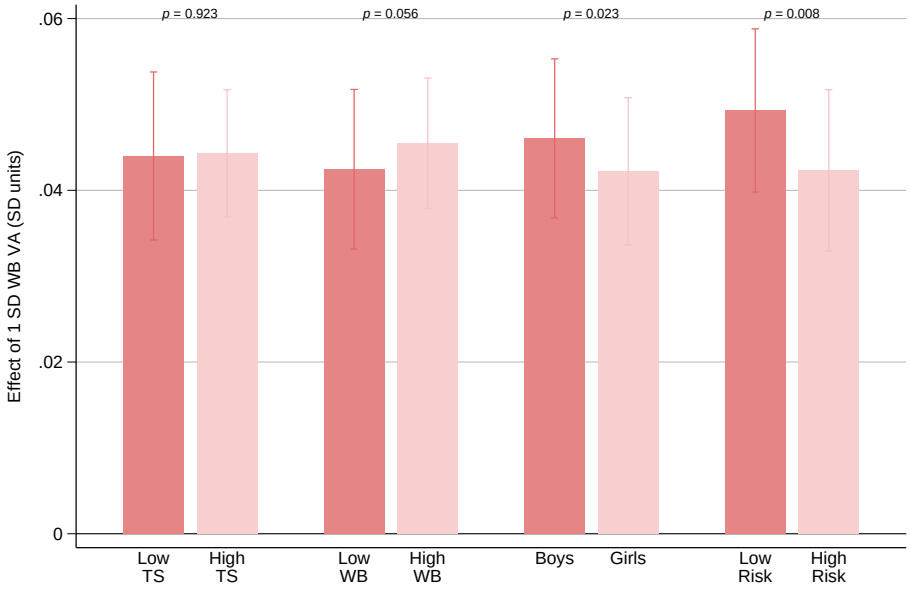
Figure 8: Effects of Well-Being Value-Added on School Experience



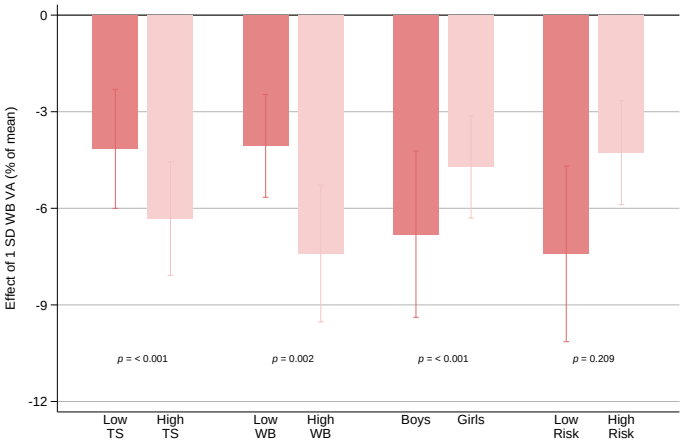
Notes: This figure reports estimates of the effect of a one standard deviation increase in well-being value-added on three school experience indices (diamonds) and on the survey items that make up each index (circles): teacher quality, student agency, and peer relations. The items are not used in our well-being measure. Item labels are abbreviated; the full wording is listed in Appendix B.3. Indices and items are measured in the same year as the teacher assignment and standardized by grade and year. Each coefficient is from a separate regression of the outcome on well-being value-added, conditional on the baseline control vector used in the value-added estimation. Error bars indicate 95 percent confidence intervals, with standard errors clustered at the student and classroom level.

Figure 9: Heterogeneous Effects

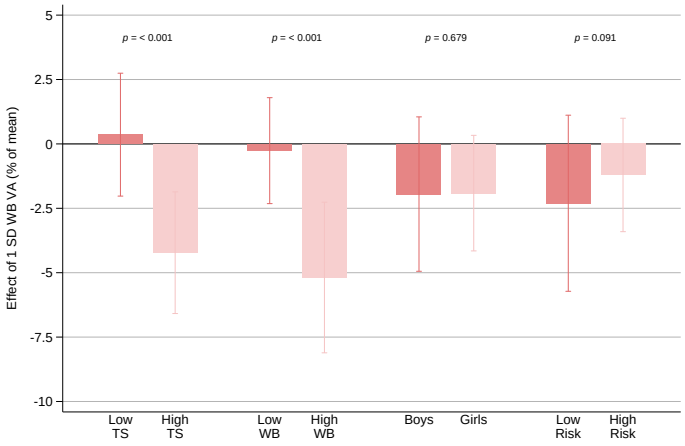
(a) High School Well-Being



(b) Mental Health Care Utilization (Ages 16–18)



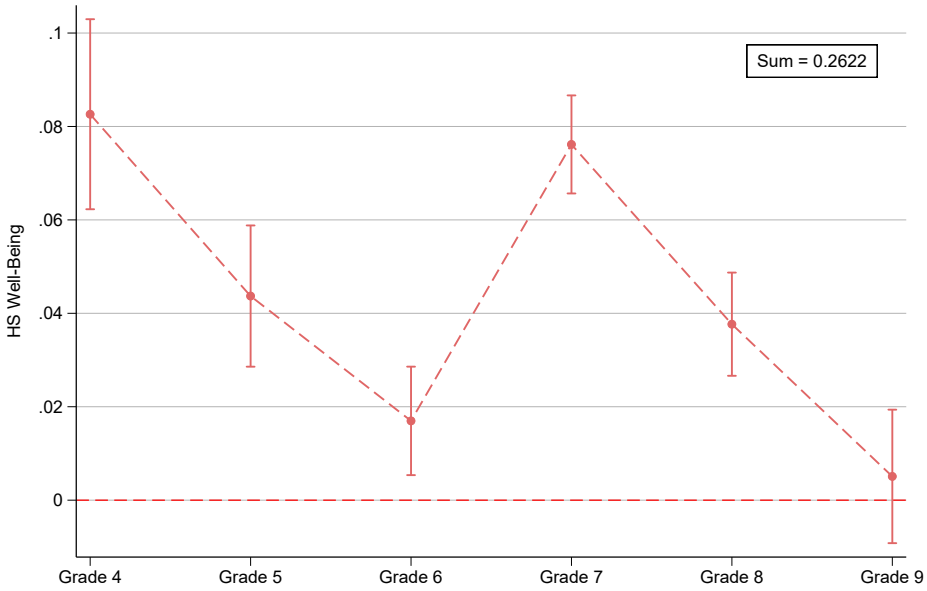
(c) Psychiatric Diagnosis (Ages 16–18)



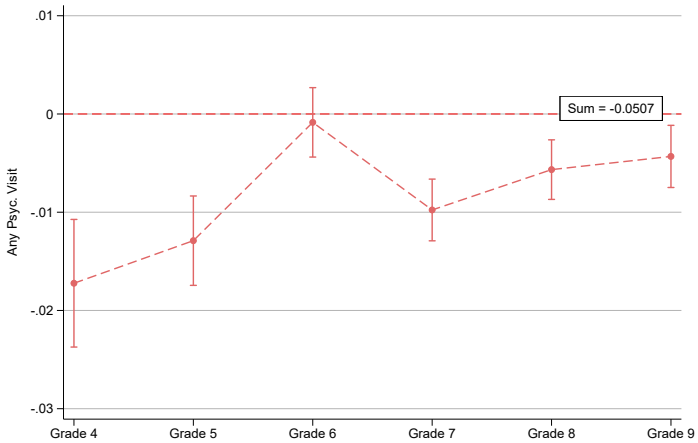
Notes: These figures report heterogeneous effects of teacher well-being value-added on high school well-being, mental health care utilization, and psychiatric diagnoses by student gender, prior academic achievement, prior well-being, and predicted risk of mental health care utilization. Each panel presents coefficients from split-sample regressions estimated separately for the indicated subgroup. Teacher value-added is standardized, and all specifications include the baseline control vector used in the value-added estimation. Effects in Panel A are in standard deviation units; effects in Panels B and C are expressed as a percentage of the subgroup mean. The *p*-values reported correspond to a test of equality of the effects in levels across subgroups, estimated from a pooled regression interacting teacher value-added with the subgroup indicator. Point estimates and standard errors are reported in Appendix Table A.6.

