Rege, Mari et al.

Working Paper

The Impact of Teaching Coping Skills in Schools on Youth Mental Health and Academic Achievement: Evidence from a Field Experiment

CESifo Working Paper, No. 11742

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Rege, Mari et al. (2025) : The Impact of Teaching Coping Skills in Schools on Youth Mental Health and Academic Achievement: Evidence from a Field Experiment, CESifo Working Paper, No. 11742, CESifo GmbH, Munich

This Version is available at:
https://hdl.handle.net/10419/316856

CESifo WORKING PAPERS

11742
2025

March 2025

The Impact of Teaching Coping Skills in Schools on Youth Mental Health and Academic Achievement: Evidence from a Field Experiment

Mari Rege, Edvin Bru, Ingeborg F. Solli, Maximiliaan W. P. T. Thijssen, Kjersti B. Tharaldsen, Lene Vestad, Sigrun K. Ertesvåg, Terje Ogden, Paul N. Stallard

CESifo

Impressum:

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

CESifo *Working Paper No. 11742*

The Impact of Teaching Coping Skills in Schools on Youth Mental Health and Academic Achievement: Evidence from a Field Experiment

Abstract

Effective and scalable strategies for promoting youth mental health are urgently needed. We conducted a randomized controlled trial to assess the impact of school-based, teacher-led coping skills instruction on youth mental health and academic achievement. The trial included 84 classes and 1,879 ninth-grade students (ages 14–15) in Norway. Findings indicate improved mental well-being and reduced emotional distress at a one-year follow-up, particularly among students with low baseline well-being. The intervention also enhanced academic motivation and increased the likelihood of choosing an academic high school track. Additionally, it had a positive effect on math performance among students with initially low academic motivation, but no significant effects on performance in English or Norwegian.

JEL-Codes: I200, J000.

Keywords: social and emotional skills, preventive mental health policy, school-based intervention, teacher led intervention, education policy.

*Mari Rege**
University of Stavanger, Business School,
Department of Economics and Finance
Stavanger / Norway
mari.rege@uis.no
https://orcid.org/0000-0002-9574-5851

Edvin Bru
Norwegian Centre for Learning Environment
and Behavioral Research in Education
University of Stavanger / Norway
edvin.bru@uis.no
https://orcid.org/0000-0003-2004-6957

Ingeborg F. Solli
Department of Economics and Finance
University of Stavanger / Norway
ingeborg.f.solli@uis.no
https://orcid.org/0000-0002-9016-9920

Maximiliaan W. P. T. Thijssen
Knowledge Center for Education
University of Stavanger / Norway
maximiliaan.thijssen@uis.no
https://orcid.org/0000-0001-8622-3106

Kjersti B. Tharaldsen
Norwegian Centre for Learning Environment
and Behavioral Research in Education
University of Stavanger / Norway
kjersti.b.tharaldsen@uis.no
https://orcid.org/0000-0003-1536-730X

Lene Vestad
Norwegian Centre for Learning Environment and Behavioral Research in Education
University of Stavanger / Norway
lene.vestad@uis.no
https://orcid.org/0000-0002-6558-0929

Sigrun K. Ertesvåg
Norwegian Centre for Learning Environment and Behavioral Research in Education
University of Stavanger / Norway
sigrun.ertesvag@uis.no
https://orcid.org/0000-0001-7804-8571

Terje Ogden
The Norwegian Center for Child Behavioral Development, Oslo / Norway
https://orcid.org/0000-0002-2353-0798

Paul N. Stallard
Department for Health, University of Bath / United Kingdom
P.Stallard@bath.ac.uk
https://orcid.org/0000-0001-8046-0784

*corresponding author

This project was funded by the Research Council of Norway (grant number 325398). We thank all the teachers, students, and school administrators who participated. We also extend our gratitude to Hilde Ness Sandvold for her role as the administrative project coordinator. The RCT was pre-registered in the AEA RCT Registry (AEARCTR0009429).

The Impact of Teaching Coping Skills in Schools on Youth Mental Health and Academic Achievement: Evidence From a Field Experiment

There is a global crisis in child mental health with an estimated 166 million adolescents (aged 10–19 years) worldwide living with a mental health disorder (United Nations Children's Fund, 2021). Depression and anxiety are the most prevalent mental health problems during adolescence (Merikangas et al., 2010). According to a recent meta-analysis the prevalence of self-reported clinically elevated (i.e. moderate to severe) symptoms of anxiety and depression among youth is 21 and 25 percent, respectively (Racine et al., 2021).

The personal and economic impacts of child mental problems are considerable. Childhood depression and anxiety are strong predictors of poor psychosocial and interpersonal functioning in both childhood and adulthood, lower educational attendance and attainment, reduced workforce participation, increased reliance on welfare, and poorer mental and physical health (see, e.g., Castelpietra et al., 2022; Currie, 2024; Currie et al., 2010). Moreover, mental health problems often seem to be transferred to the next generation (see, e.g., Bütikofer et al., 2024; Johnston et al., 2013).

Despite the significant personal and economic consequences, only a minority of children with mental health issues receive adequate treatment (Costello et al., 2014; Cuddy & Currie, 2020a, 2020b). Surveys in the United States and United Kingdom indicate that less than 50 percent of children with mental health disorders access mental health services, with those experiencing emotional disorders like anxiety or depression being the least likely to receive support (Costello et al., 2014; Ford et al., 2003). The capacity of specialized child mental health services is limited and cannot keep up with the growing demand for treatment. Additional barriers, such as limited mental health awareness, stigma, embarrassment, and concerns over confidentiality, further hinder access (Radez et al., 2021).

The significant "treatment gap" underscores an urgent need for effective, scalable prevention strategies. Schools provide a non-stigmatizing and accessible environment to

reach all children, making them ideal for mental health support initiatives (Fazel et al., 2014). Teachers already play a role in fostering social and emotional skills, many of which are closely tied to mental well-being. Strengthening and systematizing this role, by equipping teachers with targeted training and a knowledge-based curriculum, may be an effective and scalable approach to strengthen youth mental health.

In this paper, we investigate the effects of teachers teaching "coping skills" on youth mental health. We define coping skills as the subset of social and emotional competencies essential for managing challenging situations, such as academic demands, challenging social relations, loneliness, and negative emotions. This set includes relationship skills, emotion regulation, problem-solving, and a growth mindset. An extensive body of theoretical and empirical literature suggests that interventions designed to strengthen coping skills can mitigate the adverse effects of stress while promoting emotional well-being and mental health (Cipriano et al., 2023; Lazarus, 2006). However, to develop scalable prevention strategies, more evidence is needed on the effects of teacher-led coping skills instruction in schools on mental health.

Our intervention aimed to foster coping skills among students in middle school (grade 8 to 10) in Norway through a curriculum called ROBUST. This curriculum comprised 25 interactive and engaging sessions, each lasting 45-60 minutes, delivered by teachers during regular school hours throughout 9th-grade. The sessions were facilitated by one or two of the students' regular teachers, who received specialized training prior to implementing the curriculum.

We examined the effects of the ROBUST intervention using a two-armed randomized controlled trial (RCT) involving 84 classes and 1,879 students in 9th-grade (14-15 years old). Classes were randomly assigned within schools, with one group receiving the ROBUST curriculum and the other following a business-as-usual approach. To measure students' mental health outcomes, we used the Warwick-Edinburgh Mental Well-being Scale (WEMWBS: Tennant et al., 2007) and the 10-item version of the Hopkins Symptoms

Checklist for emotional distress (HSCL-10: Strand et al., 2003). Assessments took place at three intervals: before the intervention, immediately following the intervention, and at a one-year follow-up, marking the end of middle school (10th-grade). Additionally, we matched this data with school administrative records on academic performance at both middle school entry and graduation, as well as registry data from Statistics Norway on student and family characteristics.

We found that the intervention had a positive effect on mental well-being and reduced emotional distress at the one-year follow-up, with effect sizes around 10-12 percent of a standard deviation. Investigating differential treatment effects, the effects are particularly large among the tertile with lowest mental well-being at baseline, increasing mental well-being by 21 percent of a standard deviation and reducing emotional distress by 14 percent of a standard deviation. Interestingly, while the immediate post-intervention effect sizes were in the expected direction, they were smaller in magnitude and for emotional distress not statistically significant. This may be because it takes time to practice and apply coping skills before they have a measurable impact on mental health. Additionally, the follow-up assessment was conducted in the final semester of middle school, a period when students may experience increased stress and pressure to achieve good grades, which are important for high school opportunities. In this context, the skills learned from ROBUST may have become particularly relevant. Consistent with this conjecture, our mechanism investigation finds that a substantial part of the effect on mental well-being can be explained by improved relationship skills.

We also investigated how the intervention impacted motivation for school and academic grades at the end of middle school. Improved mental health can enhance a student's motivation and capacity to take advantage of the learning opportunities in school and thereby improve grade outcomes (Bas, 2021; Pekrun, 2024). Additionally, several of the coping skills taught in ROBUST, such as problem-solving and growth mindset, can have a direct impact on academic motivation and achievement (Stallard, 2010; Yeager

et al., 2014). Consistent with these conjectures, we found that the intervention increased motivation for school at the one-year follow up. Moreover, it had a positive effect on performance in math, especially for students with low baseline motivation for school, but no effect on grades in Norwegian or English.

Another indicator of school outcomes is students' choice of high school program. In the Norwegian high school system, students can select either academic programs, which prepare them for higher education, or vocational programs, which lead directly to employment. The intervention increased the likelihood of students choosing the academic track over the vocational track by 4.2 percentage points.

Our paper contributes to the literature in two key ways. First, there is a growing interest in mental health as a critical determinant of human capital development (Currie, 2024). The evidence discussed above show that youth mental health is a strong predictor of educational attainment and workforce participation in adulthood, suggesting that preventive youth mental health interventions can serve as effective investments in human capital. We contribute to this literature, by examining the effectiveness of school-based, teacher-led coping skills instruction on youth's mental health, and how this investment affects academic achievement.

Second, there are several papers documenting a growing importance of social and emotional skills for labor market outcomes (Deming, 2017; Edin et al., 2022; Lindqvist & Vestman, 2011; Woessmann, 2024). Moreover, several studies demonstrate that interventions fostering these skills may lead to lasting changes in academic and labor market outcomes, mental health and well-being (e.g., Alan & Ertac, 2018; Alan et al., 2019; Berger et al., 2025; Cipriano et al., 2023; Heckman & Kautz, 2012; Heckman et al., 2006; Sorrenti et al., 2024). However, more research is needed for us to better understand how to effectively invest in social an emotional skills (Deming, 2022). We contribute to this gap in the literature by investigating the impacts of teachers teaching coping skills in middle school. As far as we know, we are the first study investigating impacts of a universally

provided, teacher led SEL-intervention on middle school students' mental health and grade outcomes (see recent scoping review of interventions targeting teens in Métais et al., 2024).[1]

Conceptual Framework

The ROBUST curriculum supports students development in social and emotional skills particularly important for coping with common stressors in the school context, such as academic demands, challenging social relations, loneliness, and negative emotions. Specifically, it targets four key skills—relationship skills, emotion regulation, problem-solving, and growth mindset. We refer to these as coping skills.

The design and selection of the four coping skills are based on Lazarus' stress and coping theory, which posits that emotional distress results from an imbalance between external stressors and available coping resources (Lazarus, 2006). A core tenet of Lazarus' theory is that enhancing coping skills can reduce the negative impact of these stressors, promoting better mental health.

Below, we outline how the ROBUST curriculum targets the four coping skills to help mitigate the adverse effects of stressors and, in turn, improve mental health:

Relationship Skills. The curriculum helps students develop skills to build and maintain supportive relationships, such as understanding others' perspectives and providing social and emotional support. Supportive relationships fulfill essential human needs by helping adolescents feel valued and fostering a sense of belonging (Ryan & Deci, 2020). In times of stress, these relationships offer social support that enhances functioning, problem-solving, and emotion regulation, benefiting mental health (Rueger et al., 2016; Thoits, 2011). Numerous studies demonstrate a positive association between social relationships and mental health (see, e.g., Rueger et al., 2016; Thoits, 2011).

---

[1] A few studies have examined the effects of teacher-led interventions that have a narrower psychological scope, compared to the broader approach of the ROBUST curriculum, which incorporates social skills training and promotes a growth mindset. While these psychological interventions, often rooted in cognitive-behavioral therapy, tend to be effective when targeted (for young people with risk factors or symptoms), they appear less effective in improving mental health outcomes when delivered universally (Werner-Seidler et al., 2021).

Emotion Regulation. The curriculum helps students strengthen their ability to understand and manage their emotions. Drawing on principles from cognitive behavioral theory, it deepens students' understanding of the connection between thoughts and feelings, guiding them to identify and replace negative automatic thoughts with positive, constructive ones (Albert, 2010). By learning to interpret situations more constructively, students can reduce emotional distress and negative emotions, improving their mental health. A substantial body of research has shown a strong link between emotion regulation skills and mental health (Compas et al., 2017). Moreover, the ability to regulate emotions related to schoolwork can enhance students' motivation and academic achievement (Pekrun, 2024).

To further support emotion regulation, the curriculum also helps students develop mindfulness (Bishop et al., 2004). Through breathing and visualization practices, students learn methods to become aware own perceptions, reduce psychological activation, and focus their attention. A recent meta-analysis suggests that these practices can enhance well-being and promote better mental health outcomes (Dunning et al., 2019).

Problem-Solving. The curriculum teaches students problem-solving strategies for managing controllable stressors, such as academic challenges and social conflicts. Students learn to break down challenges into manageable parts and tackle them step-by-step (Stallard, 2010). This component also includes planning and scheduling skills to help balance schoolwork with leisure activities, reducing perceived time pressure, along with strategies for self-regulated learning. By developing these problem-solving skills, students can apply adaptive strategies to manage heavy workloads and learning challenges, strengthening their sense of control over academic and other demands. Existing studies demonstrate that problem-solving skills are associated with improved mental health outcomes (Aldao et al., 2010).

Growth Mindset. The curriculum aims to foster students' belief that abilities can grow through effort, effective learning strategies, and support from others, as described by

Dweck and Yeager (2019) as a growth mindset. Building on protocols from psychology, the curriculum incorporates reading and writing exercises that emphasize: (1) the brain's capacity to grow and change; and (2) how hard work on challenging exercises enhances neural connections in the brain (Rege et al., 2021; Yeager et al., 2014). Research shows that students with a growth mindset see mistakes and setbacks as essential for learning rather than as signs of low ability. Instead of giving up, they put in more effort and try new approaches to overcome challenges (Yeager & Dweck, 2012). This adaptive approach to learning challenges may also support better mental health. Several studies demonstrate a positive association between growth mindset and mental-health outcomes (see meta-analysis in Burnette et al., 2020).

In summary, we hypothesize that the ROBUST curriculum will positively impact students' mental health by increasing their access to social support, improving emotion regulation skills, strengthening their sense of control over their environment, and fostering more constructive beliefs about learning:

Hypothesis 1: The ROBUST curriculum will have a positive impact on students' mental health.

Some students already thrive with good mental health. We hypothesize that the intervention will have a greater effect on mental health for those with poor baseline mental health compared to those who already have good mental health:

Hypothesis 2: The impacts of the ROBUST curriculum on students' mental health are larger for students who scored low on mental health at baseline compared to students who scored high.

Improved mental health can help students thrive and thereby enhancing their motivation and capacity to take advantage of the learning opportunities in school (Bas, 2021; Pekrun, 2024). Consequently, the ROBUST curriculum may positively impact

students' motivation and academic outcomes. Additionally, as discussed above, emotion regulation, problem-solving skills and a growth mindset, may have direct effects on students' motivation and school performance. This motivates our third hypothesis:

Hypothesis 3: The ROBUST curriculum will have a positive impact on students' academic motivation and school performance.

Some students are already highly motivated for school and perform well. We hypothesize that the intervention will have a greater impact on academic motivation and academic performance for those with low baseline academic motivation:

Hypothesis 4: The impacts of the ROBUST curriculum on students' academic motivation and school performance are larger for students who scored low on academic motivation at baseline compared to students who scored high.

Empirical Approach

Experimental Design

To investigate the impact of the ROBUST curriculum, we conducted an RCT, involving 84 classes and 1,879 students in 9th-grade (ages 14–15) in the school year of 2021/22. The experimental design is illustrated in Figure 1. Each school with more than one 9th-grade classroom constituted a randomization block. Schools with only one classroom were grouped into blocks based on geographic proximity. We randomized to control and treatment at the classroom level within each block. Students in treated classes received the ROBUST curriculum, whereas students in control classes continued with practice as usual.

We assessed all students at baseline (May 2021), post intervention (May 2022), and follow-up just before graduation (May 2023). The same test-battery was used in all the three surveys. Additionally, we matched the data to administrative records on grades and parental characteristics. Moreover, we conducted a process evaluation assessing the quality

of implementation.

Below we provide details on the intervention, measures, and empirical strategy. We will provide details on the process evaluation in the section on Implementation.

Intervention

The ROBUST curriculum was delivered weekly to 9th-grade students through 25 lessons, each lasting 45–60 minutes, during regular school hours throughout the school year. Teachers had the flexibility to determine when, within the week, to allocate time for the curriculum. In most cases, they incorporated the lessons into their subject-specific teaching, supplemented by time from the class's form period (homeroom period) and a subject known as "Education Choice."

The curriculum covered four coping skills: relationship skills, emotion regulation, problem-solving and growth mindset (see the Conceptual Framework for details). The first 15 lessons introduced each of the topics, whereas the final 10 lessons reinforced and expanded upon the key learning objectives, allowing students to revisit and deepen their understanding of the material covered initially. Teachers were also encouraged to incorporate elements from ROBUST into regular classroom instruction.

The curriculum was taught by one of the students' regular teachers,[2] who received specialized teacher professional development before implementation. To support delivery, the curriculum was accompanied by a comprehensive package of educational materials. This package included the ROBUST book, which provided engaging learning activities for the 25 lessons, and a series of ROBUST podcasts for further learning and inspiration. Additionally, teachers had access to digital classroom resources, including brief introductory videos for each lesson, a rap music video emphasizing coping skills covered in ROBUST, PowerPoint presentations for each lesson, audio resources for breathing exercises, a digital growth mindset program, and student worksheets.

[2] In two schools, teachers from the project team had to step in to do some of the teaching because of capacity constraints.

The teacher professional development consisted of a five-day digital training program. This training took place from August through January, prior to teaching the respective topics. University staff involved in developing the intervention served as professional trainers. Training covered the theoretical foundations and research underlying each of the four targeted coping skills, which were also summarized in the introductory chapters of the ROBUST book. Teachers also practiced implementing the curriculum's interactive learning activities. Each training day lasted six hours and included presentations, group discussions, activity practice, and open discussions for questions and feedback.

Measures

We assessed student outcomes across three domains: mental health, academic motivation and achievement, and coping skills.[3]

Mental Health

Mental health was measured by students' perceived mental well-being, emotional distress, and academic stress.

Mental Well-being. Mental well-being was measured using the 14-item Warwick-Edinburgh Mental Well-Being Scale (WEMWBS: Tennant et al., 2007). This scale captures both the emotional and functional dimensions of mental well-being. The emotional dimension reflects positive feelings, such as optimism, while the functional dimension relates to engagement in daily activities and a sense of purpose. Example items include, "I have been optimistic about the future" and "I have been interested in new things." The scale has a five-point response format: (1) "None of the time," (2) "Rarely," (3) "Some of the time," (4) "Often," and (5) "All of the time." The scale demonstrated high internal consistency, with a Cronbach's alpha of 0.95.

Emotional Distress. Emotional distress was measured using the 10-item version of the Hopkins Symptoms Checklist (HSCL-10: Strand et al., 2003). This scale assesses

[3] We provide a Pearson correlation matrix in the Appendix, Table A.1.

emotional distress through symptoms of anxiety (e.g., "Sudden fear for no reason") and depression (e.g., "Feeling that everything is a waste"), with four response options: (1) "Not bothered," (2) "A little bothered," (3) "Pretty much bothered," and (4) "Very much bothered." The scale demonstrated strong internal consistency, with a Cronbach's alpha of 0.93.

Academic Stress. Academic stress was assessed by using the five-item Academic Stress scale (Murberg & Bru, 2004), designed to evaluate stress specifically related to schoolwork. Sample items include statements such as "You think schoolwork has been too demanding" and "You have worried that you will not be able to do the schoolwork well enough." The scale has a five-point response format, ranging from (0) "No stress" to (5) "Very high stress." The scale exhibited high internal consistency, with a Cronbach's alpha of 0.93.

Academic Motivation and Achievement

We assess academic motivation through an established survey measure, academic achievement based on administrative grade records, and academic aspiration through students' self reported choice of high school track.

Academic Motivation. Academic motivation was assessed using measures of behavioral and emotional engagement based on Skinner et al. (2009). Behavioral engagement was assessed through students' efforts in academic tasks (e.g., "In class, I have worked as hard as I can") on a seven-item scale, while emotional engagement was measured through students' interest in their academic work (e.g., "I have enjoyed keeping up with my schoolwork") on a six-item scale. Both scales have the same six-point response format, with options ranging from (1) "Completely disagree" to (6) "Completely agree." We combined the two scales to form a comprehensive 13-item Academic Motivation scale, demonstrating good internal consistency with a Cronbach's alpha of 0.95.

Grades. Grades represent student achievement in the three core subjects: Norwegian, English, and math, collected at baseline (8th grade) and at the end of middle

school (10th-grade). Grades range from 1 (lowest) to 6 (highest). In Norwegian, students receive three separate grades, which we averaged to form a single composite grade.

Track. Track measures students' aspirations for high school. Prior to the follow-up survey conducted at the end of 10th-grade, the final year of middle school, students had recently submitted their high school applications. In the survey, we asked students which high school program they had applied for: an academic program, which prepares students for higher education, or a vocational program, which leads directly to employment. Based on this survey question, we created an indicator variable that assigns a value of 1 if a student applied to an academic program and 0 if they applied to a vocational program.

Coping Skills

We measured students' coping skills using established measures of relationship skills, problem solving skills, growth mindset, and emotion regulation. All scales have the a six-point response format, with options ranging from (1) "Completely disagree" to (6) "Completely agree."

Relationship Skills. Relationship skills were assessed using a seven-item sub-scale from the Strength and Difficulties Questionnaire (SDQ: Goodman, 1998) to measure prosocial behavior (e.g., "I often offer to help others"), along with a five-item scale evaluating students' perceived ability to initiate and establish social contact (e.g., "I get to know others easily") (Vestad et al., 2021). We combined the two scales to create a comprehensive 12-item measure of relationship skills. The combined measure had high internal consistency with a Cronbach's alpha of 0.94.

Problem Solving Skills. Problem solving skills were assessed using a slightly adapted version of the sub-scale Planning from the COPE Scale (Carver et al., 1989). The scale measures the extent to which students employed planning as a coping strategy in relation to their schoolwork (e.g., "I have tried to come up with a strategy for what I should do"). Internal consistency was high, with a Cronbach's alpha of 0.95.

Growth Mindset. Growth mindset was measured using a five-item sub-scale from the Self-Efficacy Formative Questionnaire (SFQ: Gaumer Erickson & Noonan, 2016). It assesses an individual's belief that abilities can improve with effort ("My abilities grow with the effort I put in"). Internal consistency was high with Cronbach's alpha of 0.96.

Emotion Regulation. Emotion regulation was assessed using the five-item Cognitive Reappraisal sub-scale from the Emotion Regulation Questionnaire for Children and Adolescents (ERQ-CA: Gullone & Taffe, 2012). This scale measures students' ability to re-frame their thinking more constructively in situations that may evoke negative emotions (e.g., "I control my feelings about things by changing the way I think"). The scale demonstrated high internal consistency, with a Cronbach's alpha of 0.93.

