A Meditation on Multidisciplinarity, in the Context of a School-Based Meditation Intervention

Sarah Collier Villaume
Northwestern University

Aurelie Ouss
University of Pennsylvania

Jonathan Guryan
Northwestern University and IPR

Emma Adam
Northwestern University and IPR

Version: January 25, 2024

DRAFT
Please do not quote or distribute without permission.
Abstract

In this paper, the researchers share insights from a multidisciplinary collaboration between developmental psychologists and economists. Together, they designed and implemented a randomized controlled trial (RCT), in a school-based setting, of the effects of transcendental meditation on perceived and biological indicators of stress, as well as health and academic outcomes. Through the lens of the study that resulted, they describe the process, challenges, and advantages of their collaboration. They also present novel results on the effect of the meditation RCT on one physiological marker of stress and health: blood pressure. Overall, intent to treat impacts were modest and not statistically significant for the full sample; they observed an impact on systolic blood pressure (SBP) of -2.37 mm Hg (0.21 SD). Additionally, blood pressure reductions were large and statistically significant for certain subgroups, including for Black students and those with obese body mass index, two groups at elevated risk for early onset of hypertension. In addition to substantive results, the researchers offer insights and recommendations for others who may take on multidisciplinary research in the future.

A study like this one relies on the work of many partners to be successful. The authors thank the research teams in the COAST Lab, especially Ednah Nwafor, and at the University of Chicago Crime and Education Labs, including John Wolf and Jaureese Gaines. Data collection would not have been possible without Olga Nehme and Mike Weisensee. They are most grateful for the school leaders and teachers who supported their study’s activities, the parents who contributed their time and valuable perspectives, and the students who participated in the study. The authors also thank four anonymous reviewers for thoughtful feedback on prior versions of this manuscript. The Russell Sage Foundation (award G-6650) supported the stress sub-study discussed here. Funding from the Institute of Education Sciences supported SCV’s time for data analysis and manuscript preparation (R305B140042).
When a researcher is approached by someone from another discipline to collaborate, the easiest answer is “no.” There are many bridges to cross in interdisciplinary collaboration, including differences in: language, assumptions, research design preferences, funding sources, and writing styles and norms, among others. But such collaborations also provide a unique opportunity. Each discipline brings its own advantages (and disadvantages). An effective multidisciplinary collaboration can bring together the knowledge bases, theories, and analytical tools of multiple disciplines to create a research product that is better than would have been produced by any one discipline alone. In the current study, we describe how economic and developmental psychology researchers came together to design and implement a rigorous randomized controlled trial (RCT), in a school-based setting, of the effects of transcendental meditation on perceived and biological indicators of stress, as well as health and academic outcomes. We also present novel results on the effect of the meditation RCT on one physiological marker of stress and health: blood pressure.

The story starts in 2016, when an economist (JEG) approached a developmental psychologist (EKA) who specializes in stress biology. He described an ongoing, school-based randomized controlled trial of transcendental meditation. The RCT was fully funded and set to take place in several public high schools, all in low-income neighborhoods (Guryan & Ouss, 2019). What could we measure, he asked, that would help researchers understand how meditation was affecting the biological stress levels of these adolescents? At first, the team of economists simply asked for consulting, or expertise, from the developmental psychologists. But all parties quickly realized that a true collaboration, rather than consultation, would be a more promising route. No one discipline, or its methods, can be easily taught to another research group without deep, time-intensive engagement on both sides. The newly formed multidisciplinary
team began designing a study that was: justified by past theory and research, acceptable to the scientific communities of both disciplines, appropriate to the population of interest (low-income adolescents), acceptable to the local community, poised to yield meaningful insights, and feasible in a school-based setting.

What brought these team members together was shared interest in a scientific problem, specifically the issue of stress in adolescence. Excess stress exposure is a documented concern for young people, which can have implications for physical health and brain development, for mental health, and for academic achievement, all concerns germane to both developmental psychology and economics (Adam, Klimes-Dougan, & Gunnar, 2007; Smith, 1999). This topic is especially relevant to young people whose family or community context includes experiences of poverty, structural violence, or racism, among other destabilizing influences. There are stark differences in stress exposure, which we call stress disparities, across demographic groups and other marginalized identities in the US (Levy, Heissel, Richeson, & Adam, 2016; Su et al., 2015).