Figure 10: Effects of Well-Being Value-Added by Grade

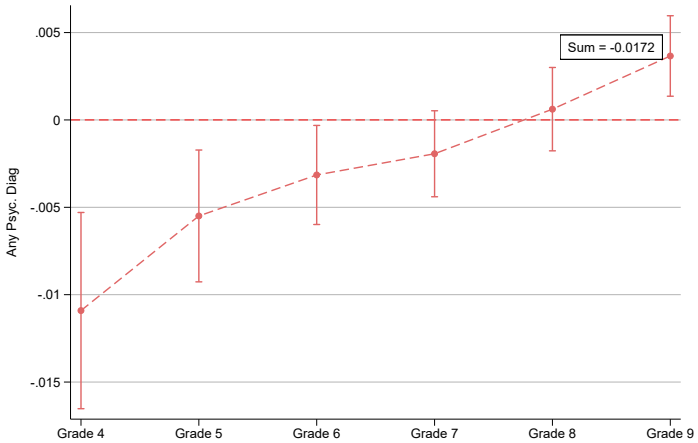
(a) High School Well-Being



(b) Mental Health Care Utilization (Ages 16–18)

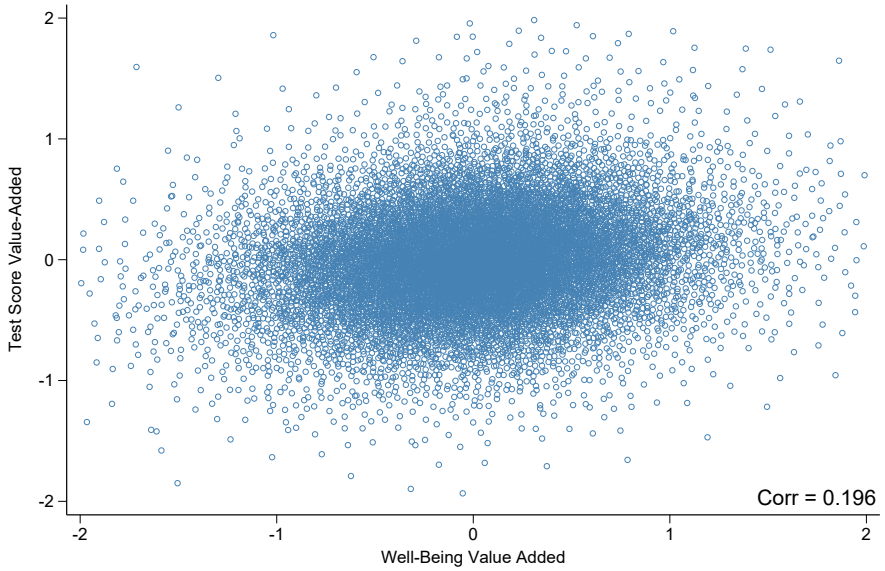


(c) Psychiatric Diagnosis (Ages 16–18)



Notes: This figure plots the effect of well-being value-added on three outcomes by the grade level of the student: self-reported well-being during the first year of high school (Panel A), an indicator for visiting a psychiatrist or psychologist between ages 16 and 18 (Panel B), and an indicator for receiving a psychiatric diagnosis between ages 16 and 18 (Panel C). The plotted coefficients and 95 percent confidence intervals are from regressions of each outcome on well-being value-added and the baseline control vector used in the value-added estimation, estimated separately for each grade. Standard errors are clustered at the classroom level. Each panel also reports the sum of coefficients across all grades, which under the assumption of no diminishing returns provides an upper bound on the cumulative effect of having a one standard deviation higher teacher VA in every grade from 4 to 9.

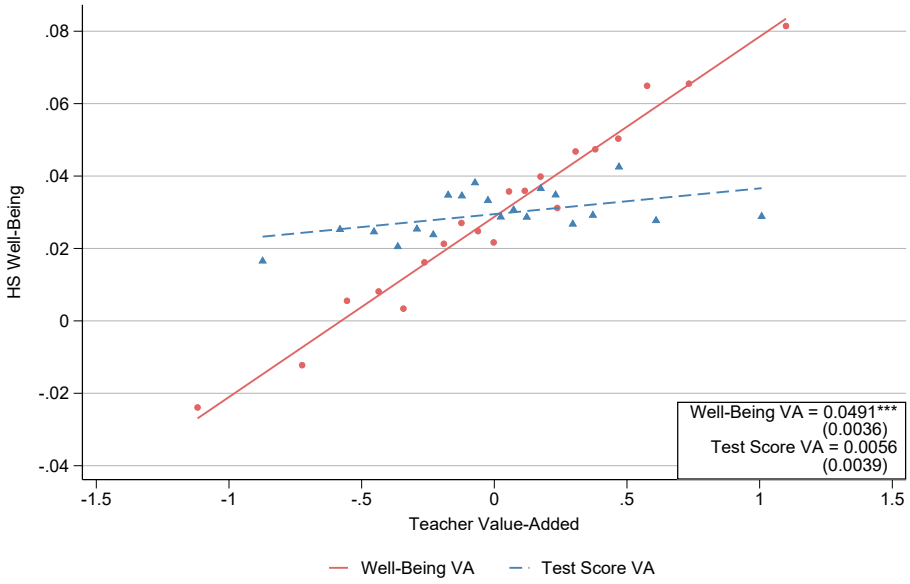
Figure 11: Within-Teacher Correlation Between VA Measures



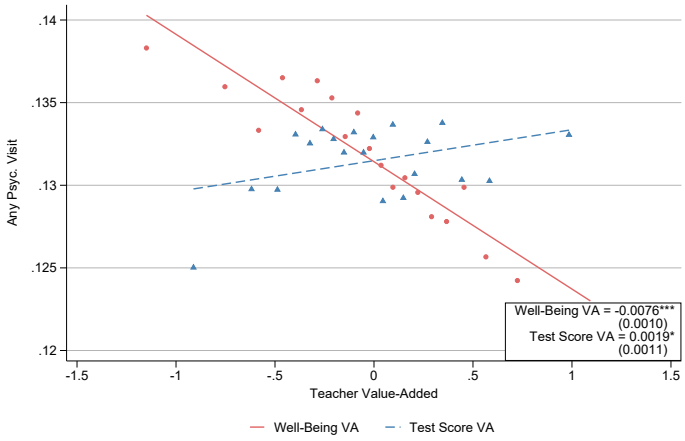
Notes: This figure plots teachers' normalized test score and well-being value-added within 2 standard deviations of the mean. For confidentiality reasons, each point shows the mean of 6 nearest-neighbor observations.

Figure 12: Effects of Teacher Value-Added on Well-Being and Mental Health (Multivariate)

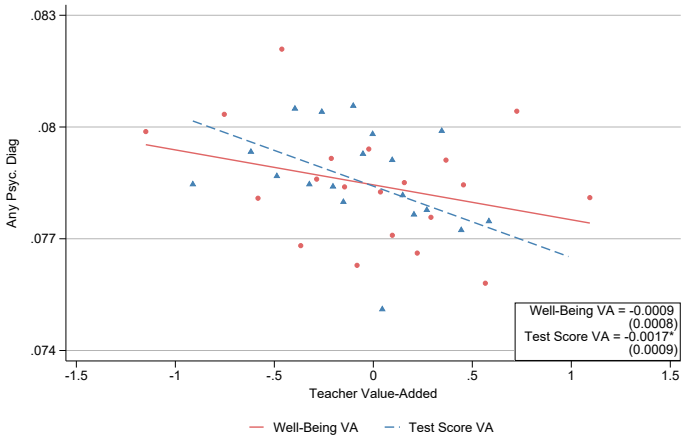
(a) High School Well-Being



(b) Mental Health Care Utilization (Ages 16–18)

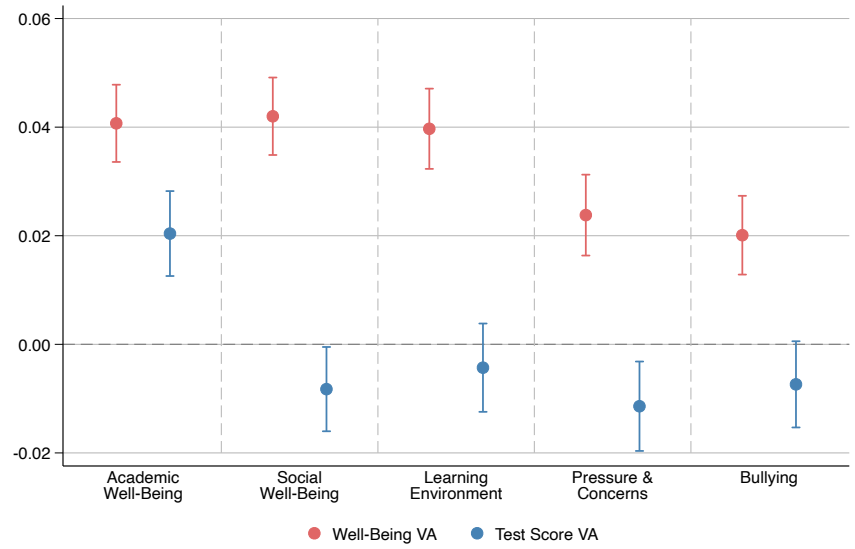


(c) Psychiatric Diagnosis (Ages 16–18)