Student and Family Characteristics

We matched our data with registry data from Statistics Norway to obtain information on student and family characteristics. These include an indicator for gender (female), an indicator for immigrant background (defined as having a mother or father not born in Norway), and family income (calculated as the sum of the mother's and father's taxable earnings). Additionally, we included indicators for maternal and paternal education levels, specifically whether they had completed a bachelor's degree or attained education beyond a bachelor's degree, with the reference category being those who had not completed a bachelor's degree.

Empirical Strategy

Estimation Models

To investigate hypotheses 1 and 3, we estimate the program's treatment effects using an ordinary least squares (OLS) regression model, controlling for a comprehensive set of covariates. These include student and family characteristics, the baseline measures of the outcome variable, and fixed effects for randomization blocks. The OLS model is specified as follows:

$$y_{ijk} = \alpha_k + \beta T_{jk} + X_i'\gamma + \epsilon_{ijk}, \tag{1}$$

where y_{ijk} denotes the outcome of interest for student i in classroom j, within randomization block k. T_{jk} is a binary treatment indicator, equal to one if classroom j is in the treatment group and zero otherwise. X_i' represents a vector of student and family characteristics (gender, immigrant background, family income, mother and father education, and baseline measure of dependent variable), and ϵ_{ijk} is the error term. Including student-level covariates increases the model's statistical power by reducing variance in outcome measures, allowing for more precise estimates of program effects. In Figures A.1 to A.4 in the Appendix, we demonstrate that our findings are robust to excluding these covariates. To account for clustered treatment assignment, standard errors are clustered at the class level.

To investigate hypotheses 2 and 4, we divide the sample into three equally sized groups along two dimensions, representing low, moderate, and high levels of mental well-being or academic motivation at baseline. We then estimate differential treatment effects across these levels using the following interaction model:

$$y_{ijk} = \alpha_k + \beta_1 T_{jk} + \beta_2 (T_{jk} \cdot M_i) + \beta_3 (T_{jk} \cdot H_i) + \sigma_M M_i + \sigma_H H_i + X_i'\gamma + \epsilon_{ijk}, \qquad (2)$$

where M_i and H_i are indicators for moderate and high mental well-being/motivation at baseline, respectively.

Finally, to quantitatively evaluate the extent to which coping skills function as a mediating mechanism for the observed effects, we employ the decomposition framework developed by Heckman et al. (2013) and Heckman and Pinto (2015). This approach assumes that the outcome can be expressed as a linear combination of mediator variables and controls, and that program-induced changes in coping skills are statistically independent of unmeasured mechanisms, conditional on the controls. Under these assumptions, the method decomposes the total treatment effect into distinct components, allowing for the attribution of specific shares to different coping skills.

Standardization

As outlined in the section on Measures, each survey construct is measured using multiple items. To aggregate these items at the student level, we compute an equally weighted average following a standardized procedure. First, we translate categories on the Likert scales into numeric values.[4] Second, we standardize each item to a z-score with mean of 0 and standard deviation of 1. Third, we calculate the equally weighted average of these z-scores within each construct for each student. Finally, to ensure consistency and comparability, we re-standardize this average, once again setting the mean to 0 and the standard deviation to 1. As a result, the reported treatment effects are expressed in percentage standard deviations, representing the difference between the treatment and control groups. To maintain comparability of effect sizes, all grade measures are standardized to a mean of zero and a standard deviation of one.

Missing Values

To address missing values in control or outcome variables, we assume that the data are missing at random (Rubin, 1976) and employ multiple imputation using chained equations (van Buuren & Groothuis-Oudshoorn, 2011). This approach involves generating 100 imputations, each with 10 iterations, to ensure convergence and stability of the imputed values.

The imputation model includes a consistent set of variables across imputations: student and family characteristics, randomization-block fixed effects, the treatment indicator, baseline measures of all dependent variables, indicators for moderate and high mental well-being/motivation, and interactions between these indicators and the treatment indicator. Additionally, to improve imputation accuracy, each survey measure's imputation model includes the corresponding measure from other periods. For instance, missing values

[4] We assume equal intervals between categories on the Likert scales. While this assumption may be imprecise, since distances between categories are theoretically unrestricted (see, e.g., Bond & Lang, 2019), we address this concern with a robustness check following Bloem (2022) detailed in the Appendix (see Table A.4 and Figures A.31 to A.43).

for the Warwick scale in post-intervention (T2) are informed by observed values from baseline (T1) and follow-up (T3) assessments.

Following imputation, treatment effects are estimated using Rubin's pooling rules (Rubin, 1987), which appropriately combine estimates and standard errors across multiple imputed datasets to account for imputation uncertainty. Additional results based on alternative methods for handling missing data (e.g., full-information maximum likelihood and indicator method) are presented in Table A.3 in the Appendix as a robustness check.

Implementation

Norwegian Context

We conducted our experiment in Norwegian middle schools. In Norway, children begin school in August of the year they turn six years old and complete compulsory education after 10th-grade at age 16. This compulsory education includes seven years of elementary school followed by three years of middle school. Education is publicly funded, with public schools free of charge for all students. Most children (95 percent) attend a public school, and students generally enroll in the school designated for their residential area.

When enrolling in public elementary and middle schools, students are placed in classes without regard to their background or abilities, and no ability tracking occurs before they enter high school. Most students begin high school immediately after completing middle school. They can choose between academic programs that prepare them for higher education or vocational programs that lead directly to employment. Students have a legal right to attend high school, but apply to specific schools and study programs based on their middle school grade point average.

The Norwegian National Curriculum was revised in 2020, the year before we implemented the ROBUST intervention. Relevant for our study, the updated plans introduced a revised general part, as well as a new interdisciplinary subject called "Health and Life Skills", both of which increased schools' responsibility for fostering students' social

and emotional development. However, the revision included no curriculum nor any support materials or guidance for teachers or schools on how to effectively provide this support for the students. ROBUST was designed to address these critical needs by offering a comprehensive, scientifically grounded curriculum, resources for both teachers and students, and professional development for teachers.

Our study focuses on 9th-grade middle school students in Norway, a critical stage of both academic and personal development. At this point, students face increasing academic demands, as their grades influence high school opportunities. They are also navigating identity formation and social belonging, making this a particularly vulnerable period.

Among Norwegian youth, self-reported symptoms of anxiety and depression are notably high, particularly among females. A meta-analysis of Norwegian studies found that 22 percent of female and 11 percent of male youth reported clinical levels of anxiety and depression symptoms (Potrebny et al., 2025). Furthermore, Norway is among the countries that have experienced the largest increase in self-reported symptoms over the past 30 years. However, the rise in such symptoms during the COVID-19 pandemic was relatively modest (Potrebny et al., 2025).

Sample and Procedures

In 2019, we signed agreements with five municipalities to join the project. Teachers and principals in these municipalities were introduced to the project through information meetings. School principals, together with their teachers, decided whether their schools would participate. To join, principals and participating teachers signed a collaboration agreement outlining their responsibilities. The municipality administration and school principals were responsible for facilitating data collection and informing parents about the project.

Teachers committed in advance to attending the teacher professional development program and delivering 25 intervention-based lessons to their 9th-grade classes if assigned to the treatment group. To encourage commitment to data collection, teachers in the

control group were offered the same professional development two years later. Of the 36 schools invited, 27 participated, encompassing a total of 84 classes.

We originally planned to implement ROBUST for 8th graders, which is the first year of middle school, in the academic year 2020–2021. However, due to the pandemic and subsequent school closures, we postponed implementation of the curriculum to the following year, when these students were in 9th-grade. During this academic year there were no school closures in the region.

In April 2021, towards the end of 8th grade, we informed parents about the project through text messages and a webpage link with comprehensive project descriptions. While students had to participate in the ROBUST-lessons as part of standard school instruction, participation in the research project was optional. We therefore requested consent for their children to participate in the research project. We received parental consent for 92 percent of the students. After collecting parental consents, we randomized classes and teachers to treatment and control groups. Schools received immediate notification about the treatment status for each class to facilitate planning for the following year.

In May 2021, we conducted the baseline survey. Students received a link to the survey by email and completed it during regular class hours. All students who received the link to the survey had received parental consent, and in addition, we asked for their own consent: On the first page of the survey, we informed students about the research project and asked for their consent to participate. 95 percent of students agreed to participate. Those who did not consent were redirected to a webpage with engaging math exercises. Students without parental consent also received a link by email, but their link led directly to the math page instead of the survey.

We conducted the post-survey in May 2022 at the end of 9th-grade and, in May 2023, just before students graduated from middle school, we conducted the follow-up survey. The same test battery was used in all three surveys, with one exception: in the follow-up survey, we also asked students about their choice of track for high school, as

discussed in the section on Academic Motivation and Achievement.

Extensive measures were taken to reduce contamination from treated to control classes. Firstly, teachers at the same grade level who collaborate extensively were randomized in pairs so that they were in the same study arm. Secondly, treated teachers had to sign a non-disclosure agreement. Thirdly, access to web resources was via a password-protected website only accessible to treated teachers. A survey of the control classroom teachers suggest that they also worked to support students social and emotional development. This is as expected as it is a part of the National Curriculum, as described above. The treatment control contrast is a strengthening and systematization of this role, by equipping treated teachers with training, support material and a more detailed curriculum.

Quality of Implementation

We developed a fidelity check list for teachers to measure compliance and dosage (Ertesvåg et al., 2020). Teacher reports indicated a high level of compliance with the intervention. On average, teachers completed 21.9 out of 25 lessons (SD = 4.45), reflecting substantial intervention dosage. For each lesson, teachers rated the extent to which they implemented each activity, using a 3-point scale with response options: (0) "None," (1) "Some," and (2) "All." The average fidelity score was 1.55 (SD = 0.35), indicating that teachers reported to have implemented the majority of activities as prescribed.

In the post-survey (T2) we measured reach[5] by asking students approximately how many Robust lessons they had attended throughout the school year. This allowed us to measure class reach by computing the average number of ROBUST lessons attended among all students in each class. The mean class reach was 18.71 (SD = 2.48), suggesting substantial variation in reach across classes. In addition to variation in teacher compliance, this is likely due to due to class variation in student absence, since the implementation was at the end of the COVID pandemic.

[5] Reach is defined as how much the eligible population participated in the intervention (Durlak, 2016)

In addition to investigating treatment effects in our full sample, we will investigate effects in the sample of schools with high reach. Notably, simply excluding classes with low reach would bias our experimental design, since reach is not at random. Instead we calculate the minimum class reach in each block (school), and refer to this as a measure of school reach. When investigating treatment impact on schools with high reach, we exclude the blocks (schools) with school reach lower than 16 lessons, which include both treated and control classes.

In Table 1 we provide a summary of our samples. A total of 2,146 students from 84 classes across 27 schools were invited to participate in the study. Of these, we obtained parental consent for 1,968 students, and subsequent student consent for 1,879 students. Among the consenting students, 1,404 attended a school with high reach.

Results

Summary Statistics and Balance

In Table 2 we present summary statistics for student and family characteristics, as well as baseline assessments, for both the full student sample and the subset of students in schools with high reach. For each listed variable, we show mean values for the treatment and control groups. Additionally, we investigate balance by regressing each variable, standardized as a z-score with a mean of 0 and standard deviation of 1, on the treatment indicator, controlling for block fixed effects.

The treatment and control groups are generally well-balanced at baseline. Students in the treatment group score slightly lower on relationship skills and emotion regulation. Given the large number of variables examined for balance, these minor differences do not raise significant concerns. Nevertheless, in our analyses we control for baseline measure of dependent variable and all variables listed in Panel A Table 2. Additionally, Figures A.1 to A.30 in the Appendix provide robustness checks showing that our results are robust to the exclusion of these control variables.

Impacts on Mental Health

In Table 3 we investigate the treatment effects on mental health outcomes (Hypothesis 1) at post-intervention (T2) and follow-up (T3) using the OLS model in Equation 1. In all models, we control for student and family characteristics, as listed in Table 2, along with the baseline measure of the dependent variable and block fixed effects. In parentheses, we report cluster-robust standard errors, clustered at the class level. Panel A presents treatment effects for the full sample, while Panel B presents treatment effects for students in schools with high reach.

In the full sample, at the one-year follow-up, the treatment increased mental well-being by 12 percent of a standard deviation ($p < .01$) and reduced emotional distress by 10 percent of a standard deviation ($p < .05$). Notably, while the immediate post-intervention effects are in the expected direction for both outcomes, they are smaller in magnitude and not statistically significant for emotional distress. When restricting the sample to students in high-reach schools, the effect sizes increase to 14 percent of a standard deviation for both outcomes ($p < .01$). However, there is no significant effect for academic stress in either the full sample or among students in high-reach schools.

In Table 4 we investigate differential treatment effects across different levels of mental well-being at baseline (Hypothesis 2),[6] using a tertile split and the OLS interaction model in Equation 2. This analysis was not pre-registered (AEARCTR0009429) and should be considered exploratory. The treatment effects are substantially larger for students with low mental well-being at baseline. For this group, the effect on mental well-being at follow-up is 21 percent of a standard deviation ($p < .01$), while the effect on emotional distress is a reduction of 14 percent of a standard deviation ($p < .10$). Additionally, for these students, the treatment reduced academic stress by 13 percent of a standard deviation ($p < .10$) at post-intervention, though this effect is no longer significant at follow-up.

[6] In Figures A.16 to A.18 in the Appendix we investigate differential treatment effects across different levels of emotional distress at baseline.

Treatment effects for students with high baseline mental well-being are significantly smaller, with estimates close to zero. For students with medium baseline mental well-being, the effects are also smaller, approximately half the size of those observed for students with low mental well-being. However, the difference in effects between students with low and medium baseline mental well-being is not statistically significant in any model specification.

Impacts on Academic Motivation and Achievement

In Table 5 we investigate the treatment effects on academic motivation, grades in math, English and Norwegian, as well as track choice (Hypothesis 3). The results indicate that the treatment had a significant effect on motivation in the full sample at follow-up, with an effect size of 11 percent of a standard deviation ($p < .05$). The treatment also seemed to increase grades in math by 8 percent of a standard deviation ($p < .10$), but had no significant effect on grades in Norwegian or English. The positive effects on academic motivation and math are similar in the sample of high-reach schools.

In the last column of Table 5 we can also see that the treatment had an effect on track choice. In the full sample, it increased the likelihood of applying to the academic track by 4.2 percentage points ($p < .10$), while in the sample of high-reach schools, it increased this likelihood by 8.0 percentage points ($p < .01$).

In Table 6, we examine differential effects on academic motivation and school outcomes across different levels of motivation at baseline, using a tertile split (Hypothesis 4).[7] This analysis was not pre-registered and should be considered exploratory. The results indicate that the effects on motivation and math were particularly strong for students with low baseline motivation. For math, the effect size is 15 percent of a standard deviation ($p < .05$) in the full sample and 22 percent ($p < .05$) in the sample of high-reach schools. For track choice, we do not observe larger effects for students in the bottom tertile. If anything, the findings provide suggestive evidence that the effect is largest for students in

[7] In Table A.2 in the Appendix we investigate differential treatment effects across different levels of mental well-being at baseline for the academic motivation and school outcomes.

the middle tertile.

Mechanism Investigation

In Table 7 we examine the treatment effects on our four measures of coping skills (registered as secondary outcomes in the pre-registration). The results indicate that the treatment had a significant effect on relationship skills at follow-up. In the full sample, treated students scored approximately 9 percent of a standard deviation higher than students in the control group ($p < .05$). Among students in high-reach schools, the treatment effect was 11 percent of a standard deviation ($p < .05$). We find no significant effects on other coping skills, neither in the full sample nor in the high-reach schools sample.

In Table 8 and Table 9, we investigate whether the impact on relationship skills serves as an important mechanism for the intervention's effects on mental health, academic motivation and school outcomes (primary outcomes). We focus these analyses on the sub-samples with highest treatment impact on primary outcomes. These analyses were not pre-registered and should be considered exploratory.

In Table 8, we investigate differential effects on coping skills based on baseline measures of mental well-being (Panel A) and motivation (Panel B), using tertile splits. This table focuses on the sample of high-reach schools. In Panel A we can see that the treatment effect on relationship skills is particularly strong among students with low baseline mental well-being, with an effect size of 28 percent of a standard deviation ($p < .01$). Similarly, in Panel B, the treatment effect is particularly large for students with low baseline motivation, with an effect size of 21 percent of a standard deviation ($p < .05$).

The differential effects on relationship skills documented in Panels A and B correspond to our differential effect investigations on mental well-being and emotional distress in Table 4 and on motivation and math in Table 6. This suggests that relationship skills may serve as an important mechanism for the intervention effects. In Table 9 we investigate the extent to which the treatment effects on our primary outcomes can be explained by the treatment effects on relationship skills, or any of the other measured

coping skills, through a decomposition analysis. As for 9 we focus only on High-Reach Schools. Furthermore, for all outcomes except Track choice, we focus only on the bottom tertile of Well-being / Motivation.

In the first two columns we decompose the treatment effects on mental well-being and emotional distress in the high-reach schools/low baseline mental well-being sample. The results indicate that a substantial part (20 percent) of the effect on mental well-being can be explained by improved relationship skills, whereas this is not the case for emotional distress. In the next two columns we decompose the treatment effects on motivation and math in the high-reach schools/low baseline motivation sample. The results suggest that neither of these effects can be explained by the effect on relationship skills or any of the other measured coping skills. Finally, in the last column we decompose the treatment effect on track choice in the high-reach schools sample. We find that this effect cannot be explained by the impact on relationship skills or any of the other coping skills measures.

Discussion and Conclusion

The growing mental health crisis among adolescents underscores the need for scalable, evidence-based prevention strategies. Mental health challenges during adolescence can have profound and lasting effects on educational attainment, labor market participation, and overall well-being. To address this challenge, our study examines the effects of school-based, teacher-led coping skills instruction on youth mental health and academic achievement. We conducted a two-armed RCT involving 84 classes and 1,879 ninth-grade students aged 14–15 years.

We found that the ROBUST intervention positively impacted students' mental health and academic outcomes, with notable heterogeneity across subgroups. At the one-year follow-up, we observed significant improvements in students' mental well-being and reductions in emotional distress. The impacts were particularly pronounced among students with low baseline mental well-being, increasing mental health by 21 percent of a standard deviation and reducing emotional distress by 14 percent of a standard deviation.

The intervention also enhanced academic motivation and increased the likelihood of choosing an academic high school track by 4.2 percentage points. Additionally, it had a positive effect on math performance, particularly among students with initially low academic motivation, with an effect estimate of 15 percent of a standard deviation, but showed no significant effects on performance in English or Norwegian.

Despite these promising findings, our study has several limitations that should be acknowledged. First, even if our intervention targeted four coping skills, we can only document lasting impacts on relationship skills. We do not know why the intervention did not impact problem solving skills, growth mindset and emotion regulation. This may be due to the curriculum not being effective in developing these skills, the possibility that these skills were already targeted by the business-as-usual curriculum, or measurement issues (e.g., our assessments may not have adequately captured the content of the ROBUST topics). While substantial progress has been made over the past decade, further research is needed to improve the measurement of social and emotional skills (McKown, 2019).

The limited effects on measured social and emotional skills also constrain our understanding of the underlying mechanisms. Our mechanism investigation suggests that improvements in relationship skills may partly explain the intervention's impact on mental well-being. The treatment significantly enhanced relationship skills, particularly among students with low baseline mental well-being and motivation. Decomposition analyses indicate that about 20 percent of the improvement in mental well-being can be attributed to these gains. However, no similar mediation was observed for emotional distress, motivation, or academic outcomes, suggesting that other unmeasured skills or other factors may contribute to these effects.

Second, as for all field experiments, the external validity of our findings warrants careful consideration. Our study was conducted in a specific context, Norwegian public middle schools, which may limit the generalizability of our results to other educational systems, because of differences in, for example, teacher training, curriculum structure, and

student support services. Additionally, participating schools self-selected into our study. These schools may be more motivated to implement the ROBUST curriculum with greater care compared to the average school. As such, in a universal roll-out of the intervention, compliance may be lower, particularly if the curriculum is mandated rather than voluntarily adopted.

Finally, the intervention was implemented at the end of the COVID pandemic (2021/2022). Students in our sample transitioned to middle school in August 2020, a period marked by frequent school closures. As a result, during their first year of middle school, the year before the intervention, they had limited social contact with their new classmates. It is well documented that youth mental health issues increased during the pandemic (Deng et al., 2023). Although this increase appears to have been less pronounced in Norway (Potrebny et al., 2025), these circumstances may have influenced how receptive students were to the intervention. In addition, student absenteeism due to COVID infections may have limited the reach of the intervention.

Still, our findings offer strong evidence that school-based, teacher-led coping skills instruction can enhance students' mental health and influence educational outcomes. The intervention significantly improved students' mental well-being, and reduced emotional distress. The largest impacts were observed among students with low baseline mental well-being. These results have important policy implications. As mental health challenges among youth continue to rise, scalable teacher led school-based prevention programs like ROBUST represent an effective strategy for promoting adolescent mental well-being. By integrating such programs into regular school curricula, policymakers can support the development of social and emotional skills that may have enduring effects on mental health, educational attainment, and labor market outcomes.

While we document impacts on mental health and academic outcomes in a one-year follow-up, future research should investigate the long-term effects of the ROBUST curriculum on educational and labor market outcomes. Understanding whether the

observed improvements in mental health and academic motivation persist into high school and beyond would provide crucial insights for policymakers considering large-scale implementation of similar interventions.

References

Alan, S., Boneva, T., & Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal Of Economics*, *134*(3), 1121–1162. https://doi.org/10.1093/qje/qjz006

Alan, S., & Ertac, S. (2018). Fostering patience in the classroom: Results from randomized educational intervention. *Journal Of Political Economy*, *126*(5), 1865–1911. https://doi.org/10.1086/699007

Albert, E. (2010). *Overcoming destructive beliefs, feelings, and behaviors: New directions for rational emotive behavior therapy*. Prometheus Books.