Stress disparities can be found in differential exposure to psychosocial stressors (Dimsdale, 2008; Rosengren, 2004), racial discrimination (Krieger & Sidney, 1996), poverty (Lippert, Evans, Razak, & Subramanian, 2017; Wickrama, O'Neal, & Lott, 2012), and violence exposure (Heissel, Sharkey, Torrats-Espinosa, Grant, & Adam, 2017), each of which has been associated with elevated blood pressure (Ford & Browning, 2014; Mayne et al., 2018) and other stress-related biomarkers in adolescents and adults (Adam, Collier Villaume, Thomas, Doane, & Grant, 2023). Additionally, the toll of stress exposure can be exacerbated by a lack of control over the situation (Steptoe & Willemsen, 2004; Vrijkotte, Van Doornen, & De Geus, 2000) and by responses to stress that include internalizing feelings of distress or anger (Krieger & Sidney,
Social status as well as individual resources and supports are inextricably linked to the amount of stress experienced—and its impacts on health risk.

For this reason, researchers have sought interventions that may decrease young people’s exposure to, perception of, or management of stress. One promising approach that has been proposed relates to meditation and mindfulness (Lane, Seskevich, & Pieper, 2007; Turakitwanakan, Meksepralard, & Busarakumtragul, 2013). Work in both laboratory and clinical settings suggests that meditation can reduce blood pressure for adults (Schneider et al., 2012) as well as in studies of high school or college students at elevated risk for hypertension (Barnes, Prendergrast, Harshfield, & Treiber, 2008; Barnes, Treiber, & Johnson, 2004; Gregoski, Barnes, Tingen, Harshfield, & Treiber, 2011; Nidich et al., 2009). But a key question is how to make an individual practice like this more widespread. If deployed at a larger scale, would it still be effective? Could it be scaled up affordably?

These questions are central to the concerns of economists, especially those who study interventions in areas of education or health. Such questions were unanswered by the existing evidence base on meditation, which largely focuses on adults, including long-term practitioners of meditation (Sudsuang, Chentanez, & Veluvan, 1991) and individuals recruited on the basis of a specific health condition (Carlson, Speca, Faris, & Patel, 2007; Hughes et al., 2013; Schneider et al., 2012). In response to this gap in the literature, the economists on our team had identified a compelling research design conducive to investigating the effects of meditation for healthy young people. They have experience conducting RCTs in school-based settings and were able to secure funding to offer meditation at a large scale through the school setting. Experts in meditation and mindfulness would be present daily in the school, providing training and ongoing support to interested students. The team planned to assess the program’s effectiveness by looking
at administrative records, like school grades and discipline data that could be obtained through an agreement with the school district. But another problem remained: administrative data would not allow our partners to assess whether the intervention had positive impacts on health—or whether any benefits observed operated by way of changes in stress and biological stress system activity. Biological measures of stress are objectively measured (not self-reported) stress indicators. They can also be investigated as potential mechanisms or pathways linking environmental exposures to health outcomes. This problem and the advantages of biological stress measurement motivated the economist team to seek out developmental psychologists with expertise in developmental psychobiology as collaborators. Physical proximity also played a role, as [A3] and [A4] are faculty members in the same PhD program, who knew of each other’s work and had offices located near one another.

The next step in this collaboration was to get the stress biology sub-study funded: several proposals were submitted, and one of our four attempts was successful. Not surprisingly, the project was funded by a foundation (the Russell Sage Foundation) that (1) understands the importance of interdisciplinary research and (2) has access to reviewers who understand and appreciate the language, theoretical and analytic approaches of both disciplines.

At this point, the reader might wonder, what does it mean to have a collaboration between economics and developmental psychology? How does each discipline contribute? While we tend to avoid stereotyping members of a discipline, given the great variability in methodological approaches within, as well as between, disciplines, there are some features that one can broadly recognize as coming from economics and others from psychology. Some of these are detailed in a vocabulary list (Appendix 1) that will help the reader translate the assumptions and terminology used by each discipline in this paper. A few key features of
MEDITATION ON MULTIDISCIPLINARITY

economics include a central focus on causal identification (isolating, as cleanly as possible, whether variable x has a causal association with variable y), a series of sophisticated analytic techniques designed to reach that goal of causal identification (Angrist & Pischke, 2010), and a tendency to conduct a series of robustness tests to help rule out alternative explanations (Oster, 2019). Some key features of developmental psychology and developmental psychobiology include a central focus on good psychometric measurement of variables, considering the role of age and/or developmental stage in both theoretical and statistical models, placing greater emphasis on insights from correlational research, sophisticated data-analytic methods designed to maximize the insights gleaned from longitudinal research, and a focus on how varying ecological contexts matter for the processes they are observing. The emphasis on contextual determinants of development also leads to a focus on the importance of ecological validity in developmental psychology (Shiffman & Stone, 1998), even relative to other branches of psychology.