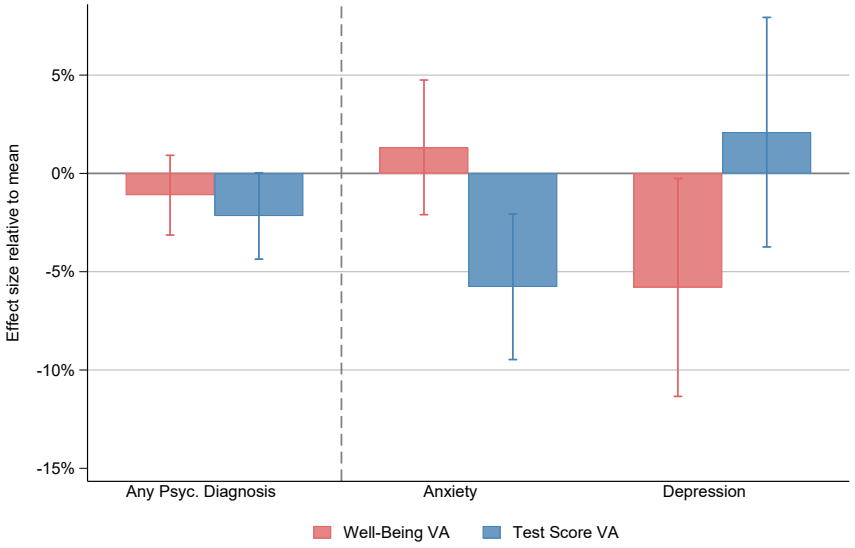
Notes: This figure plots the partial relationship between teachers' value-added in middle school and students' well-being and mental health outcomes. In particular, we focus on three outcomes: (i) self-reported well-being during the first year of high school, (ii) indicator that takes value one if the student visited a psychiatrist or psychologist between ages 16 and 18, and (iii) indicator that takes value one if the student was diagnosed with any psychiatric condition between ages 16 and 18. The figure displays two series corresponding to teacher well-being value-added and test score value-added, respectively; both series are based on a single multivariate specification in which the outcome is regressed simultaneously on both value-added measures. The unit of observation is a student–subject–year. Both the outcome and the teacher value-added measure are residualized with respect to the baseline control vector used in the value-added estimation. Teacher value-added estimates are divided into 20 equal-sized bins, and the figure plots the mean residualized outcome within each bin against the mean value-added estimate, with the unconditional means added back. We also plot the best linear fit estimated before binning the data and report its slope coefficient and standard error, clustered at the student and classroom level.

Figure 13: Effects of Teacher Value-Added on High School Well-Being Sub-Indices



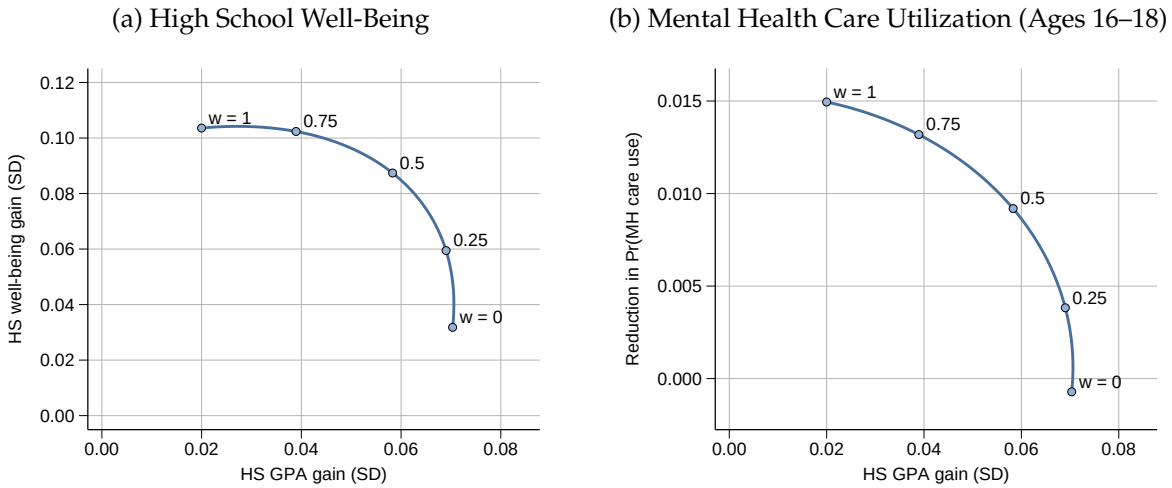
Notes: This figure reports estimates of the effects of well-being value-added and test score value-added on the five high school well-being sub-indices constructed by the Ministry of Education: academic well-being, social well-being, learning environment, pressure and concerns, and bullying. The headline high school well-being measure used in the main analysis is the average of the academic and social well-being sub-indices. All sub-indices are coded so that higher values indicate more positive experiences. For each sub-index, both coefficients come from a single regression of the standardized sub-index on well-being and test score value-added jointly, conditional on the baseline control vector used in the value-added estimation. Error bars indicate 95 percent confidence intervals, with standard errors clustered at the student and classroom level.

Figure 14: Joint Effects of Teacher Value-Added on Psychiatric Diagnoses



Notes: This figure reports estimates of the joint effects of well-being and test score value-added on the probability of receiving specific psychiatric diagnoses between ages 16 and 18. Both value-added measures are standardized and constructed using leave-year-out estimates. All regressions include the baseline control vector used in the value-added estimation. Standard errors are clustered at the student and classroom level. Point estimates and standard errors are reported in Appendix Table A.7.

Figure 15: Gains from Teacher Selection at Different Weights on Well-Being Value-Added



Notes: This figure traces the gains from replacing the bottom 5 percent of teachers with an average teacher, where the bottom 5 percent is identified by the index $I_j = w\,\mu_j^{WB} + (1-w)\,\mu_j^{TS}$, as the weight w on well-being value-added varies from zero to one. Marker labels give the value of w. Gains are per affected student for a single year of exposure. Panel A plots the gain in high school well-being against the gain in high school GPA, both in standard deviations. Panel B plots the reduction in the probability of mental health care utilization between ages 16 and 18 against the gain in high school GPA. Both panels are computed from the joint estimates of each outcome on the two value-added measures and the estimated correlation between them, and at $w = 1/2$ they reproduce the third row of Table 5.

Tables

Table 1: Descriptive Statistics

	Full Sample			Restricted Sample		
	Mean	SD	Obs.	Mean	SD	Obs.
A. Demographics						
Female	0.49	0.50	1,920,705	0.50	0.50	1,006,673
Age	12.14	1.79	1,920,705	12.00	1.76	1,006,673
Migration Background	0.12	0.32	1,920,704	0.11	0.32	1,006,673
Mother's Education	0.40	0.49	1,871,910	0.42	0.49	986,072
Father's Education	0.27	0.45	1,816,042	0.29	0.45	959,753
B. Short-Run Outcomes						
Well-Being Score	0.02	0.99	1,421,451	0.04	0.98	1,006,736
Danish Test Score	0.05	0.95	1,089,904	0.09	0.92	542,378
Math Test Score	0.04	0.97	635,548	0.10	0.94	330,346
C. Long-Run Outcomes						
High School Enrollment	0.72	0.45	1,922,957	0.73	0.45	1,006,736
High School GPA	0.00	1.00	1,054,271	0.01	1.00	500,186
High School Well-Being	0.00	0.99	756,903	0.03	0.98	425,497
Mental Health Care Utilization	0.14	0.34	1,637,792	0.13	0.34	788,832
Any Psychiatric Diagnosis	0.09	0.28	1,452,573	0.08	0.27	646,138
Any Depression Diagnosis	0.01	0.11	1,452,573	0.01	0.11	646,138
Any Anxiety Diagnosis	0.03	0.18	1,452,573	0.03	0.17	646,138

Notes: This table presents summary statistics for demographics, short-run outcomes, and long-run outcomes in the full sample and the restricted analysis sample, as described in Section 2. Parental education variables take value 1 if the parent has a tertiary degree. Well-being and test scores are standardized by cohort. Long-run outcomes unrelated to high school are measured between ages 16 and 18. In each analysis, we use the largest sample possible given when an outcome is observed.