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217–237. https://doi.org/10.1016/j.cpr.2009.11.004

Bas, G. (2021). Relation between student mental health and academic achievement revisited: A meta-analysis. In *Health and academic achievement-new findings*. IntechOpen. https://doi.org/10.5772/intechopen.95766

Berger, E. M., Fehr, E., Hermes, H., Schunk, D., & Winkel, K. (2025). The impact of working-memory training on children's cognitive and noncognitive skills. *Journal Of Political Economy*, *133*(2), 492–521. https://doi.org/10.1086/732884

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D., & Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice*, *11*(3), 230–241. https://doi.org/10.1093/clipsy.bph077

Bloem, J. R. (2022). How much does the cardinal treatment of ordinal variables matter? An empirical investigation. *Political Analysis*, *30*(2), 197–213. https://doi.org/10.1017/pan.2020.55

Bond, T. N., & Lang, K. (2019). The sad truth about happiness scales. *Journal of Political Economy*, *127*(4), 1629–1640. https://doi.org/10.1086/701679

Burnette, J. L., Knouse, L. E., Vavra, D. T., O'Boyle, E., & Brooks, M. A. (2020). Growth mindsets and psychological distress: A meta-analysis. *Clinical Psychology Review*, *77*(1), 101816. https://doi.org/10.1016/j.cpr.2020.101816

Bütikofer, A., Ginja, R., Karbownik, K., & Landaud, F. (2024). (breaking) intergenerational transmission of mental health. *Journal of Human Resources*, *59*(S), S108–S151. https://doi.org/10.3368/jhr.1222-12711R2

Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology*, *56*(2), 267–283. https://doi.org/10.1037/0022-3514.56.2.267

Castelpietra, G., Knudsen, A. K. S., Agardh, E. E., benedetta Armocida, Beghi, M., Iburg, K. M., Logroscino, G., Ma, R., Starace, F., Steel, N., Addolorato, G., Andrei, C. L., Andrei, T., Ayuso-Mateos, J. L., Banach, M., Bärnighausen, T. W., Barone-Adesi, F., Bhagavathula, A. S., Carvalho, F., … Monasta, L. (2022). The burden of mental disorders, substance use disorders and self-harm among young people in europe, 1990-2019: Findings from the global burden of disease study 2019. *The Lancet Regional Health - Europe*, *16*, 1–18. https://doi.org/10.1016/j.lanepe.2022.100341

Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development*, *94*(5), 1181–1204. https://doi.org/10.1111/cdev.13968

Compas, B. E., Jaser, S. S., Bettis, A. H., Watson, K. H., Gruhn, M. A., Dunbar, J. P., Williams, E., & Thigpen, J. C. (2017). Coping, emotion regulation, and psychopathology in childhood and adolescence: A meta-analysis and narrative review. *Psychological Bulletin*, *143*(9), 939–991. https://doi.org/10.1037/bul0000110

Costello, E. J., He, J.-p., Sampson, N. A., Kessler, R. C., & Merikangas, K. R. (2014). Services for adolescents with psychiatric disorders: 12-month data from the national comorbidity survey-adolescent. *Psychiatric Services*, *65*(3), 359–366. https://doi.org/10.1176/appi.ps.201100518

Cuddy, E., & Currie, J. (2020a, October). *Rules vs. discretion: Treatment of mental illness in U.S. adolescents* (Working Paper No. 27890). National Bureau of Economic Research. https://doi.org/10.3386/w27890

Cuddy, E., & Currie, J. (2020b). Treatment of mental illness in american adolescents varies widely within and across areas. *Proceedings of the National Academy of Sciences*, *117*(39), 24039–24046. https://doi.org/10.1073/pnas.2007484117

Currie, J. (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. https://doi.org/10.3368/jhr.2023SI-intro-child-mental-health

Currie, J., Stabile, M., Manivong, P., & Roos, L. L. (2010). Child health and young adult outcomes. *Journal of Human resources*, *45*(3), 517–548. https://doi.org/10.3368/jhr.45.3.517

Deming, D. J. (2017). The growing importance of social skills in the labor. *The Quarterly Journal of Economics*, *132*(4), 1593–1640. https://doi.org/10.1093/qje/qjx022

Deming, D. J. (2022). Four facts about human capital. *Journal of Economic Perspectives*, *36*(3), 75–102. https://doi.org/10.1257/jep.36.3.75

Deng, J., Zhou, F., Hou, W., Heybati, K., Lohit, S., Abbas, U., Silver, Z., Wong, C. Y., Chang, O., Huang, E., Zuo, Q. K., Moskalyk, M., Ramaraju, H. B., & Heybati, S. (2023). Prevalence of mental health symptoms in children and adolescents during the COVID-19 pandemic: A meta-analysis. *Annuals of the New York Academy of Sciences*, *1520*(1), 53–73. https://doi.org/10.1111/nyas.14947

Dunning, D. L., Griffiths, K., Kuyken, W., Crane, C., Foulkes, L., Parker, J., & Dalgleish, T. (2019). Research review: The effects of mindfulness-based interventions

on cognition and mental health in children and adolescents-a meta-analysis of randomized controlled trials. *Journal of Child Psychology and Psychiatry*, *60*(3), 244–258. https://doi.org/10.1111/jcpp.12980

Durlak, J. A. (2016). Programme implementation in social and emotional learning: Basic issues and research findings. *Cambridge Journal of Education*, *46*(3), 333–345. https://doi.org/10.1080/0305764X.2016.1142504

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science*, *14*(3), 481–496. https://doi.org/10.1177/1745691618804166

Edin, P.-A., Fredriksson, P., Nybom, M., & Öckert, B. (2022). The rising return to noncognitive skill. *American Economic Journal: Applied Economics*, *14*(2), 78–100. https://doi.org/10.1257/app.20190199

Ertesvåg, S. K., Tharaldsen, K., Vestad, L., Grini, N., & Bru, E. (2020). *Robust fidelity checklist* (Working Paper). University of Stavanger.

Fazel, M., Hoagwood, K., Stephan, S., & Ford, T. (2014). Mental health interventions in schools in high-income countries. *The Lancet Psychiatry*, *1*(5), 377–387. https://doi.org/10.1016/S2215-0366(14)70312-8

Ford, T., Goodman, R., & Meltzer, H. (2003). The British child and adolescent mental health survey 1999: The prevalence of DSM-IV disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(10), 1203–1211. https://doi.org/10.1097/00004583-200310000-00011

Gaumer Erickson, A. S., & Noonan, P. M. (2016). Self-efficacy formative questionnaire. In *The skills that matter: Teaching interpersonal and intrapersonal competencies in any classroom* (pp. 175–176). Thousand Oaks.

Goodman, R. (1998). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Gullone, E., & Taffe, J. (2012). The emotion regulation questionnaire for children and adolescents (ERQ-CA): A psychometric evaluation. *Psychological Assessment*, *24*(2), 409–417. https://doi.org/10.1037/a0025777

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451–464. https://doi.org/10.1016/j.labeco.2012.05.014

Heckman, J. J., & Pinto, R. (2015). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Reviews*, *34*(1-2), 6–31. https://doi.org/10.1080/07474938.2014.944466

Heckman, J. J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), 2052–2086. https://doi.org/10.1257/aer.103.6.2052

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482. https://doi.org/10.1086/504455

Johnston, D. W., Schurer, S., & Shields, M. A. (2013). Exploring the intergenerational persistence of mental health: Evidence from three generations. *Journal of health economics*, *32*(6), 1077–1089. https://doi.org/10.1016/j.jhealeco.2013.09.001

Lazarus, R. S. (2006). *Stress and emotion: A new synthesis*. Springer publishing company.

Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics*, *3*(1), 101–128. https://doi.org/10.1257/app.3.1.101

McKown, C. (2019). Challenges and opportunities in the applied assessment of student social and emotional learning. *Educational Psychologist*, *54*(3), 205–221. https://doi.org/10.1080/00461520.2019.1614446

Merikangas, K. R., He, J.-p., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K., & Swendsen, J. (2010). Lifetime prevalence of mental

disorders in us adolescents: Results from the national comorbidity survey replication-adolescent supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(10), 980–989. https://doi.org/10.1016/j.jaac.2010.05.017

Métais, C., Waters, L., Martin-Krumm, C., Tarquinio, C., & Burel, N. (2024). A scoping review of universal school-based resilience programs for adolescents. *School Psychology*, 1–19. https://doi.org/10.1037/spq0000659

Murberg, T. A., & Bru, E. (2004). School-related stress and psychosomatic symptoms among Norwegian adolescents. *School psychology international*, *25*(3), 317–332. https://doi.org/10.1177/0143034304046904

Pekrun, R. (2024). Control-value theory: From achievement emotion to a general theory of human emotions. *Educational Psychology Review*, *36*(83), 1–36. https://doi.org/10.1007/s10648-024-09909-7

Potrebny, T., Nilsen, S. A., Bakken, A., von Soest, T., Kvaløy, K., Samdal, O., Sivertsen, B., Aase, H., & Bang, L. (2025). Secular trends in mental health problems among young people in norway: A review and meta-analysis. *European Child & Adolescent Psychiatry*, *34*(1), 69–81. https://doi.org/10.1007/s00787-024-02371-4

Racine, N., McArthur, B. A., Cooke, J. E., Eirich, R., Zhu, J., & Madigan, S. (2021). Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: A meta-analysis. *JAMA Pediatrics*, *175*(11), 1142–1150. https://doi.org/10.1001/jamapediatrics.2021.2482

Radez, J., Reardon, T., Creswell, C., Lawrence, P. J., Evdoka-Burton, G., & Waite, P. (2021). Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. *European Child & Adolescent Psychiatry*, *30*(1), 183–211. https://doi.org/10.1007/s00787-019-01469-4

Rege, M., Hanselman, P., Solli, I. F., Dweck, C. S., Ludvigsen, S., Bettinger, E., Crosnoe, R., Muller, C., Walton, G., Duckworth, A., & Yeager, D. S. (2021). How can we inspire nations of learners? An investigation of growth mindset and challenge-seeking in two countries. *American Psychologist*, *76*(5), 755–767. https://doi.org/10.1037/amp0000647

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.*

Rueger, S. Y., Malecki, C. K., Pyun, Y., Aycock, C., & Coyle, S. (2016). A meta-analytic review of the association between perceived social support and depression in childhood and adolescence. *Psychological Bulletin*, *142*(10), 1017–1067. https://doi.org/10.1037/bul0000058

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*(1), 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, *69*(3), 493–525. https://doi.org/10.1177/0013164408323233

Sorrenti, G., Zölitz, U., Ribeaud, D., & Eisner, M. (2024). The causal impact of socio-emotional skills training on educational success. *Review of Economic Studies*, *92*(1), 506–552. https://doi.org/10.1093/restud/rdae018

Stallard, P. (2010). Mental health prevention in UK classrooms: The FRIENDS anxiety prevention programme. *Emotional and Behavioural Difficulties*, *15*(1), 23–35. https://doi.org/10.1080/13632750903512381

Strand, B. H., Dalgard, O. S., Tambs, K., & Rognerud, M. (2003). Measuring the mental health status of the norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic Journal of Psychiatry*, *57*(2), 113–118. https://doi.org/10.1080/08039480310000932

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, *5*, 1–13. https://doi.org/10.1186/1477-7525-5-63

Thoits, P. A. (2011). Mechanisms linking social ties and support to physical and mental health. *Journal of Health and Social Behavior*, *52*(2), 145–161. https://doi.org/10.1177/0022146510395592

United Nations Children's Fund. (2021, October). *The state of the world's children 2021: On my mind – promoting, protecting and caring for children's mental health.* UNICEF.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Vestad, L., Bru, E., Virtanen, T. E., & Stallard, P. N. (2021). Associations of social and emotional competencies, academic efficacy beliefs, and emotional distress among students in lower secondary school. *Social Psychology of Education*, *24*(1), 413–439. https://doi.org/10.1007/s11218-021-09624-z

Werner-Seidler, A., et al. (2021). The impact of school-based depression and anxiety prevention programs: A meta-analysis of universal interventions. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(1), 33–47. https://doi.org/10.1016/j.cpr.2016.10.005

Woessmann, L. (2024). *Skills and earnings: A multidimensional perspective on human capital* (CESifo Working Paper No. 11428). Munich, CESifo GmbH. https://hdl.handle.net/10419/307358

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist*, *47*(4), 302–314. https://doi.org/10.1080/00461520.2012.722805

Yeager, D. S., Henderson, M. D., Paunesku, D., Walton, G. M., D'Mello, S., Spitzer, B. J., & Duckworth, A. L. (2014). Boring but important: A self-transcendent purpose for learning fosters academic self-regulation. *Journal of Personality and Social Psychology*, *107*(4), 559–580. https://doi.org/10.1037/a0037637

Table 1
Sample Description

	Munic.	Blocks	Schools	Classes	Students
Full sample	5	23	27	84	2,146
100.00					
Parental consent	5	23	27	84	1,968
91.72					
Student consent	5	23	27	84	1,879
87.57					
High reach schools	5	16	20	59	1,404
65.41					

Table 2
Summary Statistics and Balance

	All Students ($N = 1,879$)				High-Reach Schools ($N = 1,404$)			
	T	C	Coef.	SE	T	C	Coef.	SE
	A. Student and Family Characteristics							
Student is female	.507	.489	.042	(.035)	.496	.475	.039	(.044)
Family income	12.57	12.11	.069	(.048)	12.61	12.17	.063	(.057)
F/M: Immigrant	.251	.272	-.067	(.048)	.246	.276	-.095	(.059)
F: < Bachelor's degree	.594	.605	-.015	(.049)	.584	.605	-.022	(.062)
F: Bachelor's degree	.248	.240	.008	(.047)	.254	.240	.011	(.059)
F: > Bachelor's degree	.158	.155	.011	(.042)	.161	.155	.017	(.049)
M: < Bachelor's degree	.445	.460	-.020	(.043)	.440	.452	-.005	(.052)
M: Bachelor's degree	.436	.417	.033	(.044)	.431	.415	.015	(.053)
M: > Bachelor's degree	.119	.123	-.020	(.044)	.129	.133	-.015	(.051)
	B. Student Mental Health							
Emotion distress	-.017	.030	-.058	(.051)	-.014	-.012	-.016	(.058)
Mental well-being	-.040	.013	-.041	(.045)	-.029	.016	-.033	(.054)
Academic stress	-.004	.016	-.013	(.047)	-.018	-.037	.015	(.056)
	C. Student Motivation and Academic Achievement							
Math	.005	-.006	-.003	(.052)	.019	-.031	.035	(.065)
English	.038	-.041	.085	(.060)	.042	-.053	.105	(.070)
Norwegian	.024	-.018	.046	(.061)	.055	-.044	.106	(.070)
Motivation	-.069	.030	-.092	(.047)*	-.047	.018	-.056	(.055)
	D. Student Coping Skills							
Relationship skills	-.056	.022	-.066	(.042)	-.055	.032	-.080	(.048)*
Emotion regulation	-.055	.042	-.087	(.046)*	-.067	.048	-.109	(.054)**
Growth mindset	-.068	.006	-.059	(.044)	-.064	.006	-.056	(.051)
Problem solv. skills	-.036	.004	-.019	(.046)	-.027	-.030	.017	(.052)

Note. For each variable listed, we report mean values for both the treatment (T) and control (C) groups. Additionally, we report regression results where each variable is standardized and regressed on the treatment indicator and block fixed effects. All variables are standardized to a mean of 0 and a standard deviation of 1. Baseline survey variables are constructed in two steps: first, each item is standardized to a z-score (mean of 0, variance of 1); then, an equally weighted average of these standardized items is calculated and re-standardized to yield an aggregate measure with mean 0 and variance 1. Missing data are addressed through multiple imputation. We report cluster-robust standard errors, clustered at the class level, in parentheses. Statistical significance of random imbalances are denoted by ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

Table 3
Impact on Mental Health

	Mental Well-Being		Emotional Distress		Academic Stress	
	T2	T3	T2	T3	T2	T3
	A. All Students					
Trt.	.085**	.122***	-.027	-.095**	-.062	.010
	(.037)	(.043)	(.040)	(.045)	(.044)	(.052)
N	1,879	1,879	1,879	1,879	1,879	1,879
	B. High-Reach Schools					
Trt.	.089**	.136***	-.045	-.139***	-.064	-.017
	(.043)	(.050)	(.049)	(.052)	(.052)	(.062)
N	1,404	1,404	1,404	1,404	1,404	1,404
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note. In all models, we control for student and family characteristics, along with the baseline measure of the dependent variable. Cluster-robust standard errors, clustered at the class level, are reported in parentheses. Panel A shows treatment effects (trt) for the full sample ($N = 1,879$), while Panel B shows treatment effects for high-reach schools ($N = 1,404$). To address missing data, multiple imputation is applied. T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10). Statistical significance is denoted ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

Table 4
Differential Impact on Mental Health Across Baseline Mental Well-being

	Mental Well-Being		Emotional Distress		Academic Stress	
	T2	T3	T2	T3	T2	T3
	A. All Students					
Trt.	.159**	.214***	-.097	-.141*	-.125*	-.068
	(.067)	(.079)	(.073)	(.082)	(.070)	(.083)
Trt. × Mod.	-.063	-.107	.037	-.026	.064	.062
	(.106)	(.114)	(.104)	(.108)	(.110)	(.119)
Trt. × High	-.167*	-.177	.160	.166	.124	.168
	(.100)	(.116)	(.099)	(.115)	(.108)	(.131)
N	1,879	1,879	1,879	1,879	1,879	1,879
	B. High-Reach Schools					
Trt.	.178**	.270***	-.165*	-.212**	-.145*	-.097
	(.076)	(.087)	(.088)	(.096)	(.086)	(.097)
Trt. × Mod.	-.074	-.142	.112	.016	.125	.079
	(.118)	(.127)	(.123)	(.124)	(.124)	(.138)
Trt. × High	-.203*	-.276**	.252**	.222*	.113	.159
	(.112)	(.123)	(.116)	(.134)	(.124)	(.152)
N	1,404	1,404	1,404	1,404	1,404	1,404
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note. In all models, we control for student and family characteristics, along with the baseline measure of the dependent variable. Cluster-robust standard errors, clustered at the class level, are reported in parentheses. Panel A shows treatment effects for the full sample ($N = 1,879$), while Panel B shows treatment effects for high-reach schools ($N = 1,404$). Students are categorized into three equally sized groups with low, moderate, or high levels of mental well-being at baseline. To address missing data, multiple imputation is applied. T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10). Statistical significance is denoted ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

Table 5
Impact on Academic Motivation and School Outcomes

	Motivation		10th-grade			
	T2	T3	Mat.	Eng.	Norw.	Track
	A. All Students					
Trt.	.033	.112**	.078*	.015	-.005	.042*
	(.035)	(.043)	(.043)	(.050)	(.049)	(.023)
N	1,879	1,879	1,879	1,879	1,879	1,879
	B. High-Reach Schools					
Trt.	.036	.115**	.089*	.056	.036	.080***
	(.041)	(.052)	(.052)	(.061)	(.052)	(.025)
N	1,404	1,404	1,404	1,404	1,404	1,404
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note. In all models, we control for student and family characteristics, as well as the baseline measure of the dependent variable, except in the case of students' track choice. Cluster-robust standard errors, clustered at the class level, are reported in parentheses. Panel A shows treatment effects (trt) for the full sample ($N = 1,879$), while Panel B shows treatment effects for high-reach schools ($N = 1,404$). To address missing data, multiple imputation is applied. T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10). Statistical significance is denoted ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

Table 6

Differential Impact on Academic Motivation and School Outcomes Across Baseline Motivation

	Motivation		10th-grade			
	T2	T3	Mat.	Eng.	Norw.	Track
	A. All Students					
Trt.	.140*	.182**	.152**	-.007	-.008	.005
	(.078)	(.092)	(.073)	(.080)	(.084)	(.041)
Trt. × Mod.	-.159	-.103	-.167*	.048	-.012	.086
	(.109)	(.127)	(.095)	(.104)	(.116)	(.059)
Trt. × High	-.164	-.110	-.045	.026	.035	.035
	(.118)	(.123)	(.098)	(.091)	(.108)	(.054)
N	1,879	1,879	1,879	1,879	1,879	1,879
	B. High-Reach Schools					
Trt.	.138*	.163	.220**	.046	.004	.030
	(.083)	(.100)	(.085)	(.097)	(.099)	(.044)
Trt. × Mod.	-.204*	-.082	-.246**	.035	.018	.116*
	(.119)	(.140)	(.111)	(.124)	(.141)	(.064)
Trt. × High	-.105	-.069	-.136	.000	.088	.038
	(.119)	(.127)	(.115)	(.103)	(.124)	(.062)
N	1,404	1,404	1,404	1,404	1,404	1,404
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note. In all models, we control for student and family characteristics, along with the baseline measure of the dependent variable. Cluster-robust standard errors, clustered at the class level, are reported in parentheses. Panel A shows treatment effects for the full sample ($N = 1,879$), while Panel B shows treatment effects for high-reach schools ($N = 1,404$). Students are categorized into three equally sized groups with low, moderate, or high levels of motivation at baseline. To address missing data, multiple imputation is applied. T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10). Statistical significance is denoted ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

Table 7
Impact on Coping Skills

	Relationship Skills		Emotional Regulation		Growth Mindset		Problem Solv. Skills	
	T2	T3	T2	T3	T2	T3	T2	T3
	A. All Students							
Trt.	.070	.085**	-.004	.016	.042	.045	-.070	.017
	(.043)	(.042)	(.038)	(.037)	(.040)	(.045)	(.037)	(.041)
N	1,879	1,879	1,879	1,879	1,879	1,879	1,879	1,879
	B. High-Reach Schools							
Trt.	.087*	.114**	.022	.034	.065	.071	-.032	.023
	(.052)	(.047)	(.047)	(.044)	(.047)	(.050)	(.041)	(.051)
N	1,404	1,404	1,404	1,404	1,404	1,404	1,404	1,404
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. In all models, we control for student and family characteristics, along with the baseline measure of the dependent variable. Cluster-robust standard errors, clustered at the class level, are reported in parentheses. Panel A shows treatment effects (trt) for the full sample ($N = 1,879$), while Panel B shows treatment effects for high-reach schools ($N = 1,404$). To address missing data, multiple imputation is applied. T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10). Statistical significance is denoted ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

Table 8
Differential Impact on Coping Skills

	Relationship Skills		Emotional Regulation		Growth Mindset		Problem Solv. Skills	
	T2	T3	T2	T3	T2	T3	T2	T3
	A. High-Reach Schools by Baseline Mental Well-being							
Trt.	.106	.282***	.079	.093	.121	.105	-.001	-.025
	(.101)	(.095)	(.096)	(.094)	(.095)	(.106)	(.097)	(.086)
Trt. × Mod.	.004	-.204	-.005	.013	.002	.026	.032	.144
	(.126)	(.126)	(.131)	(.129)	(.127)	(.137)	(.136)	(.124)
Trt. × High	-.088	-.318**	-.208	-.233*	-.194	-.157	-.123	-.008
	(.148)	(.138)	(.141)	(.140)	(.128)	(.151)	(.137)	(.122)
N	1,404	1,404	1,404	1,404	1,404	1,404	1,404	1,404
	B. High-Reach Schools by Baseline Motivation							
Trt.	.110	.212**	.127	.077	.064	.048	.127	-.002
	(.101)	(.096)	(.091)	(.106)	(.097)	(.103)	(.098)	(.100)
Trt. × Mod.	.007	-.050	-.157	-.021	.065	.057	-.198	.101
	(.136)	(.135)	(.126)	(.142)	(.138)	(.149)	(.135)	(.135)
Trt. × High	-.081	-.245*	-.167	-.116	-.066	.014	-.263*	-.013
	(.125)	(.132)	(.138)	(.145)	(.119)	(.129)	(.137)	(.138)
N	1,404	1,404	1,404	1,404	1,404	1,404	1,404	1,404
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. In all models, we control for student and family characteristics, along with the baseline measure of the dependent variable. Cluster-robust standard errors, clustered at the class level, are reported in parentheses. Panel A shows treatment effects for high-reach schools by baseline mental well-being, while Panel B shows treatment effects for high-reach schools by baseline motivation. Students are categorized into three equally sized groups with low, moderate, or high levels of mental well-being and motivation at baseline. To address missing data, multiple imputation is applied. T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10). Statistical significance is denoted ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

Table 9

Treatment Effect Decomposition for High-Reach Schools at the Follow-Up

	Bottom well-being Tertile		Bottom Motivation Tertile		
	Wellb.	Distr.	Motv.	Math	Track
	A. Treatment Effect				
Total effect	.278***	-.227**	.166	.212**	.084***
	(.094)	(.101)	(.114)	(.096)	(.032)
	B. Decomposition				
Relationship skills	.052*	.012	.019	-.004	.003
	(.028)	(.027)	(.020)	(.014)	(.003)
emotion regulation	.021	-.010	.014	-.007	-.001
	(.024)	(.019)	(.018)	(.015)	(.002)
Growth mindset	.011	-.002	.014	.007	.006
	(.013)	(.004)	(.024)	(.016)	(.006)
Problem solv. skills	-.001	.000	.004	-.000	-.000
	(.021)	(.005)	(.037)	(.007)	(.001)
Direct effect	.195**	-.227**	.117	.217**	.076**
	(.079)	(.100)	(.102)	(.096)	(.031)
N	626	626	641	641	1,404
Controls	Yes	Yes	Yes	Yes	Yes

Note. This table presents a treatment effect decomposition of mental well-being, emotional distress, motivation, math, and students' decision to pursue the academic track at the follow-up evaluation (T3), using our preferred specification. Our preferred model specification includes block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. Cluster-robust standard errors, clustered at the class level, are reported in parentheses. Panel A presents the total treatment effect, whereas Panel B breaks down the total treatment effect into indirect effects and the direct effect. To address missing data, multiple imputation was applied. T3 represents the follow-up evaluation (Grade 10). Statistical significance is denoted ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

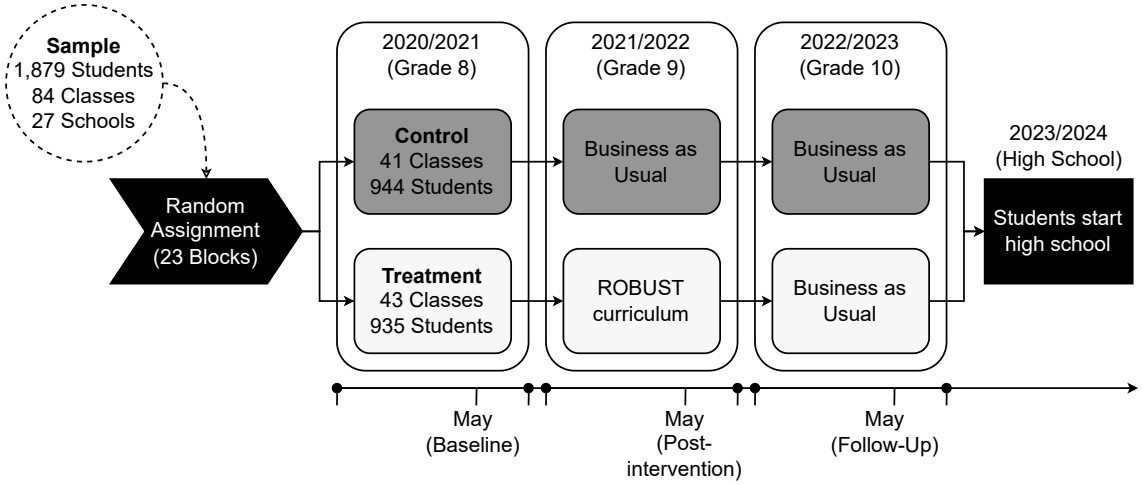

Figure 1
Experimental Design

Online Supplementary Material: The Impact of Teaching Coping Skills in Schools on Youth Mental Health and Academic Achievement: Evidence From a Field Experiment

Mari Rege[1], Edvin Bru[2], Ingeborg F. Solli[1], Maximiliaan W. P. T. Thijssen[3], Kjersti B. Tharaldsen[2], Lene Vestad[2], Sigrun K. Ertesvåg[2], Terje Ogden[4], and Paul N. Stallard[5]

[1]Department of Economics and Finance, University of Stavanger, Norway

[2]Norwegian Centre for Learning Environment and Behavioral Research in Education, University of Stavanger, Norway

[3]Knowledge Center for Education, University of Stavanger, Norway

[4]The Norwegian Center for Child Behavioral Development, Norway

[5]Department for Health, University of Bath, UK

Author Note

Mari Rege https://orcid.org/0000-0002-9574-5851,

Edvin Bru https://orcid.org/0000-0003-2004-6957,

Ingeborg F. Solli https://orcid.org/0000-0002-9016-9920,

Maximiliaan W. P. T. Thijssen https://orcid.org/0000-0001-8622-3106,

Kjersti B. Tharaldsen https://orcid.org/0000-0003-1536-730X,

Lene Vestad https://orcid.org/0000-0002-6558-0929,

Sigrun K. Ertesvåg https://orcid.org/0000-0001-7804-8571,

Terje Ogden https://orcid.org/0000-0002-2353-0798, and

Paul N. Stallard https://orcid.org/0000-0001-8046-0784.