Table 2: Relationship between Short-Run Measures and Long-Run Outcomes

	HS Well-Being	Psyc. Visit	Psyc. Diag.	HS Enrol	HS GPA
Well-Being	0.289***	-0.0304***	-0.0253***	0.0344***	0.0521***
	(0.00323)	(0.000842)	(0.000704)	(0.000790)	(0.00235)
Test Scores	0.0194***	-0.00223***	-0.00422***	0.0899***	0.332***
	(0.00265)	(0.000731)	(0.000610)	(0.000809)	(0.00250)
Observations	366,640	569,155	571,051	733,824	443,124
Mean	0.028	0.128	0.073	0.760	0.006
R-squared	0.192	0.047	0.036	0.259	0.338

Notes: Each column reports coefficients from a regression of the long-run outcomes on student well-being and test scores during middle school. Well-being and test scores are standardized by cohort. The unit of observation is a student–subject–year. Outcomes include high school well-being, mental health care utilization, psychiatric diagnoses, academic high school enrollment, and high school GPA. All specifications include the same baseline control vector used in the value-added analysis, including lagged outcomes, student demographics, classroom and school controls, and grade-by-year fixed effects. Mean reports the mean of the dependent variable in the estimation sample. Standard errors, clustered at the student and classroom level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 3: Variance Estimates

	Well-Being			Test Scores		
	Danish	Math	Pooled	Danish	Math	Pooled
Total SD	–	–	0.774	0.628	0.614	0.627
Individual-level SD	–	–	0.741	0.585	0.564	0.582
Class-level SD	–	–	0.194	0.173	0.135	0.150
Teacher SD	0.120	0.101	0.109	0.151	0.202	0.181
Estimate Teacher SD	0.134	0.094	0.108	0.129	0.144	0.132

Notes: This table reports the standard deviation of outcome residuals and decomposes the total variation into individual-, class-, and teacher-level components for test scores and well-being outcomes in Danish and Mathematics. The first row reports the total standard deviation of residuals. Rows 2–4 decompose this variation into components attributable to individual-level variation, class-level shocks, and teacher-level variation, respectively. The variances corresponding to the individual-, class-, and teacher-level components sum to the total variance. In Row 4, we estimate the standard deviation of teacher effects as the square root of the covariance of mean score residuals across a random pair of classrooms within the same year. In cases where teachers are observed teaching only one classroom per year and class-level and teacher-level shocks cannot be separately identified, we report a lower bound on the within-year standard deviation of teacher effects, computed as the square root of the autocovariance of classroom residuals across a one-year lag (Row 5).

Table 4: Quasi-Experimental Estimates: Effects of Well-Being Value-Added on Long-Run Outcomes

	HS Well-Being	Psyc. Visit	Psyc. Diag.	HS Enrol	HS GPA
Well-Being VA	0.0893***	-0.0072**	-0.0042*	0.0108***	0.0187**
	(0.0079)	(0.0029)	(0.0024)	(0.0022)	(0.0075)
Observations	23,569	23,773	24,060	33,745	25,491
R-squared	0.008	0.001	0.000	0.035	0.001

Notes: This table reports estimates of the effect of well-being value-added on long-run outcomes using quasi-experimental variation in teacher quality generated by staffing changes across adjacent cohorts within school–grade cells. The unit of observation is a school–grade–cohort. Each column reports the coefficient from a regression of the change in the school–grade–cohort mean of the outcome indicated in the column header on the change in the corresponding mean of teacher well-being value-added, and can be read as the effect of a one standard deviation increase in well-being value-added. Teacher value-added is estimated using leave-two-year-out measures of teacher quality. All specifications include year fixed effects and are weighted by cohort size. Standard errors, clustered at the school–cohort level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5: Effect of Replacing the Bottom 5 Percent of Teachers

Bottom 5% selected on	HS Well-Being	Psyc. Visit	Psyc. Diag.	HS Enrol	HS GPA
Test score VA only	0.032***	0.001	−0.004**	0.031***	0.070***
	(0.008)	(0.002)	(0.002)	(0.002)	(0.007)
Well-being VA only	0.104***	−0.015***	−0.002	0.014***	0.020***
	(0.008)	(0.002)	(0.002)	(0.002)	(0.007)
Equal weights	0.087***	−0.009***	−0.004**	0.029***	0.058***
	(0.008)	(0.002)	(0.002)	(0.003)	(0.008)
Mean	0.030	0.131	0.078	0.737	0.005

Notes: Each cell reports the effect on the outcome indicated in the column header of being assigned an average teacher instead of a teacher in the bottom 5 percent, per affected student and for a single year of exposure. Rows differ in how the bottom 5 percent is selected: on test score value-added only (row 1), on well-being value-added only (row 2), or with equal weights on the two measures (row 3). The simulation uses the estimated effects of both measures on each outcome reported in Section 7.2 and the within-teacher correlation between the two measures, and assumes that the two value-added measures are normally distributed. The sample is restricted to students whose teacher has both value-added measures. The last row reports the mean of each outcome in this sample. Teachers are ranked on true value-added, so the estimates are an upper bound on what a feasible policy could achieve. The standard errors on the estimated forecast are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

References

Adhvaryu, Achyuta, James Fenske, and Anant Nyshadham. 2019. "Early life circumstance and adult mental health." *Journal of Political Economy*, 127(4): 1516–1549.

Bacher-Hicks, Andrew, Mark J. Chin, Thomas J. Kane, and Douglas O. Staiger. 2019. "An experimental evaluation of three teacher quality measures: Value-added, classroom observations, and student surveys." *Economics of Education Review*, 73: 101919.

Backes, Ben, James Cowan, Dan Goldhaber, and Roddy Theobald. 2024. "How to Measure a Teacher: The Influence of Test and Nontest Value-Added on Long-Run Student Outcomes." *Journal of Human Resources*.

Baird, Sarah, Jacobus De Hoop, and Berk Özler. 2013. "Income shocks and adolescent mental health." *Journal of Human Resources*, 48(2): 370–403.

Barrios-Fernández, Andrés, and Marc Riudavets-Barcons. 2024. "Teacher value-added and gender gaps in educational outcomes." *Economics of Education Review*, 100: 102541.

Bau, Natalie, and Jishnu Das. 2020. "Teacher value added in a low-income country." *American Economic Journal: Economic Policy*, 12(1): 62–96.

Beuchert, Louise V, and Anne B Nandrup. 2018. "The Danish national tests at a glance." *Nationaløkonomisk tidsskrift*, 2: 1–37.

Bütikofer, Aline, Rita Ginja, Katrine V Løken, and Fanny Landaud. 2023. "Higher-achievement schools, peers and mental health." *The Economic Journal*, 133(655): 2580–2613.

Bütikofer, Aline, Rita Ginja, Krzysztof Karbownik, and Fanny Landaud. 2024. "(Breaking) intergenerational transmission of mental health." *Journal of Human Resources*, 59(S): S108–S151.

Chen, Kuan-Ming, Janet Currie, Hui Ding, and Wei-Lun Lo. 2026. "School milestones impact child mental health in Taiwan." *The Economic Journal*, ueag084.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014*a*. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014*b*. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9): 2633–2679.

Currie, Janet. 2024. "The economics of child mental health: Introducing the causes and consequences of child mental health special issue." *Journal of Human Resources*, 59(S): S1–S13.

Currie, Janet, N. Meltem Daysal, Mette Gørtz, and Jonas Cuzulan Hirani. 2026. "Child Disability and Effects on Sibling Mental Health." Working Paper.

Dalsgaard, Søren, Erla Thorsteinsson, Betina B Trabjerg, Jörg Schullehner, Oleguer Plana-Ripoll, Isabell Brikell, Theresa Wimberley, Malene Thygesen, Kathrine Bang Madsen, Allan Timmerman, et al. 2020. "Incidence rates and cumulative incidences of the full spectrum of diagnosed mental disorders in childhood and adolescence." *JAMA psychiatry*, 77(2): 155–164.

Danish Ministry of Education. 2016. "Methods Brief: Calculating the Well-being Indicators (Metodenotat om beregning af trivselsindikatorer, 2014/2015)." https://www.uvm.dk/-/media/filer/uvm/stat/pdf16/160316-metodenotat-vedr-indikatorberegningerne-24022016.pdf, Accessed: 2025-10-03.

Danish Ministry of Education. 2023. "Methods Brief: Factor Analysis and Calculating the Well-being Indicators (Metodebeskrivelse – faktoranalyse og indikatorfastlæggelse for den gymnasiale trivselsmåling)." https://www.uvm.dk/-/media/filer/uvm/udd/gym/pdf20/jan/200123-metodenotat-indikatorberegning-gymnasial-trivsel.pdf, Accessed: 2025-10-03.

Dee, Thomas S, and Hans Henrik Sievertsen. 2018. "The gift of time? School starting age and mental health." *Health Economics*, 27(5): 781–802.

Denning, Jeffrey T., Rachel L. Nesbit, Nolan G. Pope, and Merrill Warnick. 2026. "Easy A's, Less Pay: The Long-Term Effects of Grade Inflation." National Bureau of Economic Research Working Paper 34952.

Fleche, Sarah. 2026. "Teacher Effects on Students' Socio-Emotional Skills and Young Adult Outcomes." *The Economic Journal*. Forthcoming.

Gilraine, Michael, and Nolan G. Pope. 2023. "Making Teaching Last: Long-Run Value-Added." Working Paper.

Hanushek, Eric A. 2011. "The economic value of higher teacher quality." *Economics of Education Review*, 30(3): 466–479.

Jackson, C. Kirabo. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes." *Journal of Political Economy*, 126(5): 2072–2107. Publisher: The University of Chicago Press.

Kinsler, Josh. 2016. "Teacher complementarities in test score production: Evidence from primary school." *Journal of Labor Economics*, 34(1): 29–61.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-out estimation of variance components." *Econometrica*, 88(5): 1859–1898.

Koedel, Cory, Kata Mihaly, and Jonah E. Rockoff. 2015. "Value-added modeling: A review." *Economics of Education Review*, 47: 180–195.

Kraft, Matthew A. 2019. "Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies." *Journal of Human Resources*, 54(1): 1–36.

Lavy, Victor, and Rigissa Megalokonomou. 2024. "Alternative Measures of Teachers' Value Added and Impact on Short and Long-Term Outcomes: Evidence From Random Assignment." National Bureau of Economic Research.

Liu, Jing, and Susanna Loeb. 2021. "Engaging teachers: Measuring the impact of teachers on student attendance in secondary school." *Journal of Human Resources*, 56(2): 343–379.

Marcus, Jan, Simon Reif, Amelie Wuppermann, and Amélie Rouche. 2020. "Increased instruction time and stress-related health problems among school children." *Journal of Health Economics*, 70: 102256.

Montonen, Jerry. 2025. "Shaping Young Minds: The Effects of Teachers on Socio-Emotional Skills." Unpublished manuscript.

Persson, Petra, and Maya Rossin-Slater. 2018. "Family ruptures, stress, and the mental health of the next generation." *American Economic Review*, 108(4-5): 1214–1252.

Petek, Nathan, and Nolan G. Pope. 2023. "The Multidimensional Impact of Teachers on Students." *Journal of Political Economy*, 131(4): 1057–1107. Publisher: The University of Chicago Press.

Rockoff, Jonah E. 2004. "The impact of individual teachers on student achievement: Evidence from panel data." *American Economic Review*, 94(2): 247–252.

Rose, Evan K., Jonathan Schellenberg, and Yotam Shem-Tov. 2025. "The Effects of Teacher Quality on Adult Criminal Justice Contact." *Econometrica*. Forthcoming.

Rossin-Slater, Maya, Molly Schnell, Hannes Schwandt, Sam Trejo, and Lindsey Uniat. 2020. "Local exposure to school shootings and youth antidepressant use." *Proceedings of the National Academy of Sciences*, 117(38): 23484–23489.

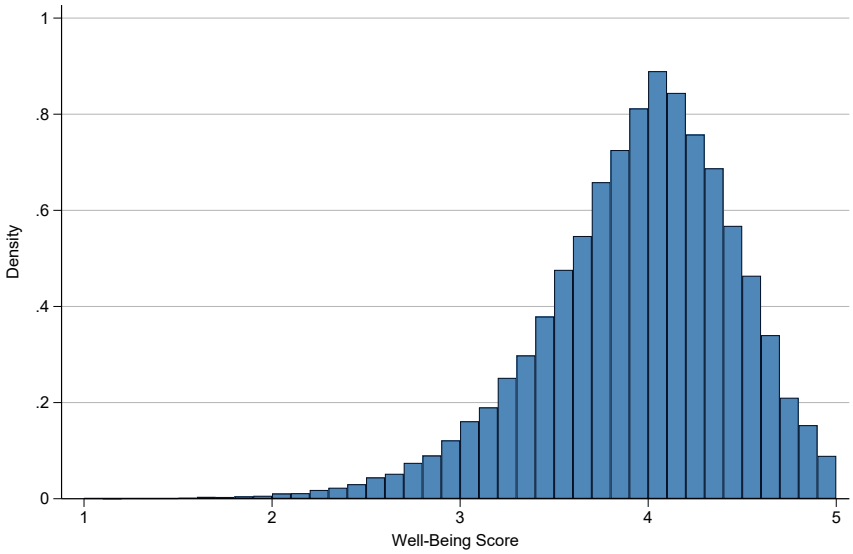
Schaller, Jessamyn, and Mariana Zerpa. 2019. "Short-run effects of parental job loss on child health." *American Journal of Health Economics*, 5(1): 8–41.

Statistics Denmark. 2024. "FORLOB10: From Primary and Lower Secondary School to Further Education." *Statistics Denmark StatBank*.

WHO. 2025. "Mental Health of Adolescents." *World Health Organization Fact Sheet*.

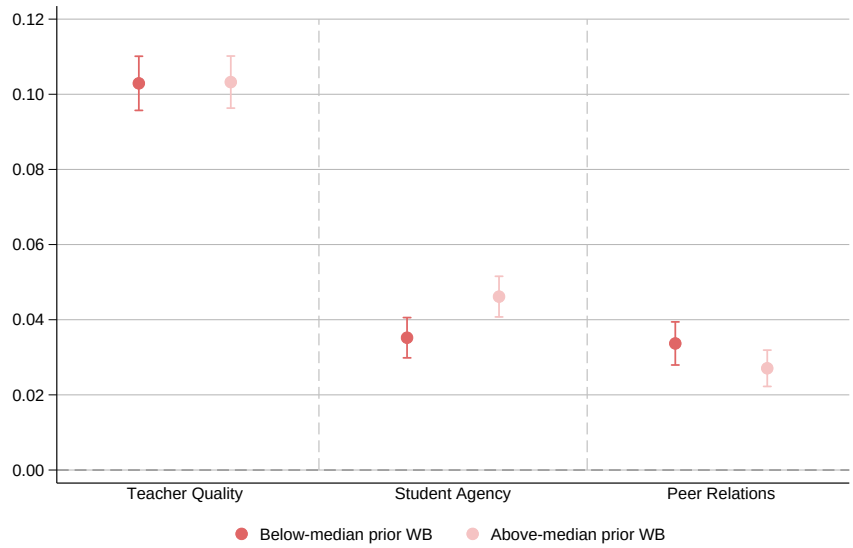
Appendix A: Additional Figures and Tables

Figure A.1: Distribution of Raw Well-Being Score



Notes: This figure shows the distribution of the raw well-being score calculated for each student in a given year. To construct the raw well-being score, we follow the definitions by the Danish Ministry of Education (2016) of academic and social well-being and compute the within-individual average of the answers provided.