This project was funded by the Research Council of Norway (grant number 325398). We thank all the teachers, students, and school administrators who participated. We also extend our gratitude to Hilde Ness Sandvold for her role as the administrative project coordinator. The RCT was pre-registered in the AEA RCT Registry (AEARCTR0009429).

Corresponding author: Mari Rege, University of Stavanger, Business School, Department of Economics and Finance, N-4036 Stavanger, Norway. E-mail: mari.rege@uis.no

Online Supplementary Material:
The Impact of Teaching Coping Skills in Schools on Youth Mental Health and Academic Achievement: Evidence From a Field Experiment

Contents

List of Tables

List of Figures

Mental Health Measures Pearson Correlation Matrix

Table A.1
Mental Health Measures Pearson Correlation Matrix

		Mental Well-Being			Emotional Distress			Academic Stress		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
	A. All Students ($N = 1,879$)									
Mental well-being	T1	1	.52	.46	-.45	-.31	-.26	-.27	-.14	-.12
	T2	.52	1	.52	-.41	-.43	-.37	-.27	-.25	-.20
	T3	.46	.52	1	-.33	-.33	-.37	-.21	-.22	-.16
Emotional distress	T1	-.45	-.41	-.33	1	.50	.40	.45	.23	.24
	T2	-.31	-.43	-.33	.50	1	.53	.30	.40	.28
	T3	-.26	-.37	-.37	.40	.53	1	.26	.28	.36
Academic stress	T1	-.27	-.27	-.21	.45	.30	.26	1	.34	.29
	T2	-.14	-.26	-.22	.23	.40	.28	.34	1	.34
	T3	-.12	-.20	-.16	.24	.28	.36	.29	.34	1
	B. High-Reach Schools ($N = 1,404$)									
Mental well-being	T1	1	.53	.50	-.48	-.31	-.26	-.28	-.14	-.10
	T2	.53	1	.53	-.41	-.42	-.38	-.29	-.26	-.19
	T3	.50	.53	1	-.34	-.33	-.36	-.23	-.21	-.15
Emotional distress	T1	-.48	-.41	-.34	1	.51	.40	.45	.25	.24
	T2	-.31	-.42	-.33	.51	1	.53	.30	.40	.28
	T3	-.26	-.38	-.36	.40	.53	1	.25	.28	.35
Academic stress	T1	-.28	-.29	-.23	.45	.30	.25	1	.35	.29
	T2	-.14	-.26	-.21	.25	.40	.28	.35	1	.34
	T3	-.10	-.19	-.15	.24	.28	.35	.29	.34	1

Note. This table presents Pearson correlation coefficients for the mental health measures. To address missing data, multiple imputation is applied. T1 refers to the baseline assessment (Grade 8). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

Treatment Impacts Under Different Model Specifications

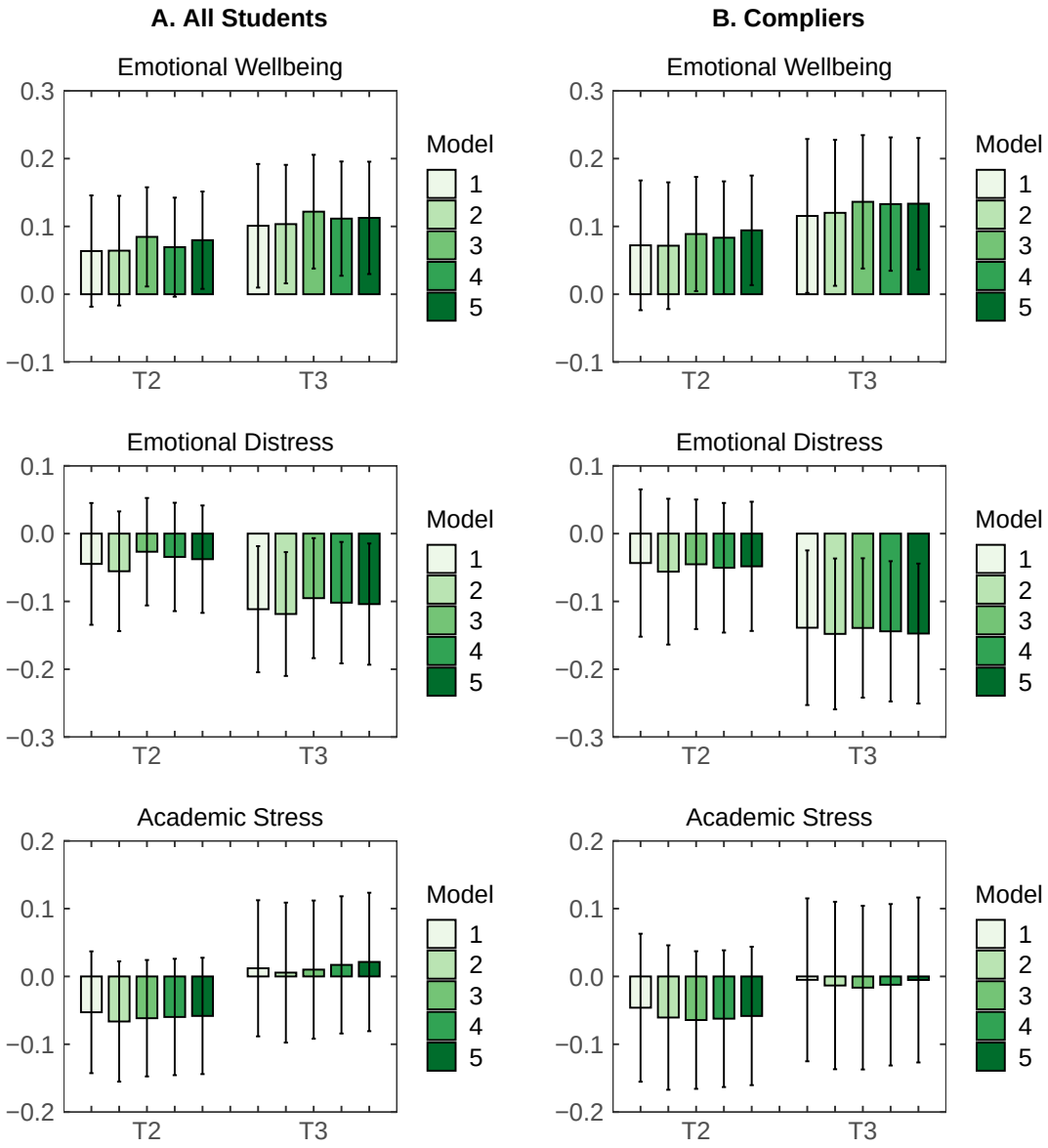

Figure A.1

Impact on Mental Health Under Different Model Specifications

Note. This figure displays the estimated treatment effects on mental health, along with their 95% confidence intervals, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to mental health. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1976). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

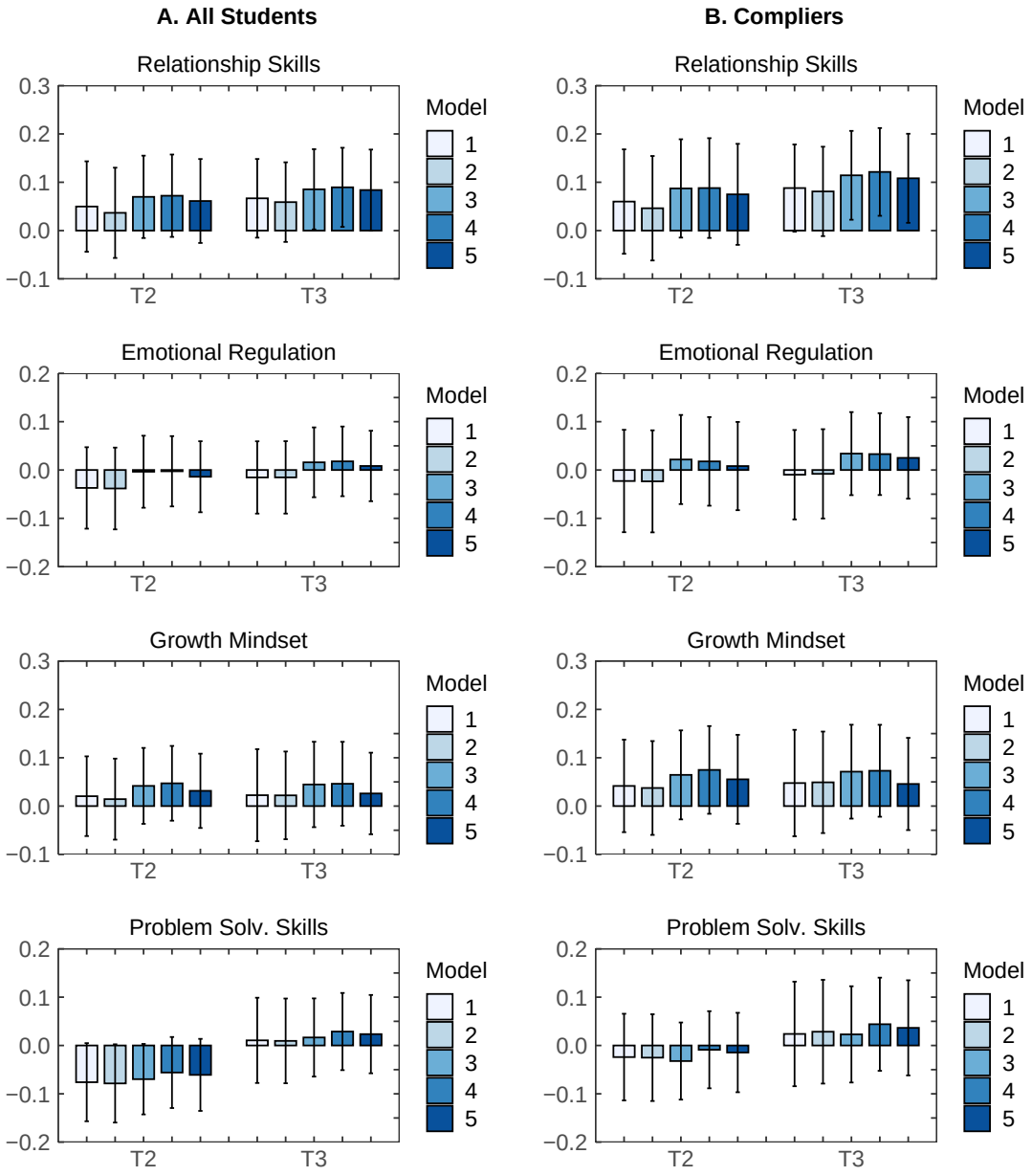

Figure A.2

Impact on Coping Skills Under Different Model Specifications

Note. This figure displays the estimated treatment effects on coping skills, along with their 95% confidence intervals, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

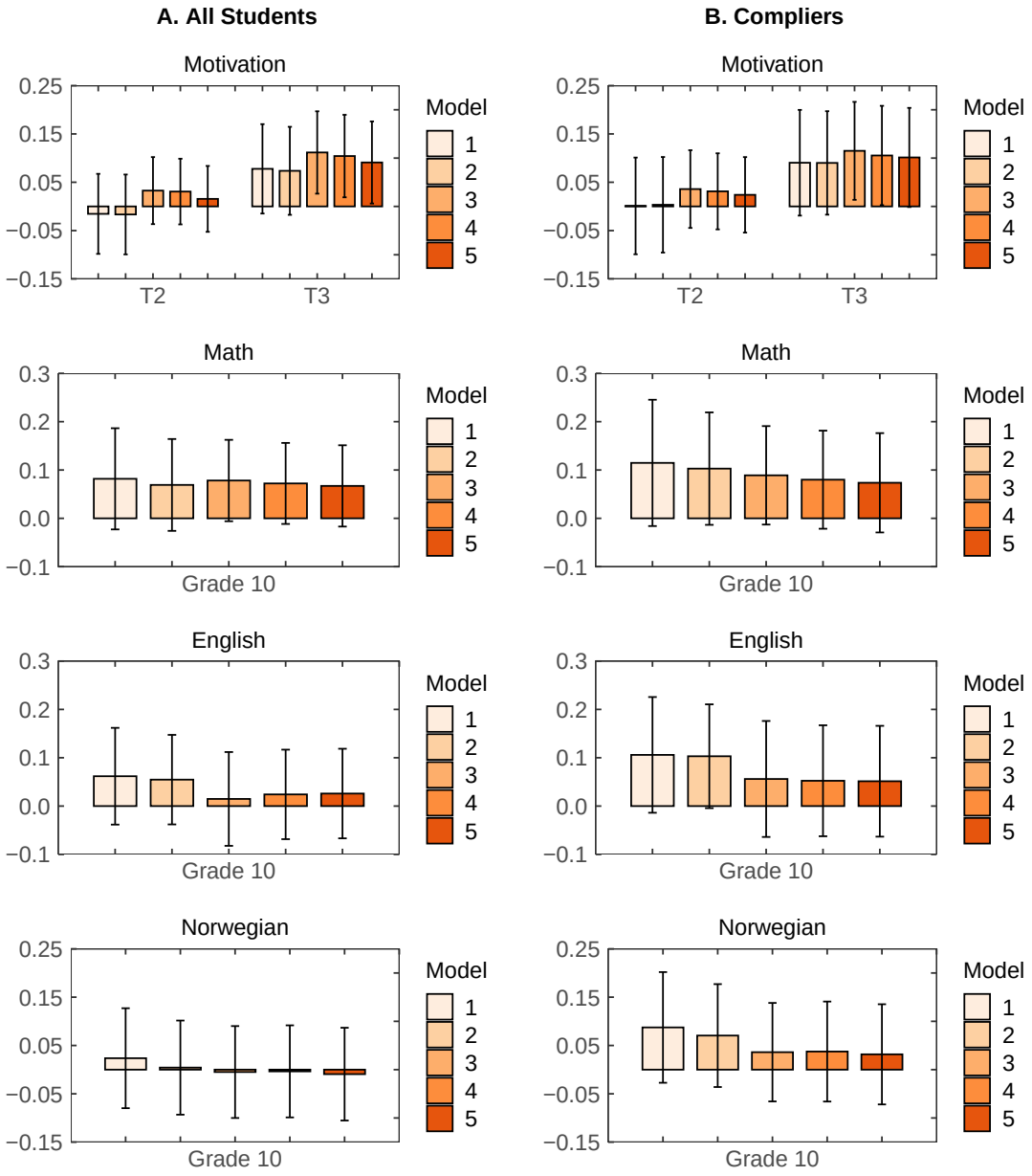

Figure A.3

Impact on Motivation and School Outcomes Under Different Model Specifications

Note. This figure displays the estimated treatment effects on motivation and school outcomes, along with their 95% confidence intervals, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

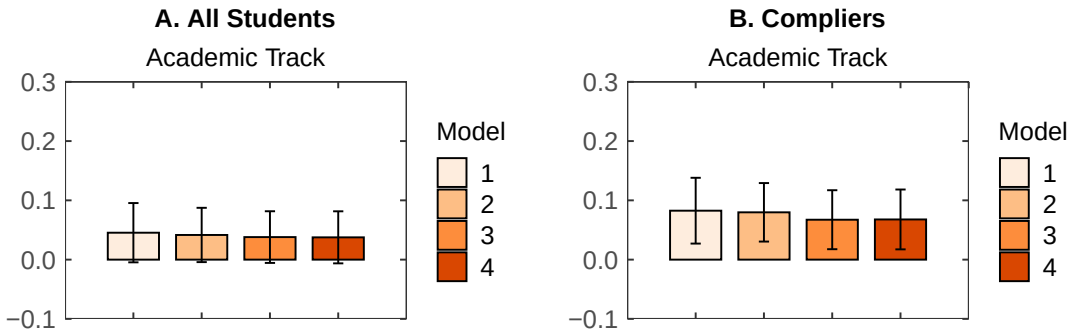

Figure A.4

Impact on Track Choice Under Different Model Specifications

Note. This figure displays the estimated treatment effects on track choice, along with their 95% confidence intervals, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

Differential Treatment Effects Under Different Model Specifications

Table A.2
Differential Impact on Motivation and School Outcomes

	Motivation		10th-grade			
	T2	T3	Mat.	Eng.	Norw.	Track
	A. All Students					
Trt.	.061	.136	.086	.001	-.044	.009
	(.071)	(.085)	(.072)	(.076)	(.084)	(.042)
Trt. × Mod.	-.013	.033	.032	.055	.089	.052
	(.107)	(.119)	(.093)	(.100)	(.115)	(.061)
Trt. × High	-.088	-.128	-.063	-.015	.032	.046
	(.112)	(.120)	(.100)	(.100)	(.118)	(.059)
N	1,879	1,879	1,879	1,879	1,879	1,879
	B. High-Reach Schools					
Trt.	.057	.107	.130	.055	-.017	.048
	(.083)	(.091)	(.086)	(.095)	(.098)	(.045)
Trt. × Mod.	.012	.152	.013	.057	.124	.068
	(.120)	(.121)	(.111)	(.123)	(.135)	(.067)
Trt. × High	-.096	-.171	-.151	-.064	.029	.022
	(.131)	(.132)	(.119)	(.124)	(.138)	(.065)
N	1,404	1,404	1,404	1,404	1,404	1,404
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note. In all models, we control for student and family characteristics, along with the baseline measure of the dependent variable. Cluster-robust standard errors, clustered at the class level, are reported in parentheses. Panel A shows treatment effects for the full sample ($N = 1,879$), while Panel B shows treatment effects for high-reach schools ($N = 1,404$). Students are categorized into three equally sized groups with low, moderate, or high levels of mental well-being at baseline. To address missing data, multiple imputation is applied. T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10). Statistical significance is denoted ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

By Baseline Mental Well-Being

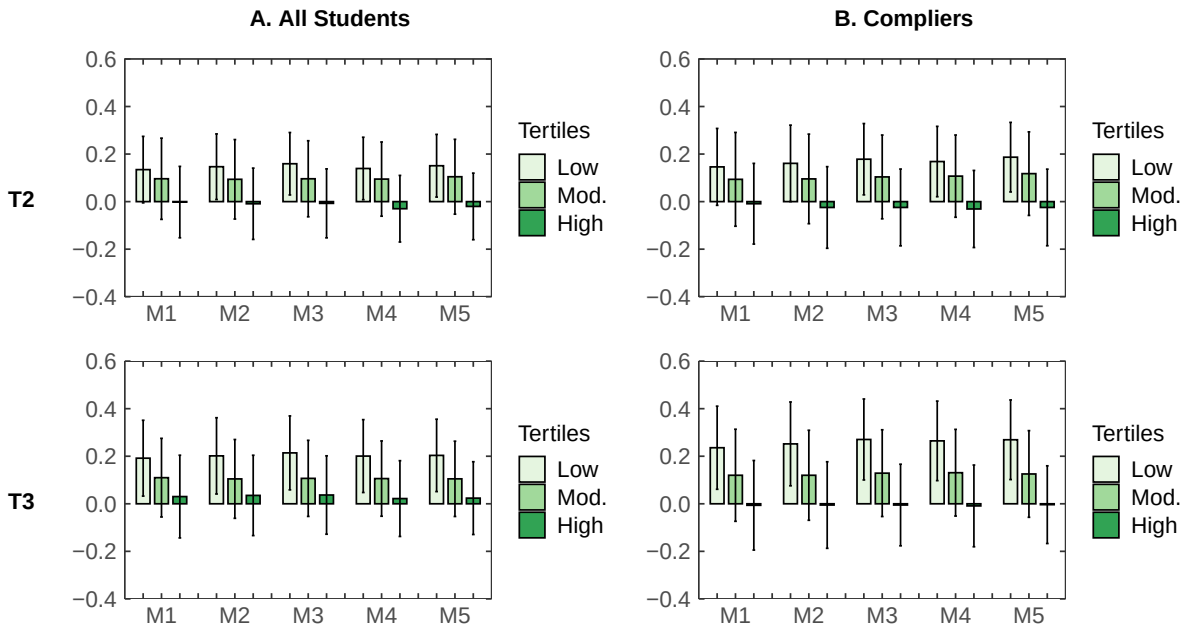

Figure A.5

Differential Impact on Mental Well-Being by Baseline Mental Well-Being Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on mental well-being, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