Figure A.2: Effects of Well-Being Value-Added on School Experience Indices, by Prior Well-Being

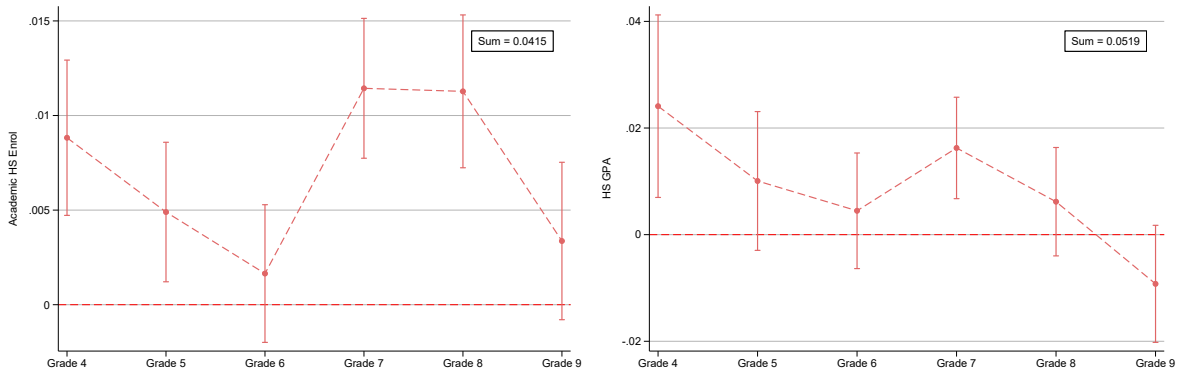


Notes: This figure reports estimates of the effects of well-being value-added on three indices constructed from student well-being survey items not used in our well-being measure: teacher quality, student agency, and peer relations (see Appendix B.3). Estimates are shown separately for students with below- and above-median well-being in the previous year. Each coefficient is from a separate regression of the standardized index on well-being value-added, conditional on the baseline control vector used in the value-added estimation. Error bars indicate 95 percent confidence intervals, with standard errors clustered at the student and classroom level. We test equality of effects across groups by interacting well-being value-added with an indicator for above-median prior well-being. The difference is statistically significant only for student agency ($p < 0.001$).

Figure A.3: Effects of Well-Being Value-Added on Academic Outcomes by Grade

(a) Academic HS Enrollment

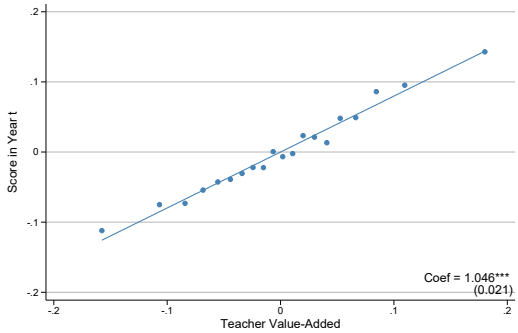
(b) High School GPA



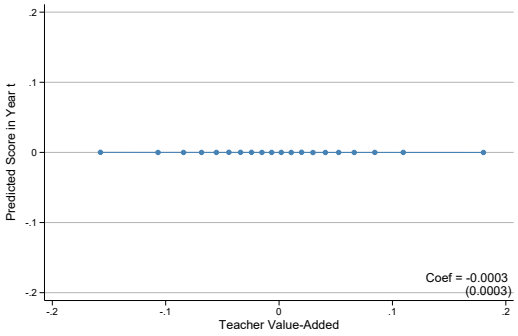
Notes: This figure plots the effect of well-being value-added on two academic outcomes by the grade level of the student: an indicator for enrolling in academic high school (Panel A) and high school GPA, standardized by graduation year and academic track (Panel B). The plotted coefficients and 95 percent confidence intervals are from regressions of each outcome on well-being value-added and the baseline control vector used in the value-added estimation, estimated separately for each grade. Standard errors are clustered at the classroom level. Each panel also reports the sum of coefficients across all grades, which under the assumption of no diminishing returns provides an upper bound on the cumulative effect of having a one standard deviation higher teacher VA in every grade from 4 to 9.

Figure A.4: Effects of Test Score Teacher Value-Added on Actual and Predicted Scores

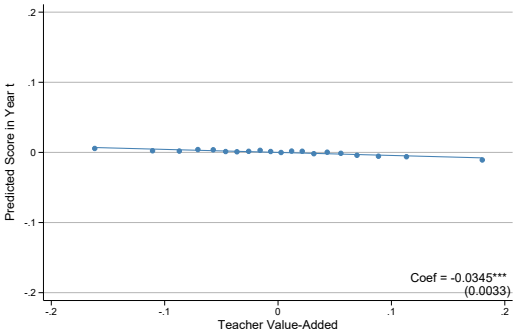
(a) Actual Score



(b) Predicted Score Using Maternal Income



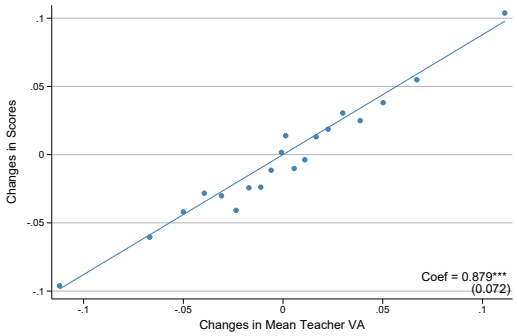
(c) Predicted Score Using Twice-Lagged Test Score



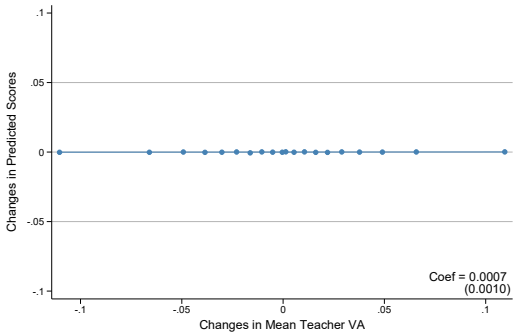
Notes: This figure assesses the validity of the test score value-added estimates. Panel A plots contemporaneous residualized test scores against teacher test score value-added. Under the stationarity assumption, the coefficient should equal one. Panels B and C assess whether students sort to teachers based on observable characteristics excluded from the value-added model. Panel B plots predicted outcomes based on maternal income; Panel C plots predicted outcomes based on twice-lagged test scores. Predicted outcomes are obtained by residualizing the predictor with respect to the baseline control vector and regressing residualized outcomes on this residualized predictor. In all panels, teacher value-added estimates are divided into 20 equal-sized bins and the figures plot the mean outcome within each bin against the mean value-added estimate. Solid lines show best linear fits estimated on the underlying microdata using OLS. Standard errors are clustered at the school-cohort level.

Figure A.5: Effects of Changes in Teaching Staff on Student Outcomes Across Cohorts: Test Score Value-Added

(a) Actual Outcomes



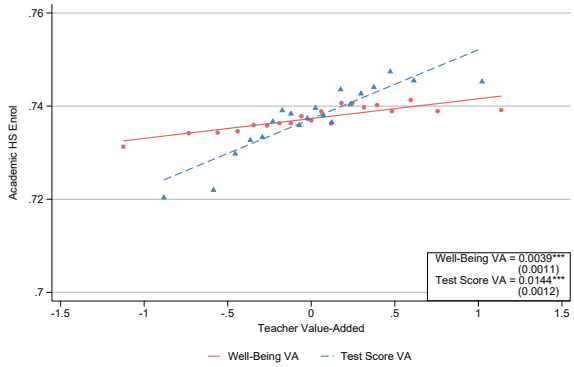
(b) Predicted Outcomes (Placebo)



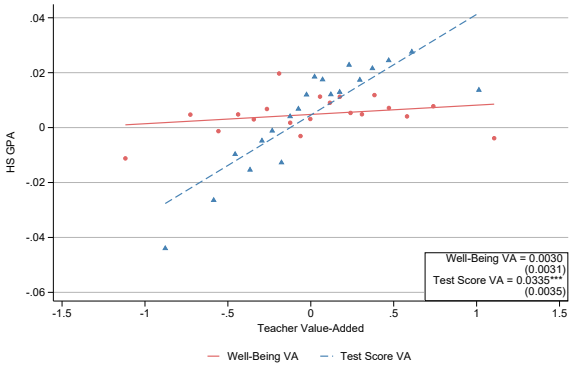
Notes: This figure plots changes in cohort-level mean outcomes against changes in mean test score value-added within school–grade–subject cells, exploiting staffing changes across cohorts. Panel A uses actual outcomes, while Panel B uses predicted outcomes based on maternal income as a placebo check. Teacher value-added is constructed using leave-two-year-out estimates. Lines show best linear fits estimated on the underlying microdata.

Figure A.6: Effects of Teacher Value-Added on Academic Outcomes (Multivariate)

(a) Academic HS Enrollment

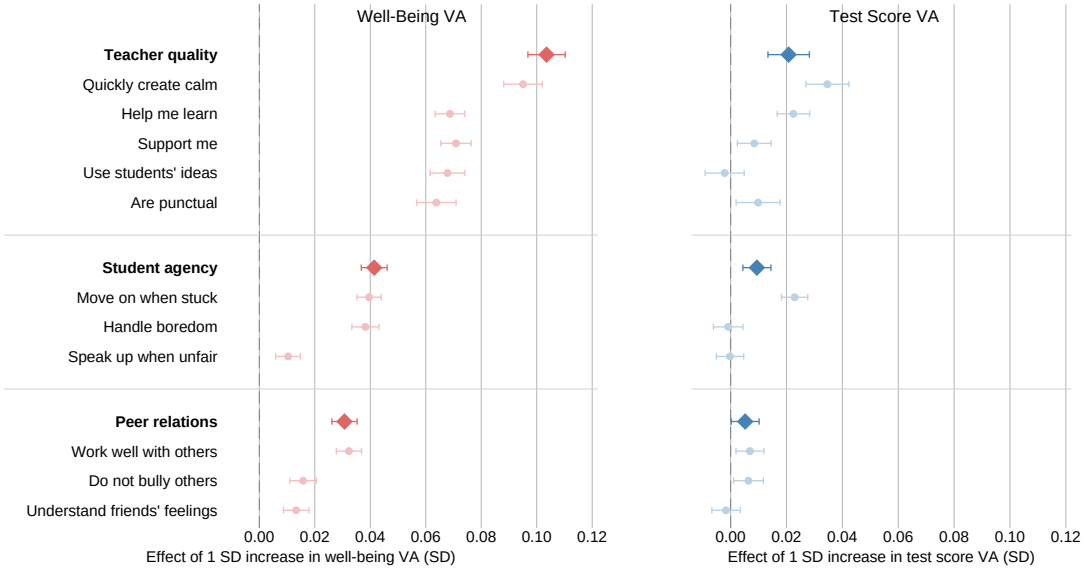


(b) High School GPA



Notes: This figure plots the partial relationship between teachers' value-added in middle school and students' academic outcomes. In particular, we focus on two outcomes: (i) indicator that takes value one if the student enrolled in academic high school and (ii) high school GPA, standardized by graduation year and academic track. The figure displays two series corresponding to teacher well-being value-added and test score value-added, respectively; both series are based on a single multivariate specification in which the outcome is regressed simultaneously on both value-added measures. The unit of observation is a student–subject–year. Both the outcome and the teacher value-added measure are residualized with respect to the baseline control vector used in the value-added estimation. Teacher value-added estimates are divided into 20 equal-sized bins, and the figure plots the mean residualized outcome within each bin against the mean value-added estimate, with the unconditional means added back. We also plot the best linear fit estimated before binning the data and report its slope coefficient and standard error, clustered at the student and classroom level.

Figure A.7: Effects of Well-Being and Test Score Value-Added on School Experience



Notes: This figure reports estimates of the effects of well-being value-added (left panel) and test score value-added (right panel) on three school experience indices (diamonds) and on the survey items that make up each index (circles): teacher quality, student agency, and peer relations. The items are not used in our well-being measure. Item labels are abbreviated; the full wording is listed in Appendix B.3. Indices and items are measured in the same year as the teacher assignment and standardized by grade and year. For each outcome, both coefficients come from a single regression of the outcome on the two value-added measures jointly, conditional on the baseline control vector used in the value-added estimation. Error bars indicate 95 percent confidence intervals, with standard errors clustered at the student and classroom level.

Table A.1: Selection into Well-Being Survey Response: Student Characteristics

	Survey Response
Female	0.0092***
	(0.0005)
Migration background	−0.0049***
	(0.0009)
Mother university degree	0.0098***
	(0.0005)
Father university degree	0.0092***
	(0.0006)
Mean	0.743
Observations	1,779,891

Notes: This table reports coefficients from a regression of an indicator for responding to the well-being survey on student characteristics, including school-by-grade-by-year fixed effects. The sample includes all students enrolled in grades 4 through 9, with one observation per student and year. Mean reports the response rate in this sample. Standard errors, clustered at the student and classroom level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.2: Selection into Well-Being Survey Response: Teacher Well-Being Value-Added

	Survey Response
Well-being VA	0.0032***
	(0.0009)
Mean	0.908
Observations	1,985,802

Notes: This table reports coefficients from a regression of an indicator for responding to the well-being survey on the teacher's standardized leave-year-out well-being value-added, conditional on the baseline control vector used in the value-added estimation. The sample includes students in classes in the value-added sample, with one observation per student, subject, and year. Mean reports the response rate in this sample. Standard errors, clustered at the student and classroom level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.3: Standard Deviation of Teacher Effects: Alternative Estimators

	Danish	Math	Pooled
A. Well-Being			
Within-year covariance	0.120	0.101	0.109
Lag-1 autocovariance	0.134	0.094	0.108
Kline et al. (2020); Rose et al. (2025)	0.111	0.075	0.093
B. Test Scores			
Within-year covariance	0.151	0.202	0.181
Lag-1 autocovariance	0.129	0.144	0.132
Kline et al. (2020); Rose et al. (2025)	0.133	0.153	0.138

Notes: This table compares estimates of the standard deviation of teacher effects on well-being (Panel A) and test scores (Panel B) across estimators. Within-year covariance is our main estimate (Row 4 of Table 3), computed as the square root of the covariance of mean residuals across pairs of classrooms taught by the same teacher in the same year. Lag-1 autocovariance (Row 5 of Table 3) is computed from the covariance of classroom mean residuals of the same teacher in adjacent years. The last row in each panel applies the unbiased variance estimator of Kline, Saggio and Sølvsten (2020) and Rose, Schellenberg and Shem-Tov (2025), computed from products of teacher-level mean residuals across all pairs of different years. These estimates are computed on teacher-years with both well-being and test score outcomes.

Table A.4: Effects of Well-Being Value-Added on Psychiatric Diagnoses

	Any Diagnosis	Anxiety	Depression
Well-Being VA	-0.0016^{**}	-0.0002	-0.0008^{***}
	(0.0007)	(0.0005)	(0.0003)
Observations	1,174,858	1,174,858	1,174,858
Mean	0.080	0.030	0.012
R-squared	0.032	0.015	0.006

Notes: This table reports the estimates plotted in Figure 4. Each column reports the coefficient from a regression of an indicator for receiving the psychiatric diagnosis indicated in the column header between ages 16 and 18 on standardized leave-year-out well-being value-added, conditional on the baseline control vector used in the value-added estimation. The unit of observation is a student–subject–year. Mean reports the mean of the dependent variable in the estimation sample. Standard errors, clustered at the student and classroom level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.5: Quasi-Experimental Estimates: Effects of Test Score Value-Added on Long-Run Outcomes

	HS Well-Being	Psyc. Visit	Psyc. Diag.	HS Enrol	HS GPA
Test Score VA	0.0291***	-0.0024	-0.0038	0.0258***	0.0321***
	(0.0097)	(0.0035)	(0.0026)	(0.0029)	(0.0087)
Observations	20,146	20,402	20,613	27,283	21,509
R-squared	0.001	0.000	0.000	0.023	0.001

Notes: This table reports estimates of the effect of test score value-added on long-run outcomes using quasi-experimental variation in teacher quality generated by staffing changes across adjacent cohorts within school–grade cells. The unit of observation is a school–grade–cohort. Each column reports the coefficient from a regression of the change in the school–grade–cohort mean of the outcome indicated in the column header on the change in the corresponding mean of teacher test score value-added, and can be read as the effect of a one standard deviation increase in test score value-added. Teacher value-added is estimated using leave-two-year-out measures of teacher quality. All specifications include year fixed effects and are weighted by cohort size. Standard errors, clustered at the school–cohort level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.6: Heterogeneous Effects of Well-Being Value-Added