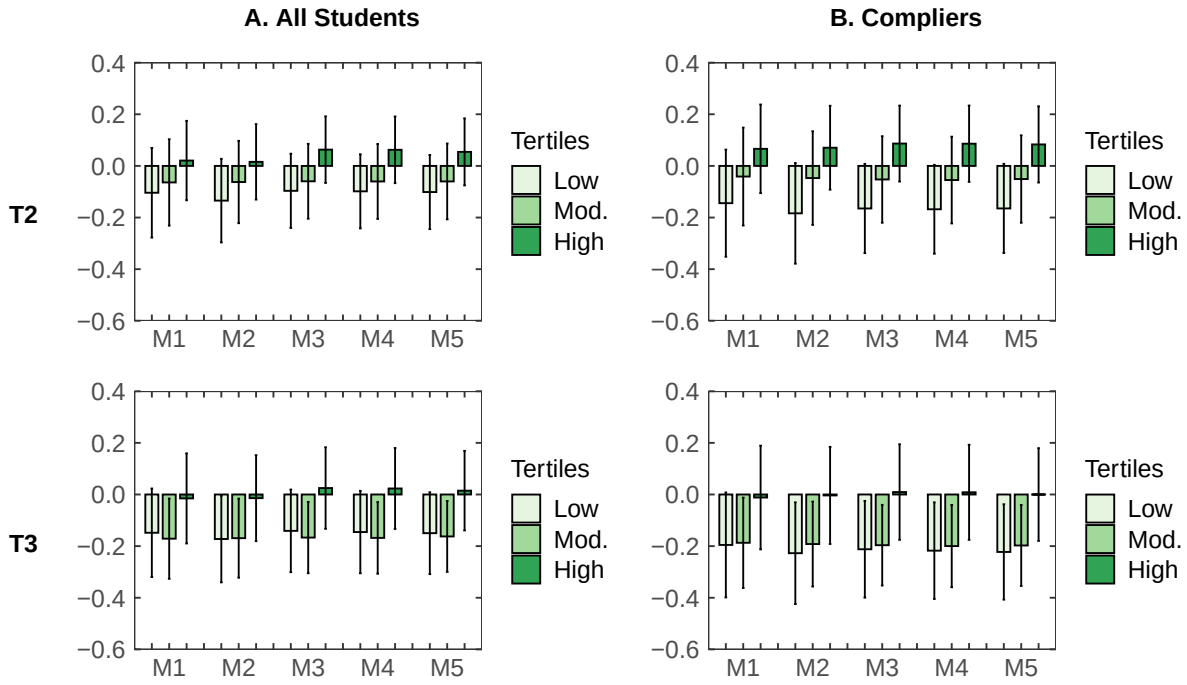

Figure A.6

Differential Impact on Emotional Distress by Baseline Mental Well-Being Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on emotional distress, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

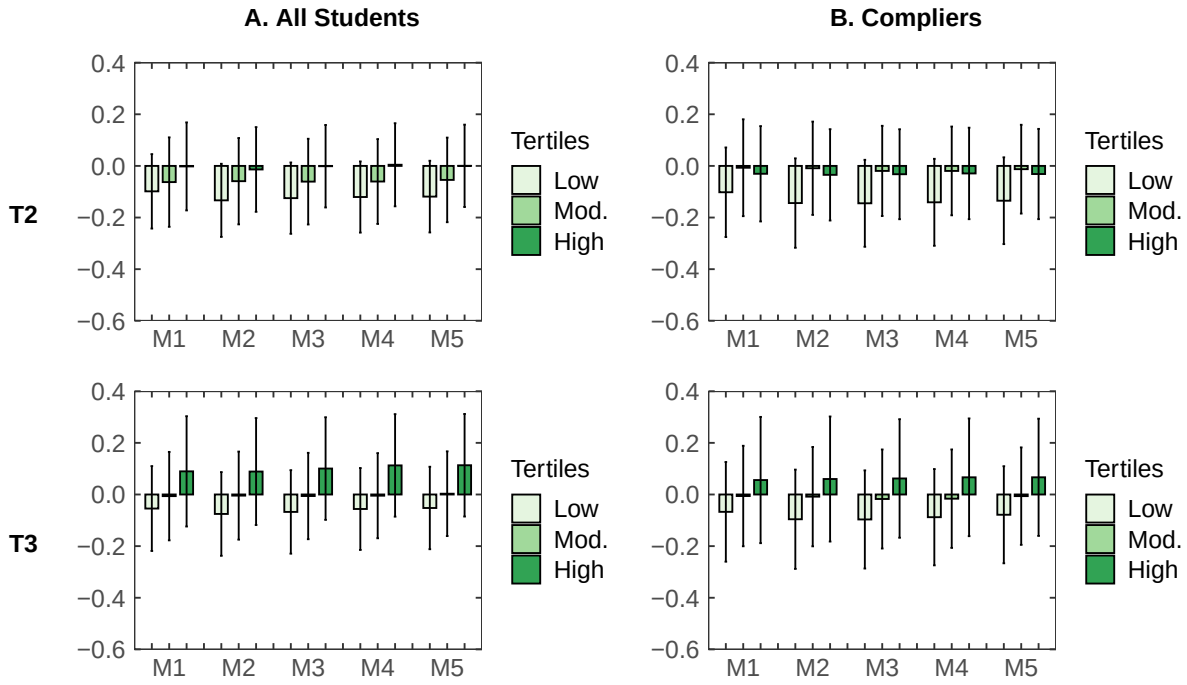

Figure A.7

Differential Impact on Academic Stress by Baseline Mental Well-Being Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on academic stress, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

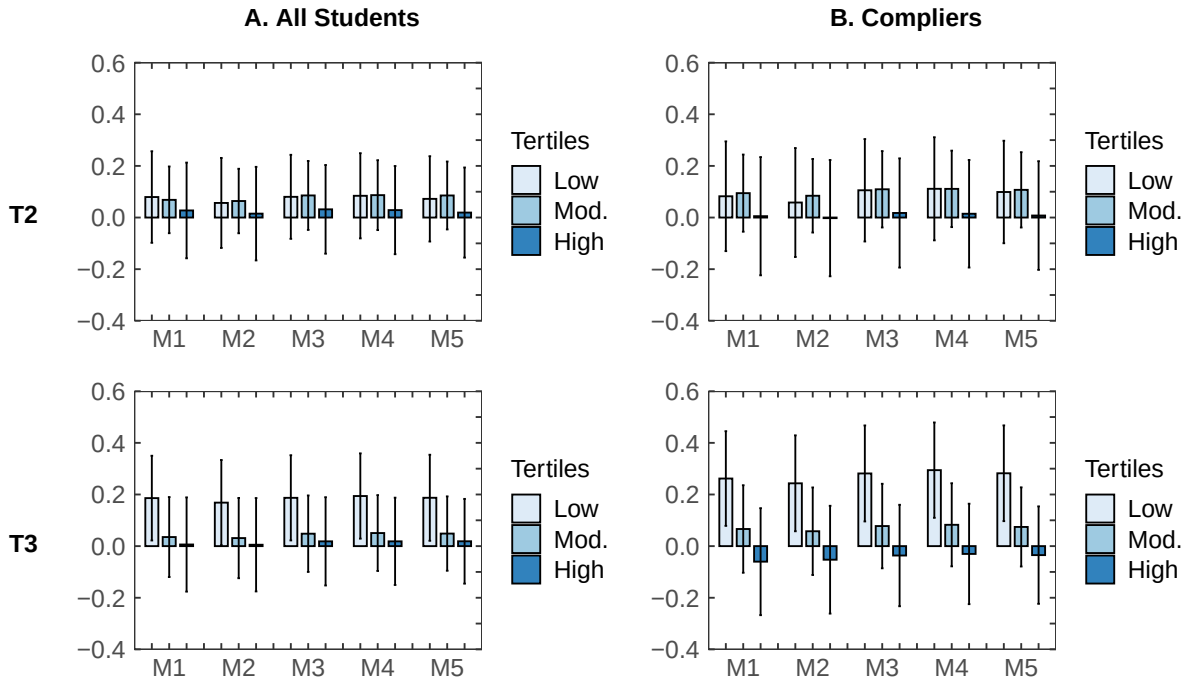

Figure A.8

Differential Impact on Relationship Skills by Baseline Mental Well-Being Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on relationship skills, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

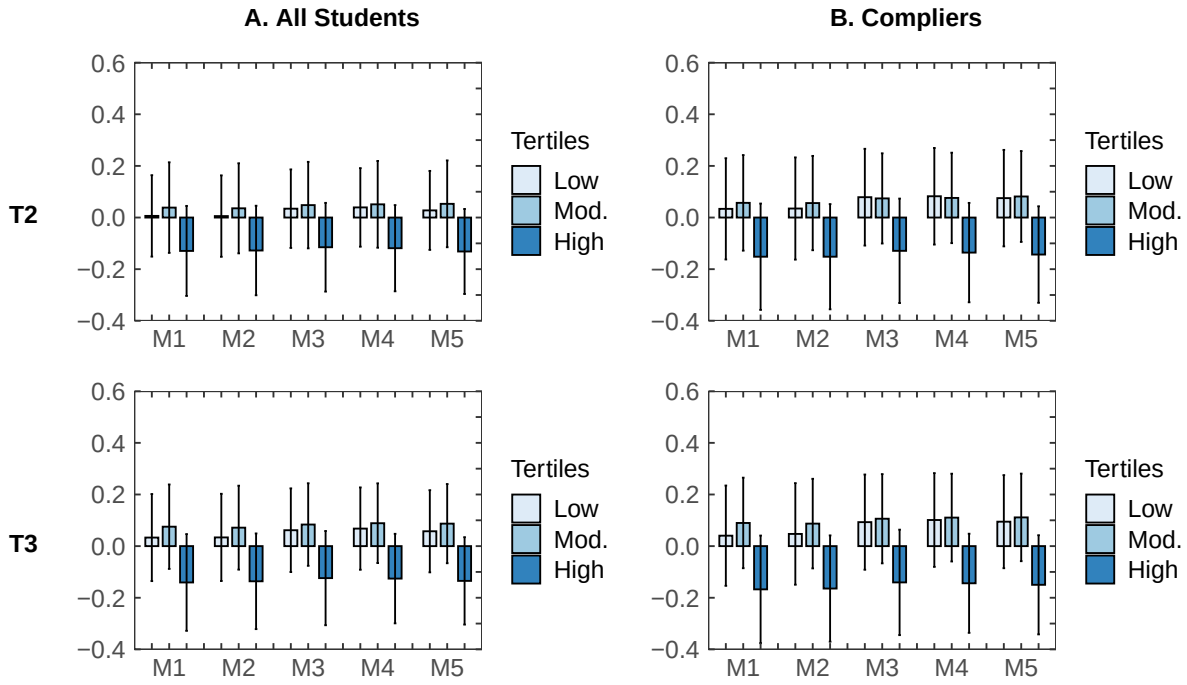

Figure A.9

Differential Impact on Emotion Regulation Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on emotion regulation, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

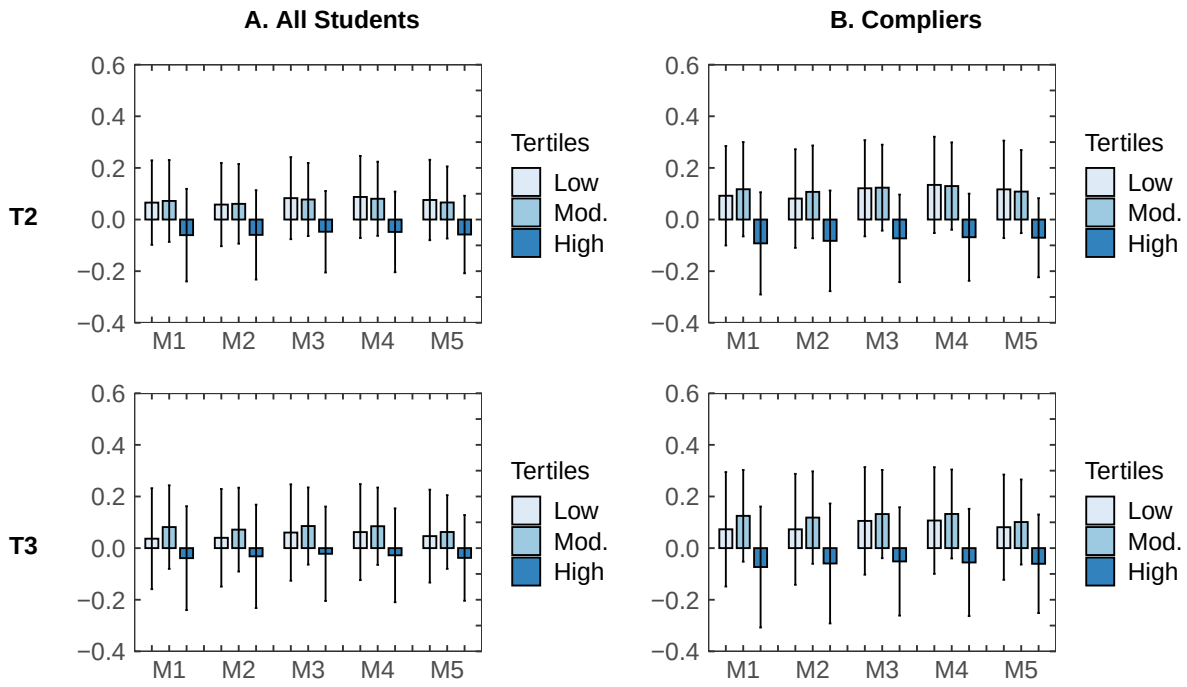

Figure A.10

Differential Impact on Growth Mindset by Baseline Mental Well-Being Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on growth mindset, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

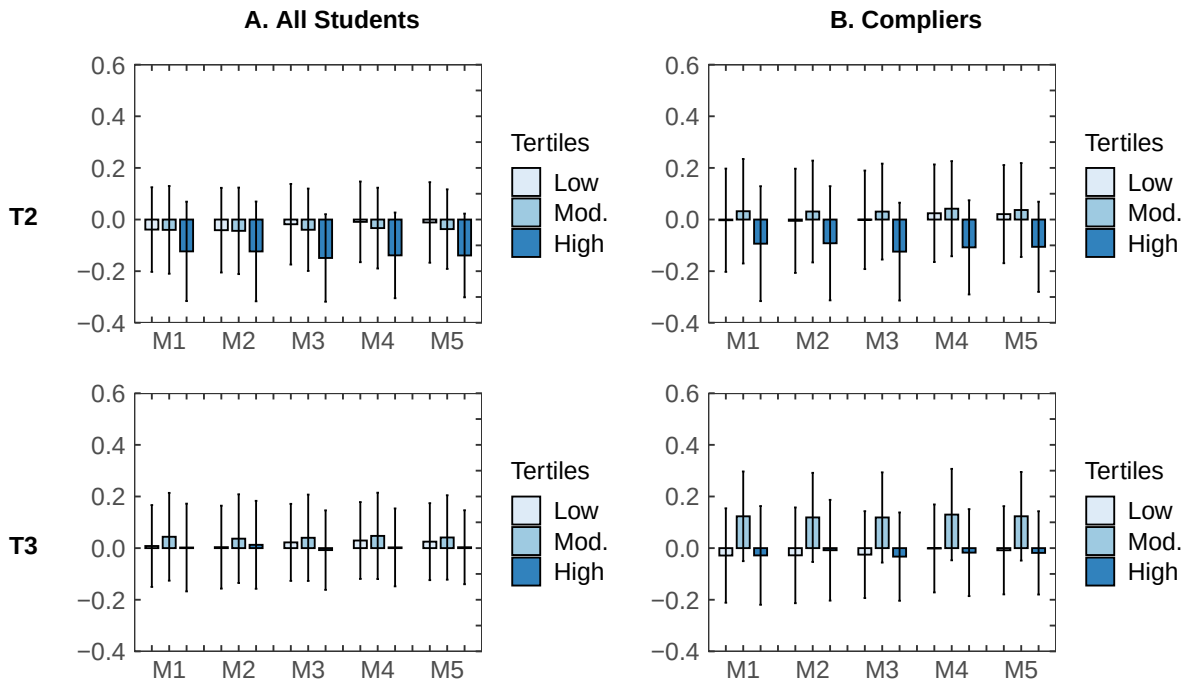

Figure A.11

Differential Impact on Problem Solving Skills Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on problem solving skills, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

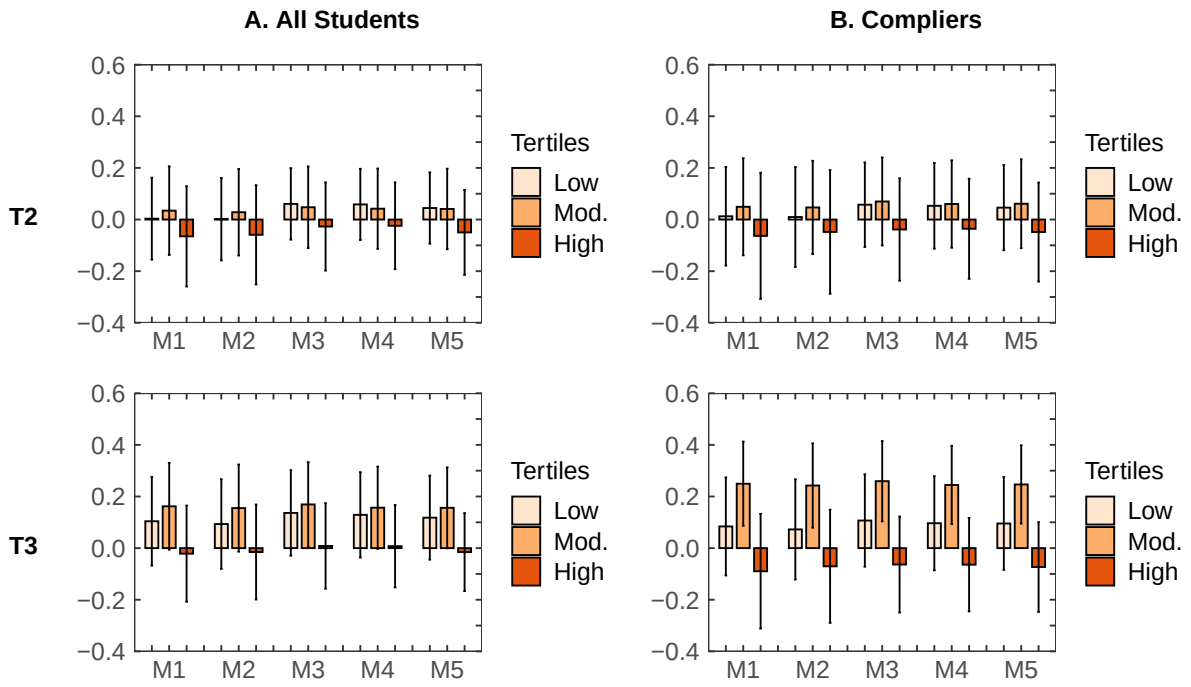

Figure A.12

Differential Impact on Motivation by Baseline Mental Well-Being Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on motivation, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

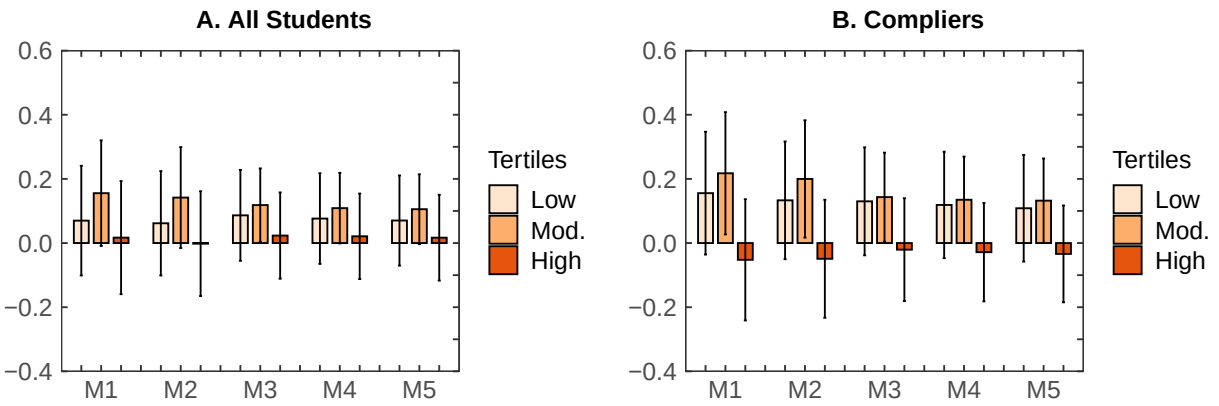

Figure A.13

Differential Impact on Math by Baseline Mental Well-Being Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on math, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

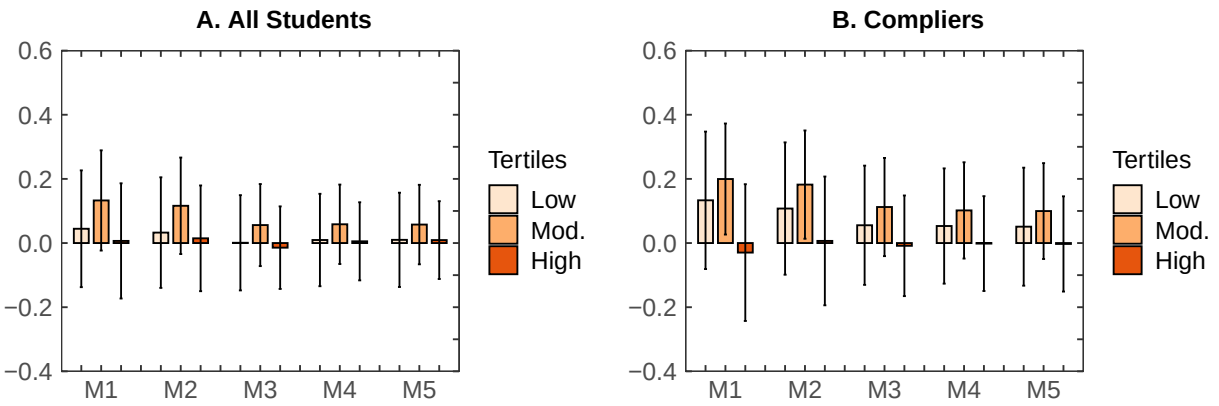

Figure A.14

Differential Impact on English Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on English, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

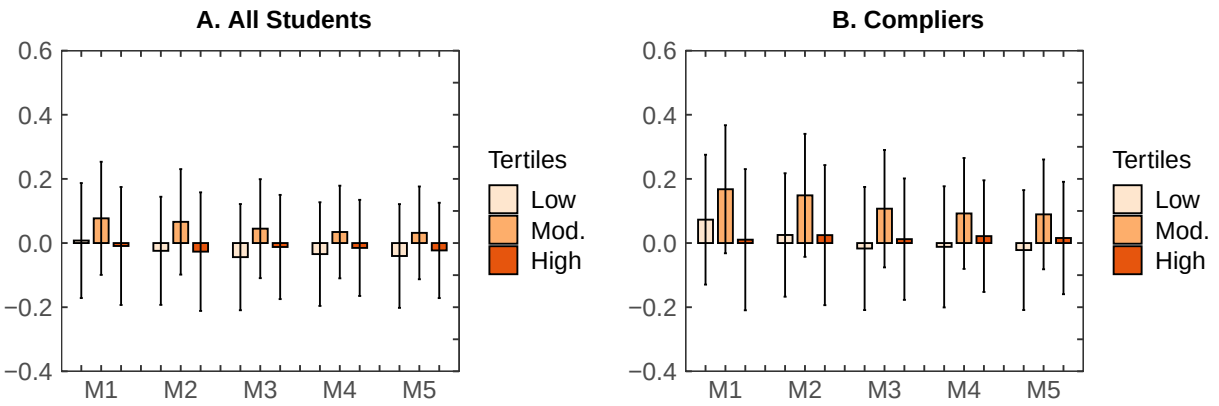

Figure A.15

Differential Impact on Norwegian by Baseline Mental Well-Being Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline mental well-being on Norwegian, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