	Prior Test Score		Prior Well-Being		Gender		Predicted Risk	
	Low	High	Low	High	Boys	Girls	Low	High
Panel A. High School Well-Being								
WB VA	0.0440***	0.0443***	0.0424***	0.0455***	0.0460***	0.0422***	0.0493***	0.0423***
	(0.0050)	(0.0038)	(0.0047)	(0.0039)	(0.0047)	(0.0044)	(0.0049)	(0.0048)
Mean	−0.078	0.094	−0.275	0.271	0.138	−0.060	0.178	−0.115
p-value	0.923		0.056		0.023		0.008	
Panel B. Mental Health Care Utilization								
WB VA	−0.0057***	−0.0081***	−0.0067***	−0.0074***	−0.0054***	−0.0088***	−0.0055***	−0.0081***
	(0.0013)	(0.0012)	(0.0013)	(0.0011)	(0.0010)	(0.0015)	(0.0010)	(0.0016)
Mean	0.136	0.129	0.165	0.100	0.080	0.187	0.075	0.190
p-value	<0.001		0.002		<0.001		0.209	
Panel C. Psychiatric Diagnosis								
WB VA	0.0003	−0.0031***	−0.0003	−0.0028***	−0.0011	−0.0020*	−0.0011	−0.0013
	(0.0011)	(0.0009)	(0.0011)	(0.0008)	(0.0009)	(0.0012)	(0.0008)	(0.0012)
Mean	0.087	0.074	0.106	0.054	0.058	0.103	0.048	0.111
p-value	<0.001		<0.001		0.679		0.091	

Notes: This table reports the estimates plotted in Figure 9. Each panel reports coefficients from split-sample regressions of the outcome indicated in the panel header on standardized leave-year-out well-being value-added, estimated separately for the indicated subgroups and conditional on the baseline control vector used in the value-added estimation. Prior test scores and prior well-being are split at the median. Predicted risk is split at the median of the predicted probability of mental health care utilization based on baseline characteristics. Mean reports the mean of the dependent variable in each subgroup (high school well-being in standard deviations). The p-value tests equality of the effects in levels across subgroups, estimated from a pooled regression interacting teacher value-added with the subgroup indicator. Standard errors, clustered at the student and classroom level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.7: Joint Effects of Teacher Value-Added on Psychiatric Diagnoses

	Any Diagnosis	Anxiety	Depression
Well-Being VA	−0.0009	0.0004	−0.0007**
	(0.0008)	(0.0005)	(0.0003)
Test Score VA	−0.0017*	−0.0017***	0.0002
	(0.0009)	(0.0006)	(0.0003)
Observations	890,009	890,009	890,009
Mean	0.079	0.030	0.011
R-squared	0.032	0.014	0.006

Notes: This table reports the estimates plotted in Figure 14. Each column reports coefficients from a single regression of an indicator for receiving the psychiatric diagnosis indicated in the column header between ages 16 and 18 on standardized leave-year-out well-being and test score value-added jointly, conditional on the baseline control vector used in the value-added estimation. The sample is restricted to observations with both value-added measures. The unit of observation is a student–subject–year. Mean reports the mean of the dependent variable in the estimation sample. Standard errors, clustered at the student and classroom level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix B: Data Appendix

B.1 Well-Being Measure

The well-being measure is constructed from the following survey items:

Academic Well-Being

1. What do your teachers think about your progress in school?
2. Are you succeeding in learning what you want to learn at school?
3. How often can you find a solution to problems if you just try hard enough?
4. How often can you accomplish what you set out to do?
5. Can you concentrate during lessons?
6. I am doing well academically at school.
7. I am making good academic progress at school.
8. If I get distracted during class, I can quickly concentrate again.

Social Well-Being

1. Are you happy with your school?
2. Are you happy with your class?
3. Do you feel lonely?
4. I feel like I belong at my school.
5. I like breaks at school.
6. Most of the students in my class are friendly and helpful.
7. Other students accept me as I am.
8. Have you been bullied this school year?
9. Are you afraid of being laughed at?
10. How often do you feel safe at school?

B.2 Mental Health Outcomes

Mental Health Care Utilization Mental health care utilization is measured using the *Health Insurance Register*, which records all reimbursed services provided by private clinics under national health insurance. We define an indicator equal to one if the individual has at least one of the following contacts during the relevant age window:

- *Psychological counseling:* Services provided by private psychologists, identified using physician specialty code 63. Mental health services provided by general practition-

ers, identified using physician specialty code 80 in combination with reimbursement codes for mental health diagnostic tests and counseling (802147–802149, 804003, 804021–804027, 804050, 804063, 804106, 804116, 806000, 806100, 806101).

- *Private psychiatric visits:* Visits to private practicing psychiatrists and child and adolescent psychiatrists, identified using physician specialty codes 24 and 26.

The Health Insurance Register does not record diagnoses, so we do not observe diagnoses made in private practice.

Psychiatric Diagnoses Psychiatric diagnoses are identified from hospital records. We combine the *Psychiatric Central Research Register*, which covers contacts with psychiatric departments and is no longer updated after 2019, with the *National Patient Register*, which covers all contacts with public and private hospitals. We define an indicator for any psychiatric diagnosis based on ICD-10 codes starting with the letter `F`, recorded in any inpatient, outpatient, or emergency room contact.

Specific Diagnostic Categories We further construct indicators for two specific types of psychiatric diagnoses:

- *Depression:* Defined using ICD-10 codes `F32.0-F33.9`
- *Anxiety disorders:* Defined using ICD-10 codes `F40.0-F48.9`

All diagnosis indicators are coded such that an individual is classified as having a given diagnosis if they receive at least one corresponding hospital-based diagnosis during the relevant age window. In the main analysis, mental health outcomes are measured between ages 16 and 18, and the sample is restricted to students observed at all ages in this window.

B.3 School Experience Indices

The school experience indices are constructed from survey items that are not included in the well-being measure. Each index is the average of the corresponding answers, coded so that higher values indicate more positive experiences.

Teacher Quality

1. If there is noise in the class, teachers can quickly create calm.
2. Do your teachers show up for class punctually?

3. Do your teachers help you learn in ways that work well?
4. The teachers are good at supporting me and helping me at school when I need it.
5. Teachers ensure that students' ideas are used in teaching.

Student Agency

1. I speak my mind when I think something is unfair.
2. If I am bored in class, I can do something to make it better.
3. If something is too difficult for me in class, I can do something myself to move on.

Peer Relations

1. I try to understand my friends when they are sad or angry.
2. I am good at working with others.
3. Have you bullied anyone at school this school year? (reverse coded)

B.4 High School Well-Being Measure

The high school well-being measure used as the main outcome in the analysis is the average of the Academic Well-Being and Social Well-Being sub-indices, which are the closest equivalent to the middle school well-being measure. The Ministry of Education constructs three additional sub-indices from the same survey—Learning Environment, Pressure & Concerns, and Bullying—which we use to disaggregate the long-run effects in Section 7. All sub-indices are coded so that higher values indicate more positive experiences. All five sub-indices are listed below.

Academic Well-Being

1. I am doing well academically at school.
2. I actively participate during lessons.
3. I submit my assignments on time.
4. I keep working until tasks are completed.
5. I am prepared for lessons.
6. I like learning new things.
7. I am motivated by the teaching.
8. I do my schoolwork because I am interested in it.
9. I am making good academic progress at school.
10. If I get distracted during class, I can quickly concentrate again.
11. If something is difficult for me in class, I can take action myself to move forward.

12. How often can you accomplish what you set out to do?
13. How often can you find a solution to problems if you just try hard enough?
14. Are you succeeding in learning what you want to learn at school?

Social Well-Being

1. Are you happy with your class?
2. I am good at working with others.
3. I feel like I belong at my school.
4. I support and help my classmates.
5. I get along well with my classmates.
6. The atmosphere in my class is characterized by understanding and respect for each other.
7. In my class, it is good to be active during lessons.
8. I can get help and support from my classmates.
9. I wish I attended a different school.
10. I am in contact with classmates outside of school hours.
11. I feel bad at school.
12. Do you feel lonely at school?

Learning Environment

1. I get feedback from teachers that I can use to improve in my subjects.
2. I have influence on teaching.
3. Teachers ensure that students' ideas are used in teaching.
4. Teaching motivates me to learn new things.
5. The teachers provide academic help when I need it.
6. The teachers respect me.
7. Teachers coordinate times for assignment submissions.
8. There are good opportunities at school to get support and guidance.

Pressure & Concerns

1. I remain calm in stressful situations.
2. How often do you feel pressured at school?
3. How often do you feel pressured because of homework?
4. How often do you feel pressured because of grades?
5. How often do you feel pressured because of your own demands and expectations?

Bullying

1. Been called nasty names, made fun of, or teased in a hurtful way?
2. Been threatened to do things I didn't want to do?
3. Been subjected to unwanted psychological or physical attention?
4. Digital bullying?
5. Been subjected to unwanted sexual attention?
6. Witnessed students or staff being pressured to do things they didn't want to do?