By Baseline Emotional Distress

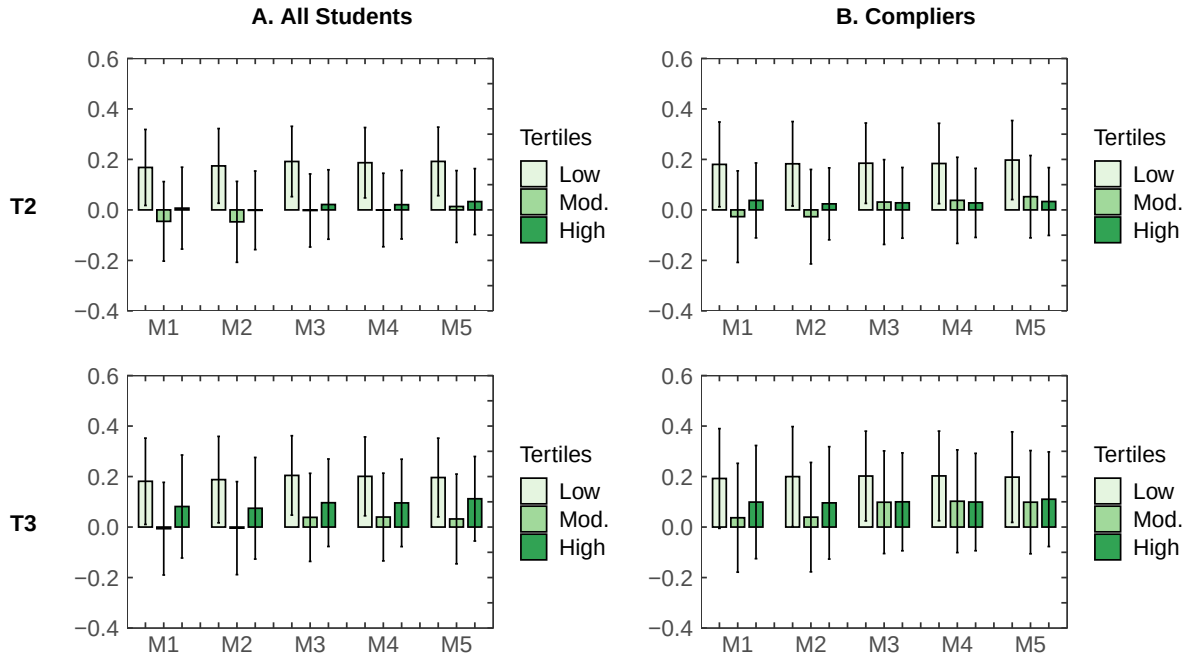

Figure A.16

Differential Impact on Mental Well-Being by Baseline Emotional Distress Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on mental well-being, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to mental health. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

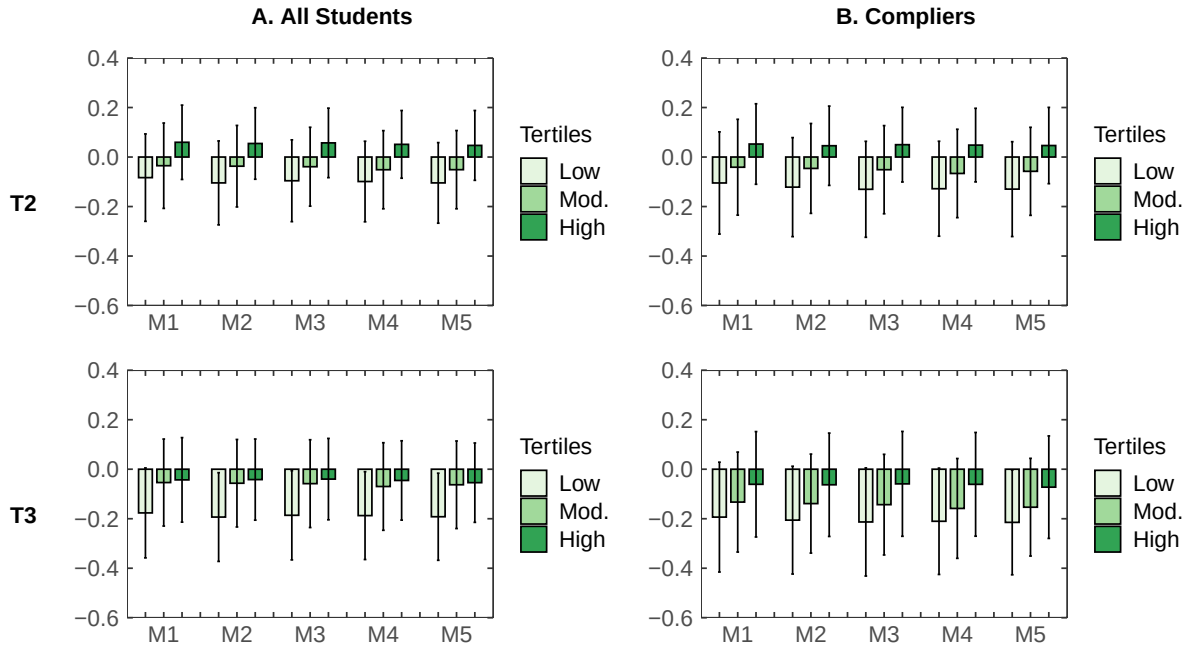

Figure A.17

Differential Impact on Emotional Distress by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on emotional distress, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to mental health. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

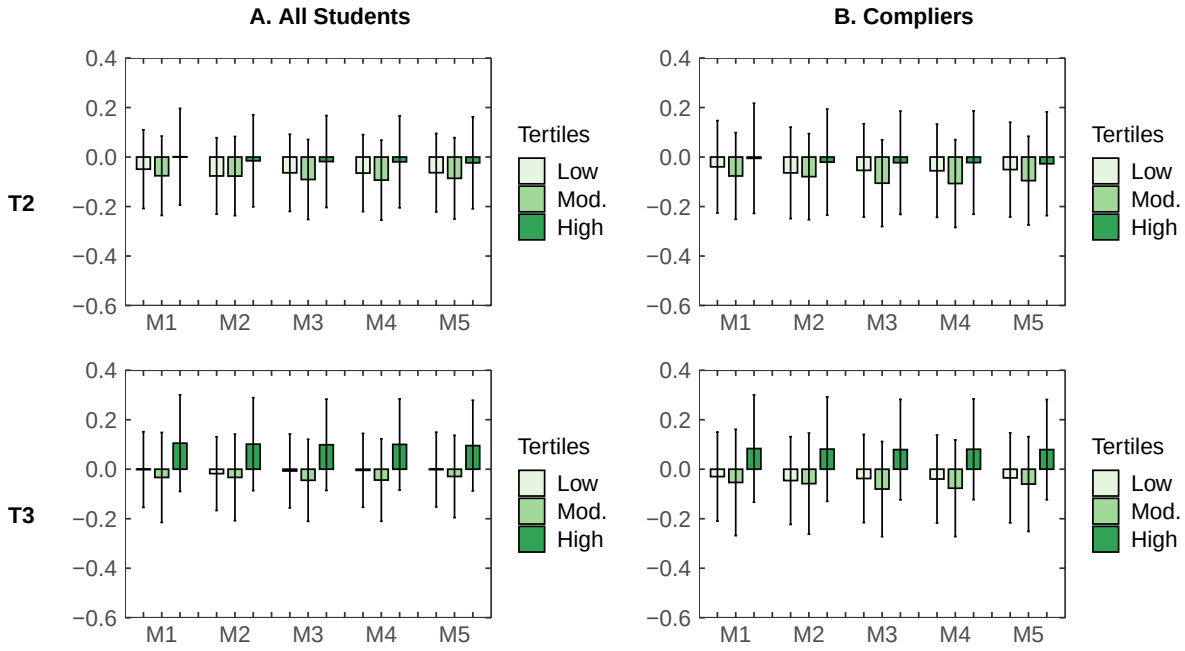

Figure A.18

Differential Impact on Academic Stress by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on academic stress, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to mental health. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

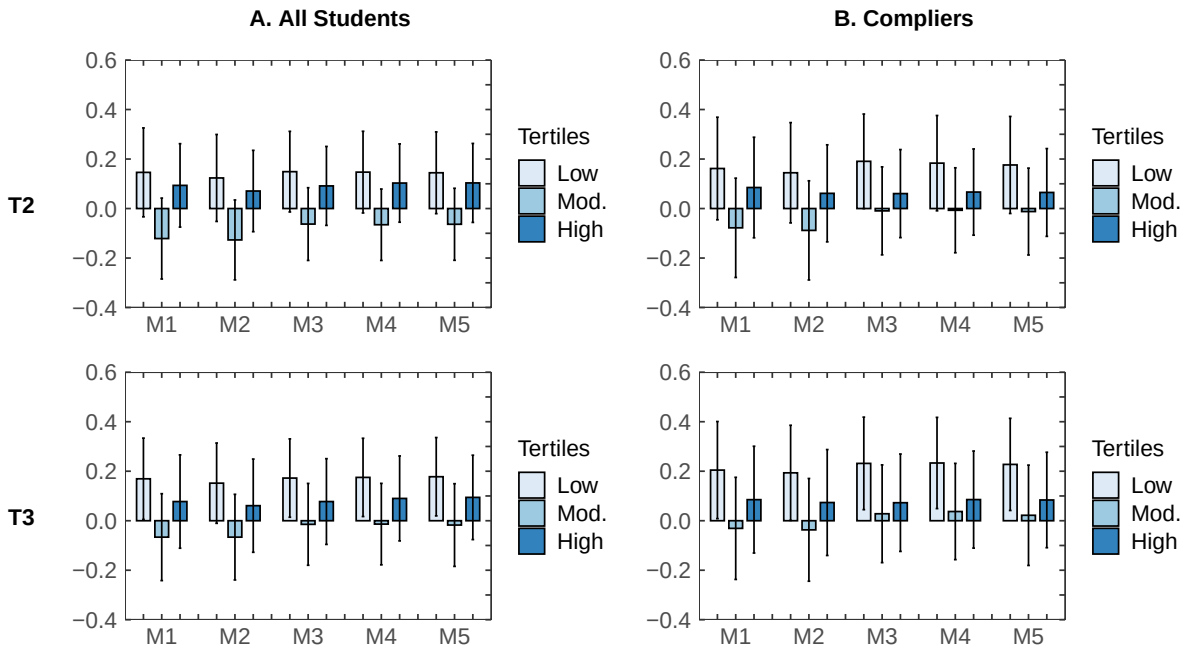

Figure A.19

Differential Impact on Relationship Skills by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on relationship skills, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

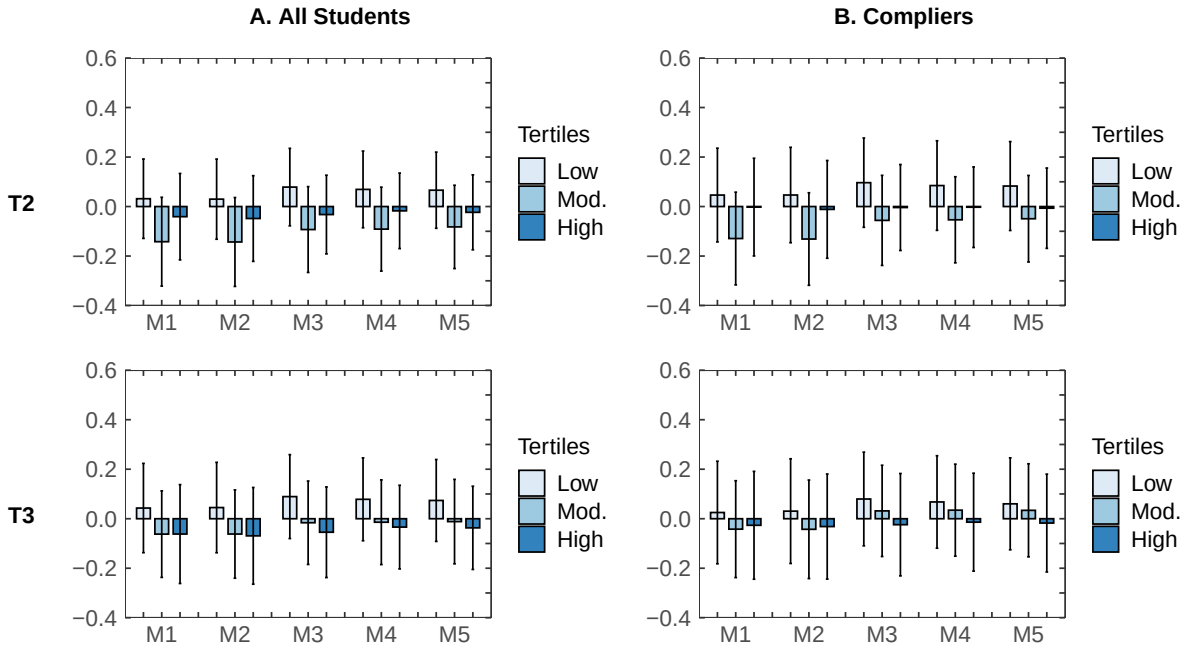

Figure A.20

Differential Impact on emotion regulation by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on emotion regulation, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

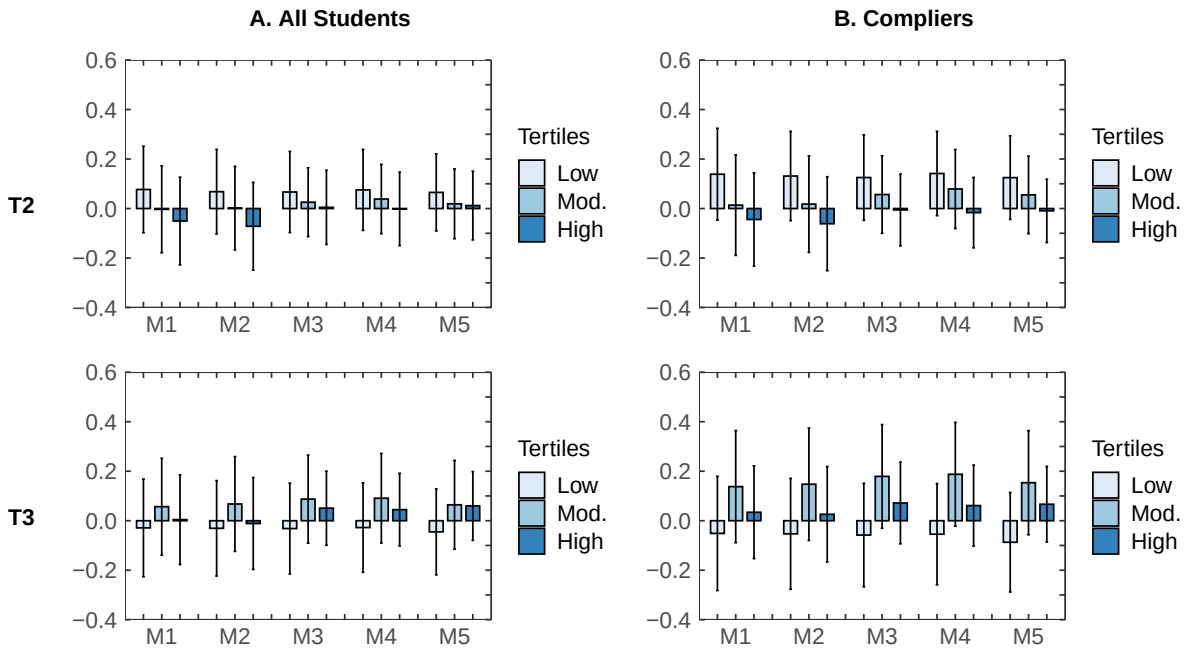

Figure A.21

Differential Impact on Growth Mindset by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on growth mindset, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

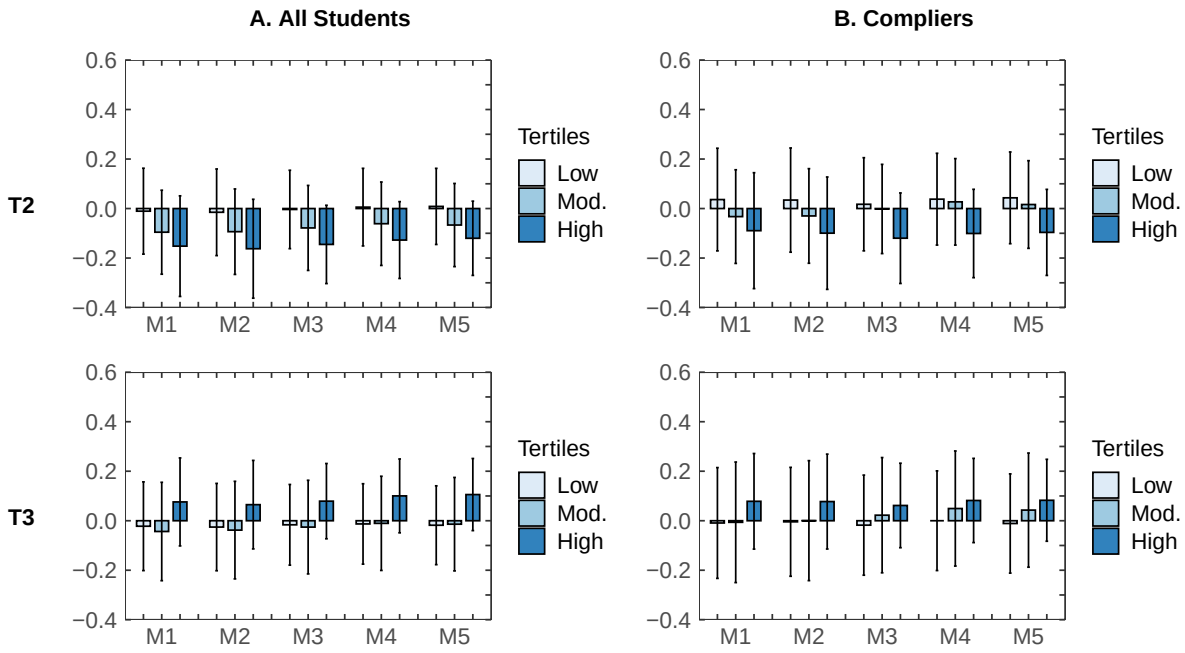

Figure A.22

Differential Impact on Problem Solving Skills by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on problem solving skills, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to coping skills. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

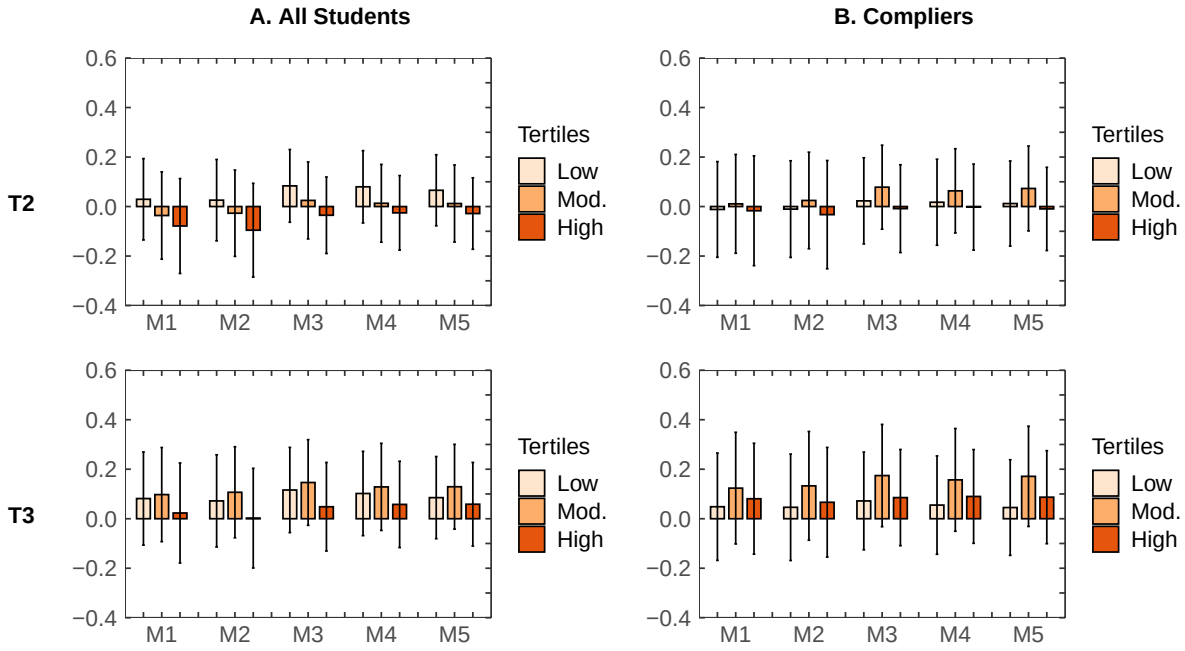

Figure A.23

Differential Impact on Motivation by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on motivation, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

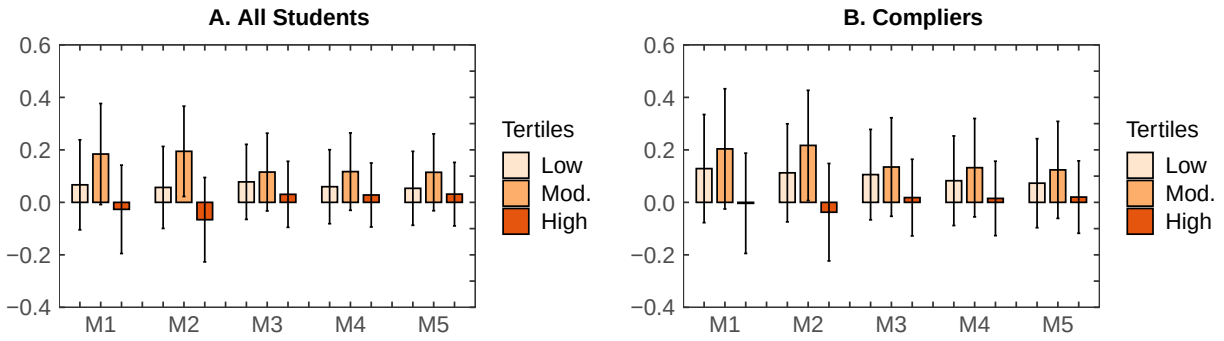

Figure A.24

Differential Impact on Math by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on math, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

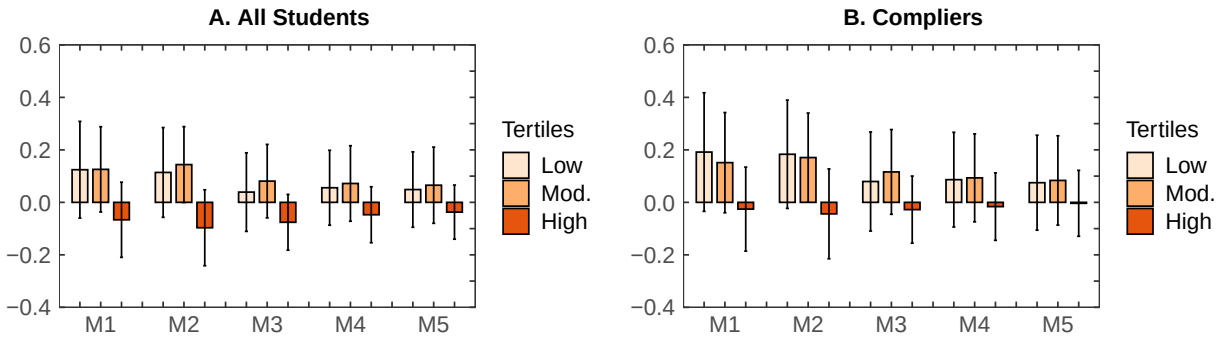

Figure A.25

Differential Impact on English by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on English, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

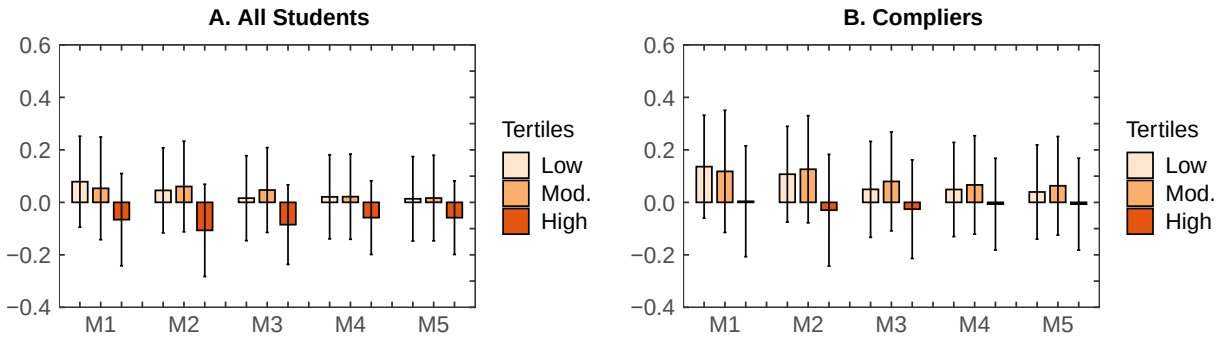

Figure A.26

Differential Impact on Norwegian by Baseline Emotional Distress Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline emotional distress on Norwegian, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

By Baseline Motivation

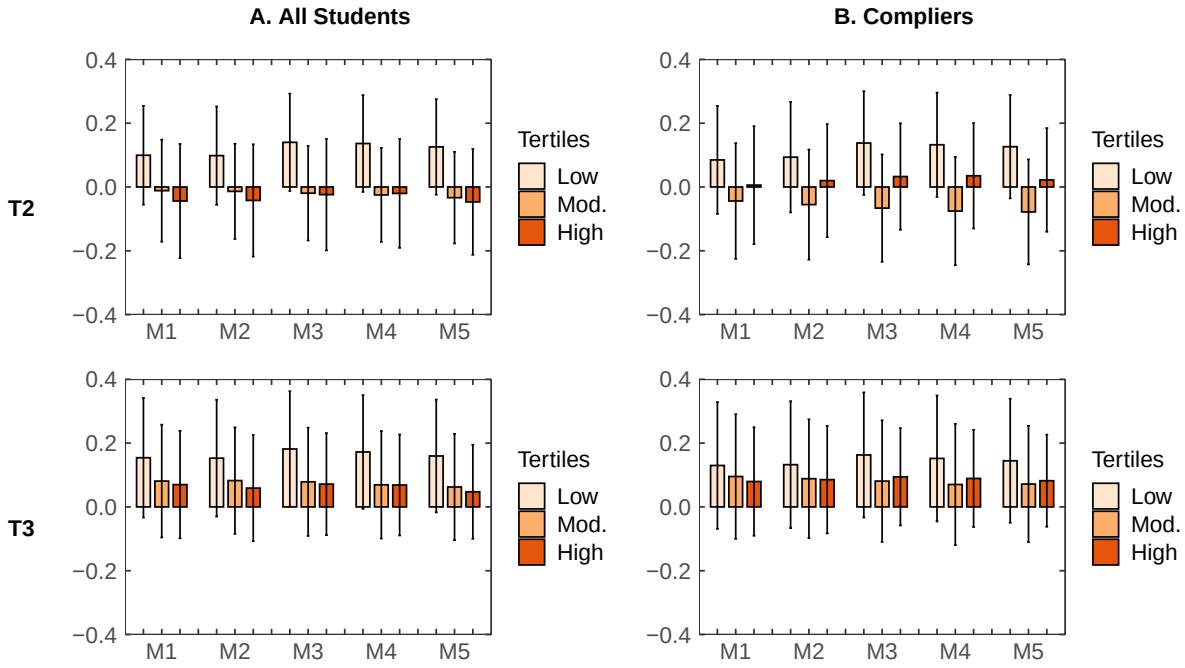

Figure A.27

Differential Impact on Motivation by Baseline Motivation Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline motivation on motivation, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

By Baseline Math

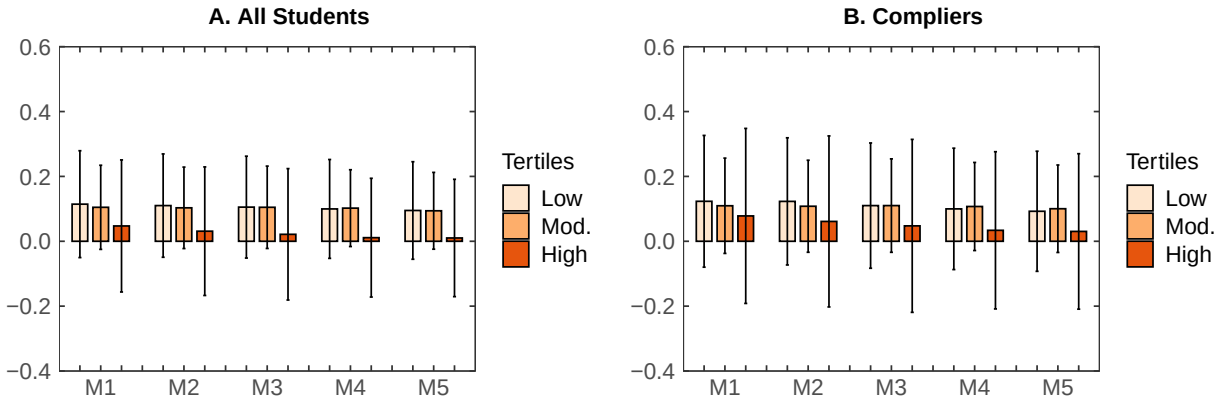

Figure A.28

Differential Impact on Math by Baseline Math Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline math on math, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

By Baseline English

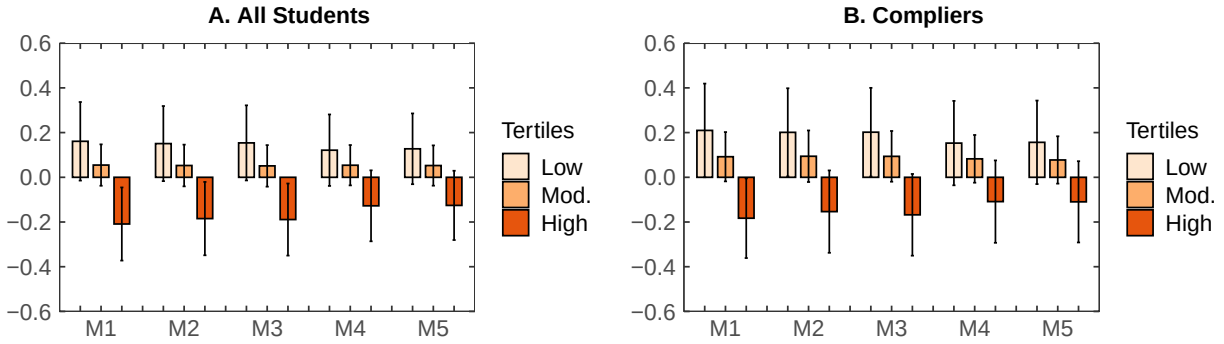

Figure A.29

Differential Impact on English by Baseline English Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline English on English, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

By Baseline Norwegian

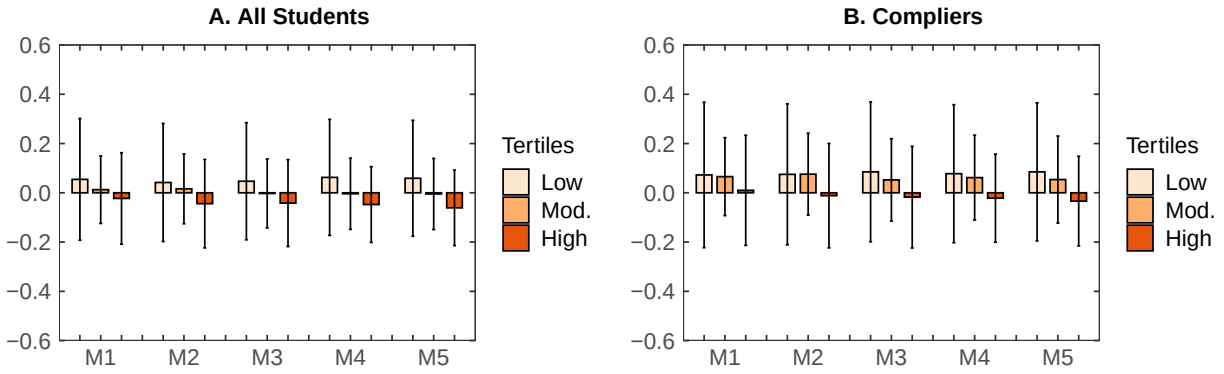

Figure A.30

Differential Impact on Norwegian by Baseline Norwegian Under Different Model Specifications

Note. This figure presents the marginal treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) by baseline Norwegian on Norwegian, across various model specifications. The first model includes block fixed effects (randomization blocks). The second model adds student and family characteristics to the block fixed effects. The third model incorporates block fixed effects, student and family characteristics, as well as the dependent variable measured at baseline. The fourth model extends this by including additional baseline measures related to motivation and school outcomes. Finally, the fifth model includes block fixed effects, student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. To address missing data, multiple imputation was applied, with statistics averaged across 100 imputed datasets and aggregated using Rubin's pooling rules (Rubin, 1987). T2 refers to the post-intervention assessment (Grade 9), while T3 represents the follow-up evaluation (Grade 10).

Treatment Effects Under Different Missing Data Handling Strategies

Table A.3
Treatment Effects Under Different Missing Data Handling Strategies

	Post-Intervention (T2)			Follow-up (T3)		
	MI	ML	Ind.	MI	ML	Ind.
	A. Mental Health					
Mental well-being	.080**	.071*	.069*	.113***	.112**	.112***
	(.037)	(.039)	(.041)	(.042)	(.044)	(.042)
Emotional distress	-.048	-.036	-.026	-.104**	-.100**	-.110**
	(.049)	(.040)	(.040)	(.046)	(.047)	(.048)
Academic stress	-.058	-.053	-.060	.021	.033	.024
	(.044)	(.043)	(.044)	(.052)	(.054)	(.056)
	B. Coping Skills					
Relationship skills	.061	.058	.054	.084**	.090**	.089**
	(.044)	(.045)	(.045)	(.043)	(.043)	(.041)
emotion regulation	-.014	-.018	-.023	.008	.005	.006
	(.037)	(.037)	(.041)	(.037)	(.036)	(.034)
Growth mindset	.032	.016	.010	.026	.009	.007
	(.039)	(.040)	(.043)	(.043)	(.041)	(.041)
Problem solv. skills	-.061	-.078**	-.076**	.023	.026	.020
	(.038)	(.037)	(.037)	(.041)	(.043)	(.044)
	C. Motivation and School Outcomes					
Motivation	.016	.008	.000	.091**	.085**	.088**
	(.035)	(.035)	(.037)	(.043)	(.043)	(.043)
Math				.067	.044	.062*
				(.043)	(.035)	(.035)
English				.026	.010	.014
				(.047)	(.035)	(.044)
Norwegian				-.009	-.052	-.031
				(.049)	(.044)	(.043)
Track				.038*	.037*	.042*
				(.022)	(.022)	(.022)
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note. This table presents treatment effects derived using various strategies for handling missing data. The analysis controls for block fixed effects, as well as student and family characteristics, along with baseline measures of academic achievement, mental health, and coping skills. Cluster-robust standard errors, adjusted for clustering at the class level, are reported in parentheses. Missing data are addressed through multiple imputation (MI), full-information maximum likelihood (ML), and the indicator method (Ind.), which replaces missing values in all independent variables with the mean and introduces a response indicator for each variable with missing data. Statistical significance is denoted ***, **, and * for the 1%, 5%, and 10% levels, respectively (two-tailed).

Treatment Effects Under Different Concave and Convex Transformations

Table A.4
Concave and Convex Transformations in Robustness Checks

											σ
Ref.	.01	1	2	3	4	5	6	7	8	9	10
1	3.95	1	.25	.063	.016	.004	.001	.000	.000	.000	.000
2	3.97	2	1.00	.500	.250	.125	.063	.031	.016	.008	.004
3	3.99	3	2.25	1.688	1.266	.949	.712	.534	.400	.300	.225
4	4.00	4	4.00	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000
1	4.92	1	.2	.04	.008	.002	.000	.000	.000	.000	.000
2	4.95	2	.8	.32	.128	.051	.020	.008	.003	.001	.001
3	4.98	3	1.8	1.08	.648	.389	.233	.140	.084	.050	.030
4	4.99	4	3.2	2.56	2.048	1.638	1.311	1.049	.839	.671	.537
5	5.00	5	5.0	5.00	5.000	5.000	5.000	5.000	5.000	5.000	5.000
1	5.89	1	.167	.028	.005	.001	.000	.000	.000	.000	.000
2	5.93	2	.667	.222	.074	.025	.008	.003	.001	.000	.000
3	5.96	3	1.500	.750	.375	.188	.094	.047	.023	.012	.006
4	5.98	4	2.667	1.778	1.185	.790	.527	.351	.234	.156	.104
5	5.99	5	4.167	3.472	2.894	2.411	2.009	1.674	1.395	1.163	.969
6	6.00	6	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000

Note. This table presents the sigma values used in the subsequent robustness analysis (see Bloem, 2022), which examines how alternative cardinalizations influence the estimated treatment effect. We apply convex and concave transformations using the following equation:

$$\mathrm{T}\{y\} = \max(y) \times \left[\frac{y}{\max(y)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{1}$$

where σ values are reported in the columns. The transformed values for each sigma are reported accordingly.

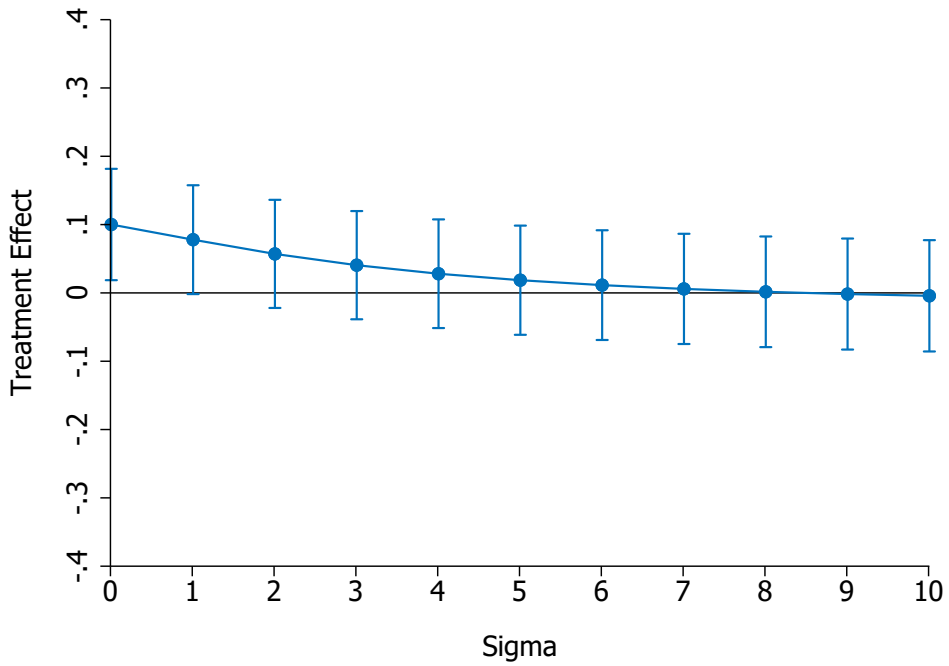

Figure A.31

Treatment Effects on Mental Well-Being (T2) Under Different Concave and Convex Transformations (All Students)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the mental well-being items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{2}$$

where $\mathrm{T}\{y_l\}$ is transformed mental well-being item l, $\max(y_l)$ is the maximum value of mental well-being item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

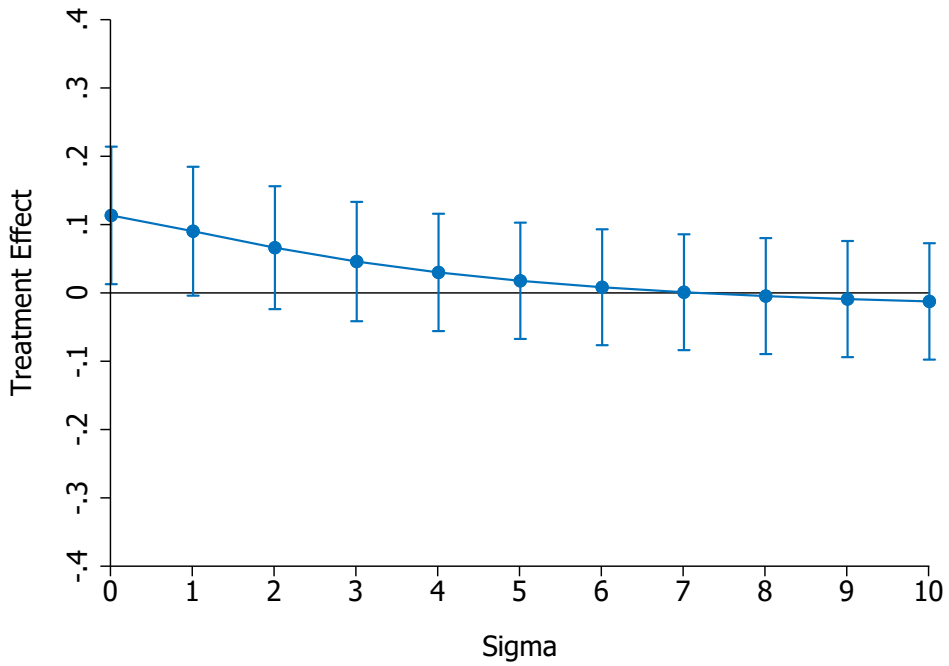

Figure A.32

Treatment Effects on Mental Well-Being (T2) Under Different Concave and Convex Transformations (High Reach Schools)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the mental well-being items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{3}$$

where $\mathrm{T}\{y_l\}$ is transformed mental well-being item l, $\max(y_l)$ is the maximum value of mental well-being item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

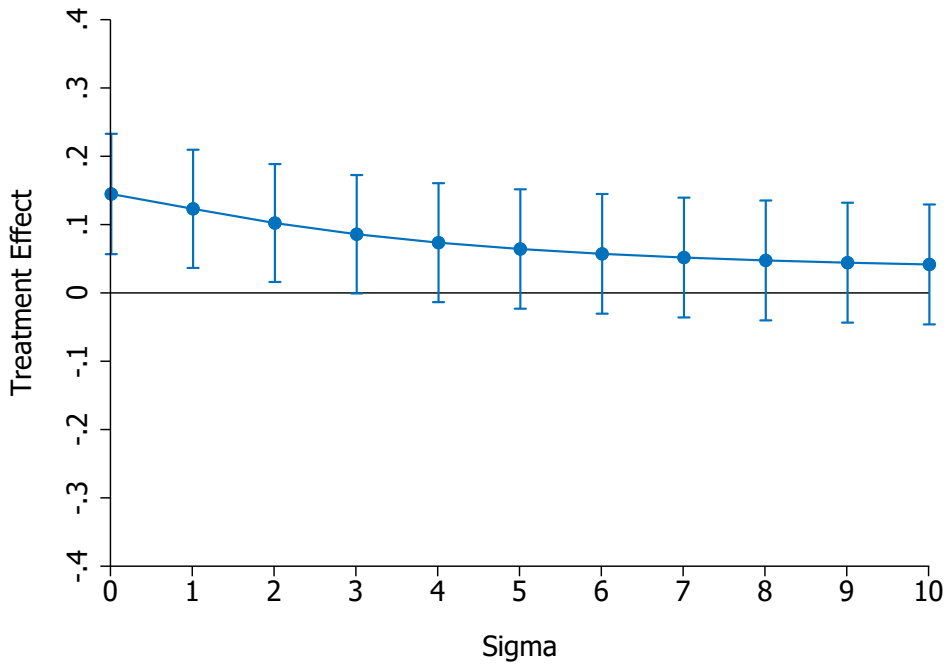

Figure A.33
Treatment Effects on Mental Well-Being (T3) Under Different Concave and Convex Transformations (All Students)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the mental well-being items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{4}$$

where $\mathrm{T}\{y_l\}$ is transformed mental well-being item l, $\max(y_l)$ is the maximum value of mental well-being item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

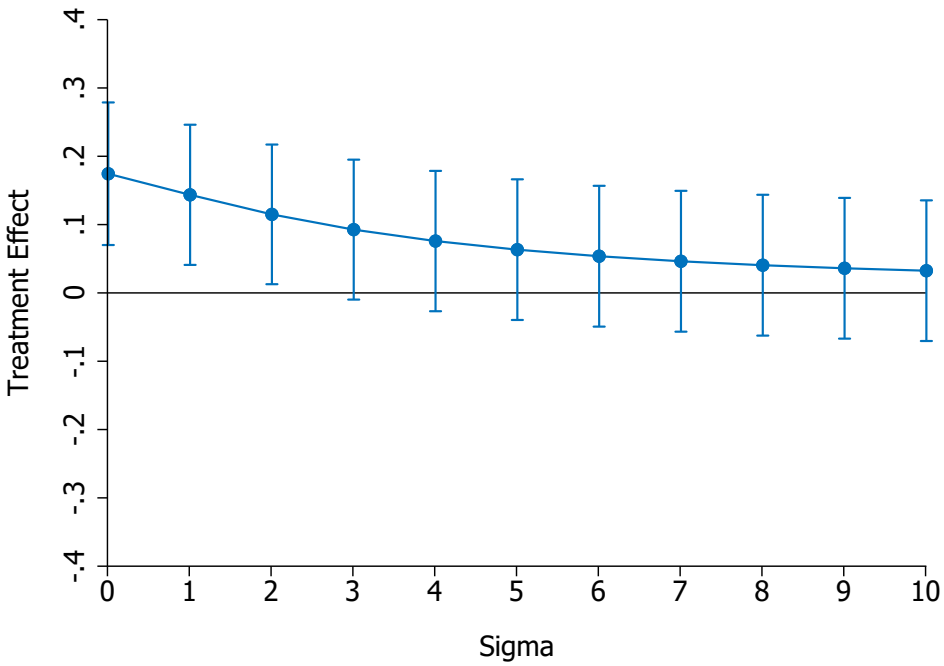

Figure A.34

Treatment Effects on Mental Well-Being (T3) Under Different Concave and Convex Transformations (High Reach Schools)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the mental well-being items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{5}$$

where $\mathrm{T}\{y_l\}$ is transformed mental well-being item l, $\max(y_l)$ is the maximum value of mental well-being item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

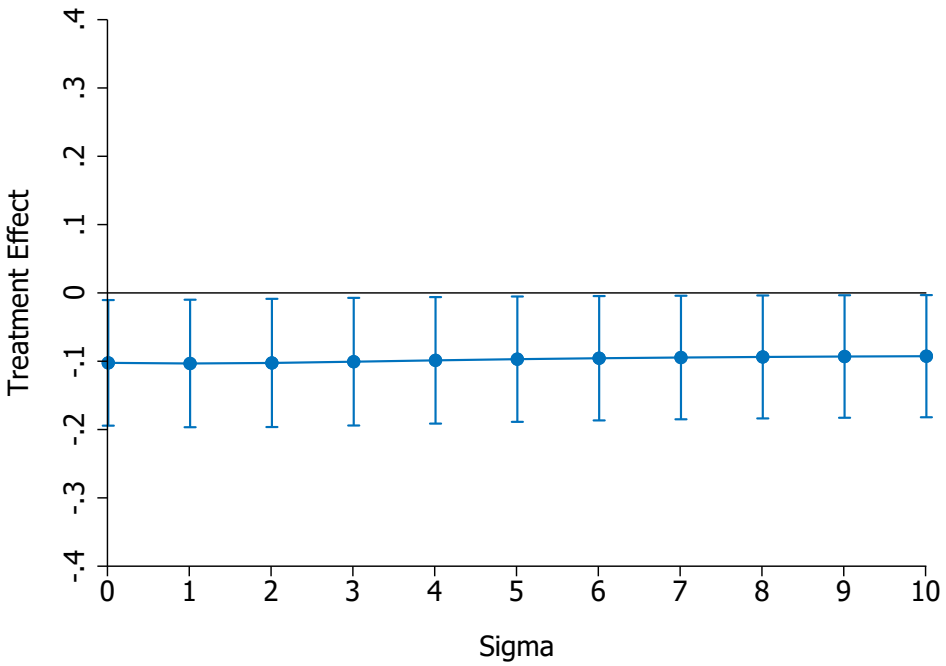

Figure A.35

Treatment Effects on Emotional Distress (T3) Under Different Concave and Convex Transformations (All Students)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the emotional distress items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{6}$$

where $\mathrm{T}\{y_l\}$ is transformed emotional distress item l, $\max(y_l)$ is the maximum value of emotional distress item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and convex transformations.

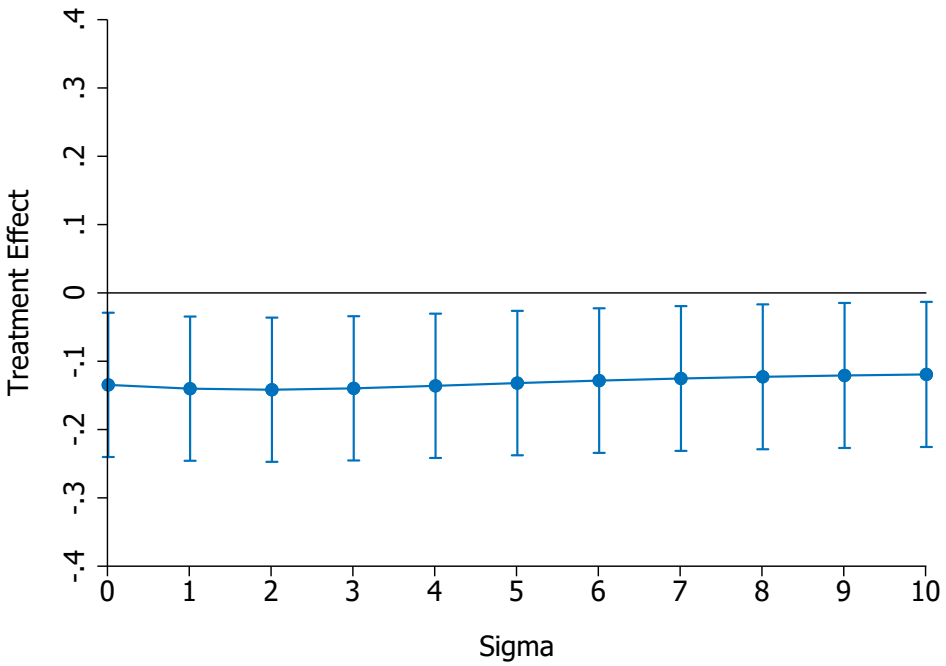

Figure A.36
Treatment Effects on Emotional Distress (T3) Under Different Concave and Convex Transformations (High Reach Schools)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the emotional distress items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{7}$$

where $\mathrm{T}\{y_l\}$ is transformed emotional distress item l, $\max(y_l)$ is the maximum value of emotional distress item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and convex transformations.

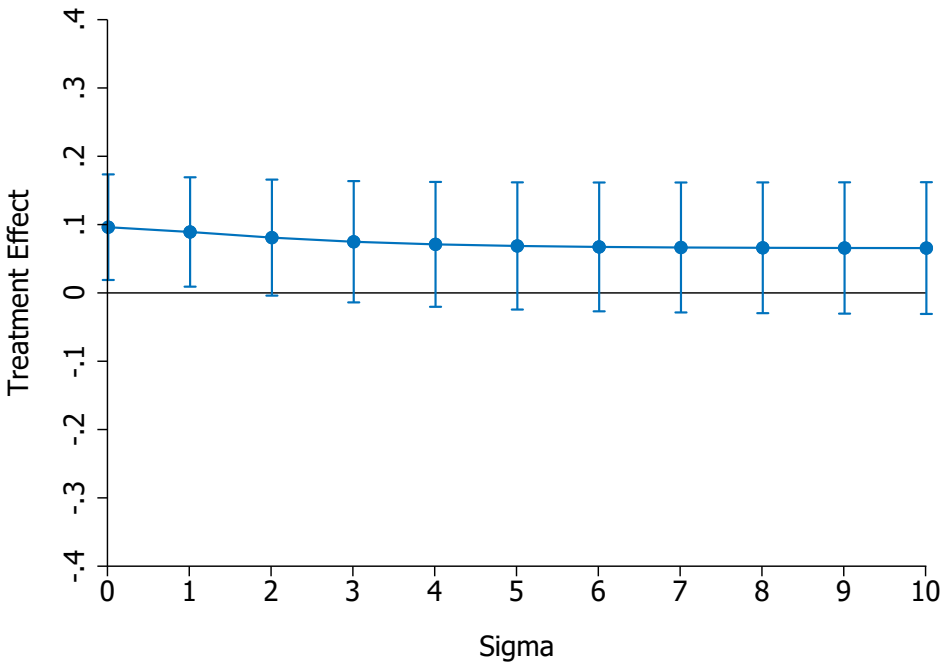

Figure A.37

Treatment Effects on Relationship Skills (T3) Under Different Concave and Convex Transformations (All Students)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the relationship skill items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{8}$$

where $\mathrm{T}\{y_l\}$ is transformed relationship skills item l, $\max(y_l)$ is the maximum value of relationship skills item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

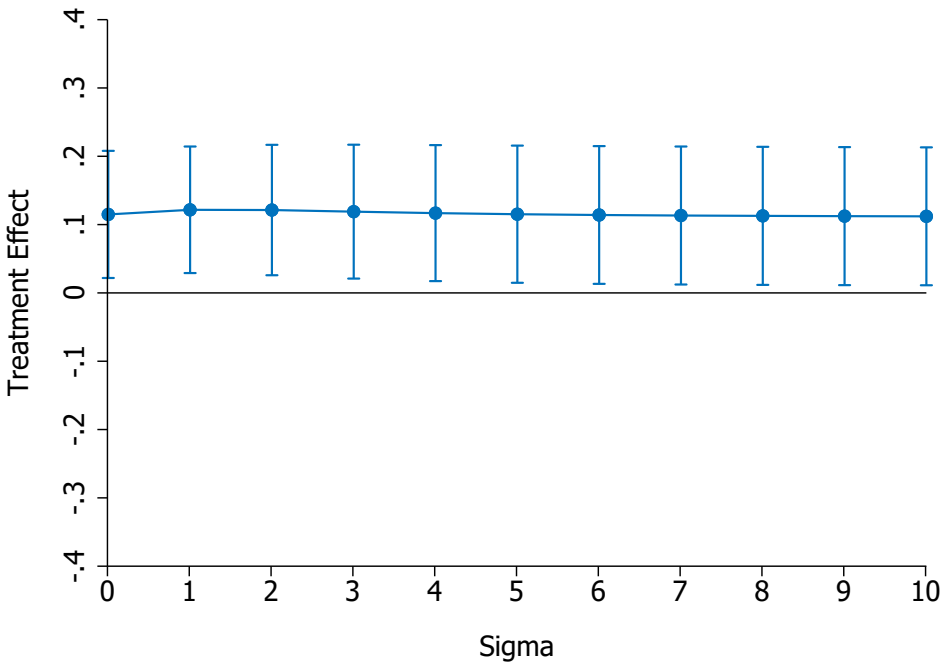

Figure A.38
Treatment Effects on Relationship Skills (T3) Under Different Concave and Convex Transformations (High Reach Schools)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the relationship skill items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{9}$$

where $\mathrm{T}\{y_l\}$ is transformed relationship skills item l, $\max(y_l)$ is the maximum value of relationship skill item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and convex transformations.

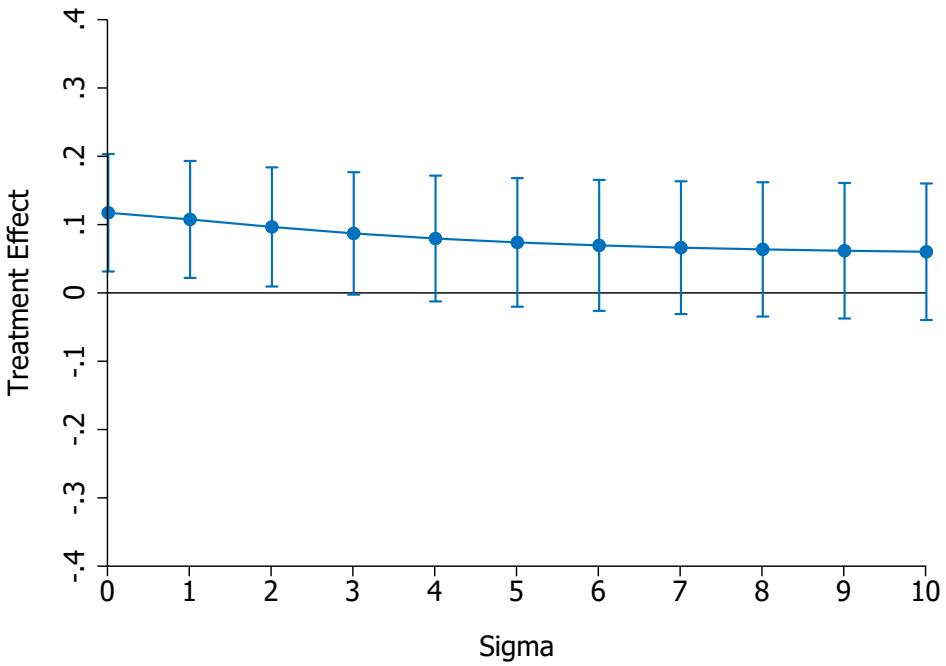

Figure A.39
Treatment Effects on Motivation (T3) Under Different Concave and Convex Transformations (All Students)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the motivation items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{10}$$

where $\mathrm{T}\{y_l\}$ is transformed motivation item l, $\max(y_l)$ is the maximum value of motivation item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

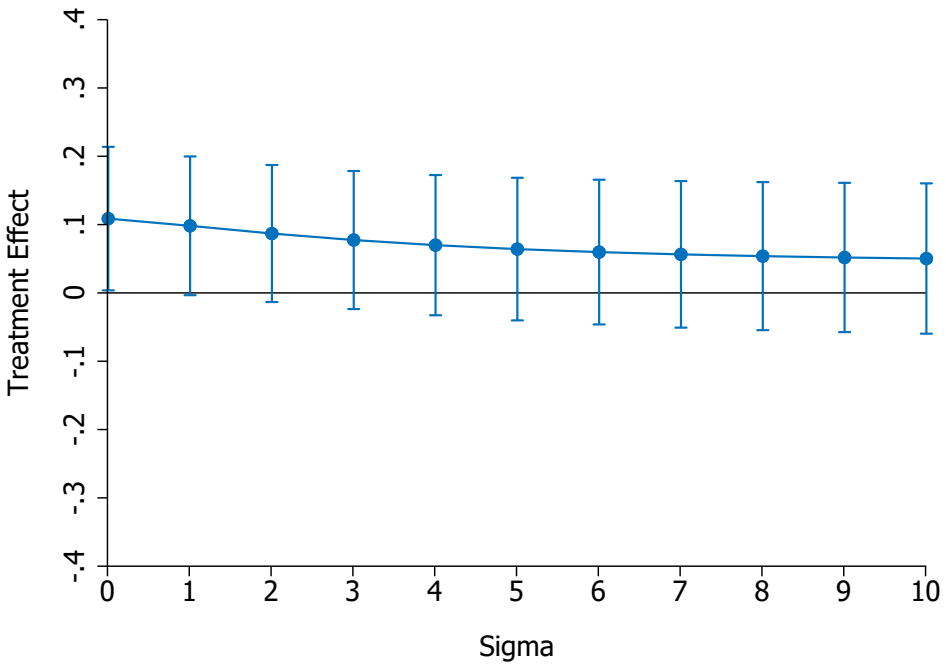

Figure A.40
Treatment Effects on Motivation (T3) Under Different Concave and Convex Transformations (High Reach schools)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the motivation items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{11}$$

where $\mathrm{T}\{y_l\}$ is transformed motivation item l, $\max(y_l)$ is the maximum value of motivation item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

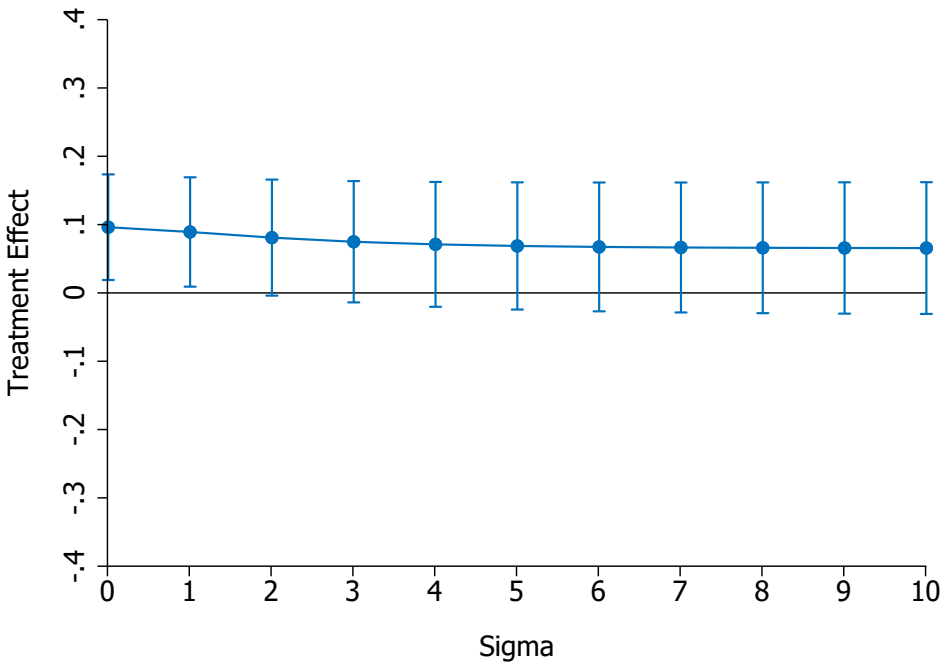

Figure A.41
Treatment Effects on Relationship Skills (T3) Under Different Concave and Convex Transformations (All Students)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the relationship skills items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{12}$$

where $\mathrm{T}\{y_l\}$ is transformed relationship skills item l, $\max(y_l)$ is the maximum value of relationship skills item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

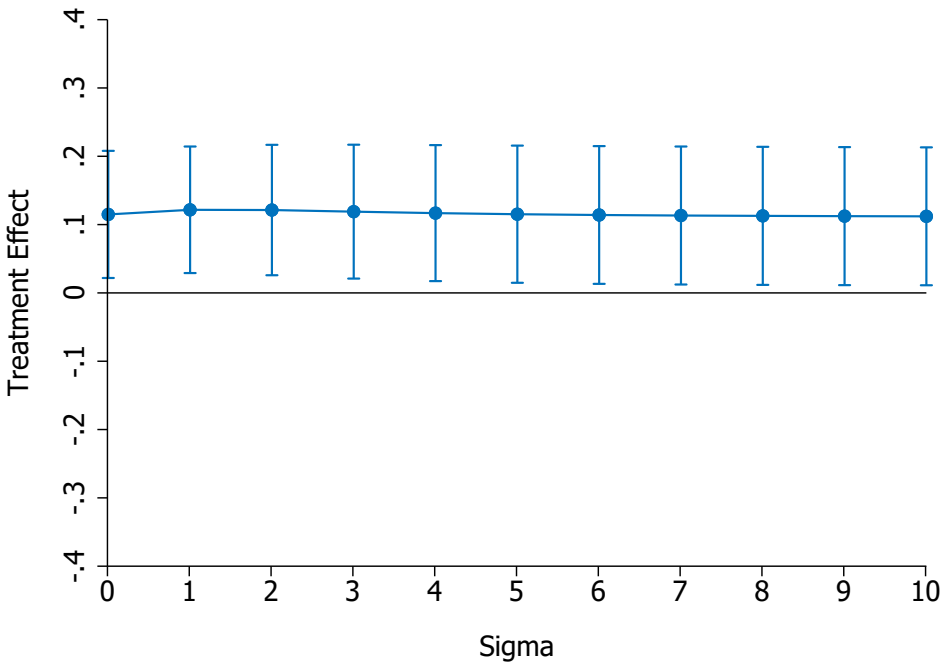

Figure A.42
Treatment Effects on Relationship Skills (T3) Under Different Concave and Convex Transformations (High Reach Schools)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the relationship skills items using,

$$\mathrm{T}\{y_l\} = \max(y_l) \times \left[\frac{y_l}{\max(y_l)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{13}$$

where $\mathrm{T}\{y_l\}$ is transformed relationship skills item l, $\max(y_l)$ is the maximum value of relationship skills item l, and σ measures the degree of monotonic transformation. Third, we standardize the transformed items to mean 0 and variance 1. Fourth, we compute an (arithmetic) average using the standardized (transformed) items. Fifth, we re-standardize the arithmetic average. Lastly, we estimate the treatment effect on the re-standardized average of (transformed) items, controlling for block fixed effects, student and family characteristics along the baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

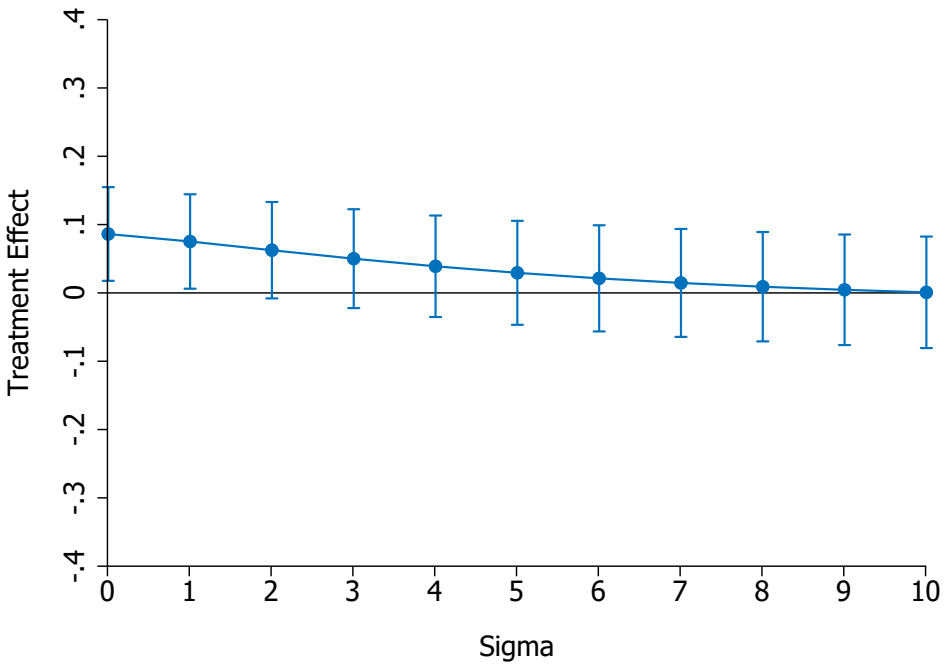

Figure A.43
Treatment Effects on Math (10th-Grade) Under Different Concave and Convex Transformations (All Students)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the math score using,

$$\mathrm{T}\{y\} = \max(y) \times \left[\frac{y}{\max(y)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{14}$$

where $\mathrm{T}\{y\}$ is the transformed math score, $\max(y)$ is the maximum value, and σ measures the degree of monotonic transformation. Third, we standardize the transformed math score to mean 0 and variance 1. Lastly, we estimate the treatment effect on the standardized (transformed) score, controlling for block fixed effects, student and family characteristics along with baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.

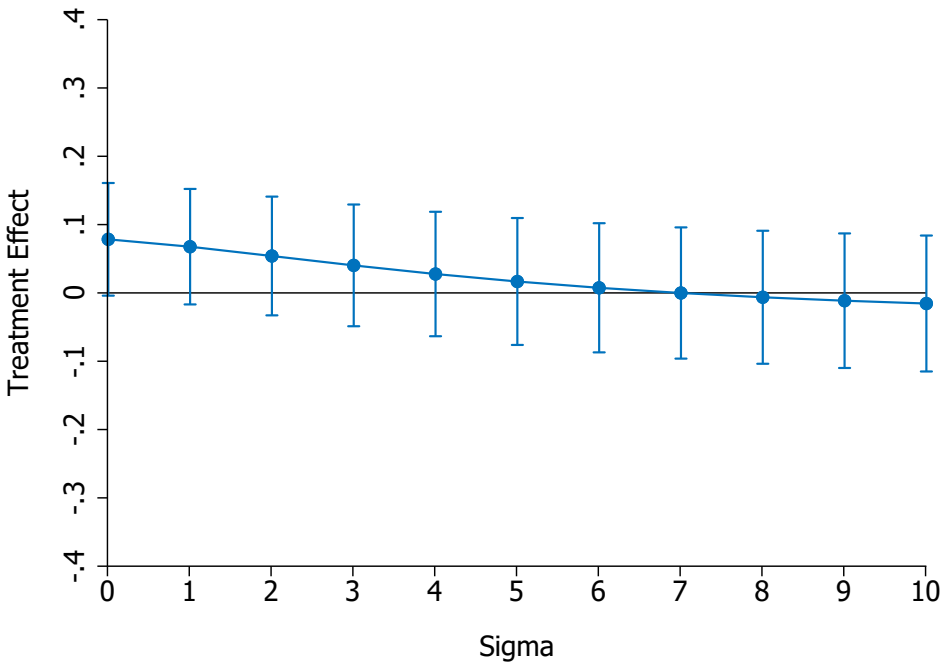

Figure A.44
Treatment Effects on Math (10th-Grade) Under Different Concave and Convex Transformations (High Reach Schools)

Note. This figure visualizes the treatment effect (and 95% confidence intervals based on cluster-robust standard errors, clustered at the class level) under alternative cardinalizations based on Bloem (2022). Refer to Table A.4 for an overview of the cardinalization implied by the different values of σ. We construct the plot as follows. First, we choose σ. Second, we transform the math score using,

$$\mathrm{T}\{y\} = \max(y) \times \left[\frac{y}{\max(y)}\right]^{\sigma} \quad \forall\ \sigma > 0, \tag{15}$$

where $\mathrm{T}\{y\}$ is the transformed math score, $\max(y)$ is the maximum value, and σ measures the degree of monotonic transformation. Third, we standardize the transformed math score to mean 0 and variance 1. Lastly, we estimate the treatment effect on the standardized (transformed) score, controlling for block fixed effects, student and family characteristics along with baseline measure of the dependent variable. To address missing data, the analysis employs the indicator method, as it avoids the computational burden associated with multiple imputation. Using multiple imputation would require running the data transformation process separately for each imputed dataset, making it considerably more resource-intensive. Despite this, the indicator method produces results comparable to those obtained with multiple imputation when estimating the main treatment effects (see Table A.3). The results remain consistent under concave transformations and certain convex transformations.