Having set the disciplinary stage, we now turn to describing the design of our study, how our disciplinary approaches informed the design, and redesign, or our study.

**Researcher positionality, reflexivity, and power sharing**

When a research team sets about raising awareness or soliciting buy-in in preparation for a planned study, they may learn of concerns that had not previously been identified. In the case of our study, an advisory parent group at the school had been consulted on the research design and had responded favorably to it. But when we later scheduled a series of meetings to share information about the study with a broader audience of parents, we began to learn of parent concerns about the risks associated with participating in research for the youth in this school. Beyond a general hesitation about participating in research, parents’ primary concern related to
the collection of salivary cortisol, which had been the primary outcome measure in our original study design. The developmental psychobiologists in our group have considerable expertise and experience in the collection of salivary cortisol data in naturalistic settings; our research group has done so successfully with participants from a range of backgrounds and age groups (e.g., Adam, 2006; Adam, Hawkley, Kudielka, & Cacioppo, 2006). While participants are at times squeamish about the process of spitting into a tube, none in our experience had ever voiced serious concerns that providing saliva might present a risk for them. But this group of parents asked whether the saliva samples we collected could be subpoenaed by police or their children’s DNA otherwise utilized for purposes other than those for which they had consented. Given the rapidly moving timeline of the study, we did not have time to seek a Certificate of Confidentiality, a mechanism administered by the National Institutes of Health which “prohibits disclosure” of identifiable or sensitive information “in response to legal demands, such as a subpoena” (NIH, 2023 see also, Beskow, Dame, & Costello, 2008).

A reflexive process (Wilkinson, 1988) allowed us to recognize that these parents’ feedback came from a different perspective than our own. The parents almost all self-identified as Black or Latino. Their neighborhoods and communities have experienced histories of structural violence and other chronic stressors. Additionally, there is a documented history of research misconduct that has harmed members of historically minoritized groups, as Schraff and colleagues discuss (2010). This history may contribute to community members’ reluctance to engage in research (George, Duran, & Norris, 2014). In this case, after hearing the parents’ input, it was clear to us that the worldview they offered was an important one to inform this research (Jamieson, Govaart, & Pownall, 2023). Instead of attempting to reassure parents and move forward with our original plan, our team concluded that it was most appropriate to propose a
change in the study’s design. We suggested changing the study’s outcome measures and observing adolescents’ sleep and blood pressure instead of collecting cortisol data. This approach would allow us to obtain high-quality measures of biological stress system functioning without physically taking any biological samples from participants. It also allowed for power sharing, a component of Community-Based Participatory Research (CBPR) and the literature on co-creation in implementation science research (Goodyear-Smith, Jackson, & Greenhalgh, 2015; Perez Jolles et al., 2022). In this case, power sharing meant entrusting authority over the study’s outcome measures to the school community (specifically the parents). We believed the parents were the most important experts on the type of research that would be acceptable to their community and would go beyond the minimum standard of not doing harm. When presented with this alternative study design, parents voiced surprise at our willingness to listen to them. They were both supportive of and interested in these measures—in part due to their greater understandability, face validity, and direct relevance to adolescent health (especially measures of blood pressure).

Following these community conversations, our team took additional steps to ensure that families received a direct benefit from their child’s participation in research (beyond monetary compensation, which was also provided). We offered to prepare a “health report” for all who participated in the study and wanted to learn more about the data they provided, in the context of recommendations for their age group. Parents were receptive to this offer. Consent forms were updated so that families whose adolescent participated in the study could opt in or out of receiving a health report. Take-up was substantial among participant families, with more than three fourths opting in. We ultimately prepared health reports for over 250 adolescents and provided them via email (or hard copy upon request) within one year of data collection.
Importantly, in several cases more rapid reporting was required and conducted, when student blood pressure was found to fall in a dangerous range.

**Blood pressure as a measure of stress biology**

Having described the process by which the study’s outcome measures were selected, we now briefly address the relevance of blood pressure as a measure of biological stress system functioning. Acute blood pressure can be considered an index of physiological stress, threat, or effort (Blascovich, Spencer, Quinn, & Steele, 2001). In particular, increased systolic blood pressure is understood to be an adaptive response to acute stress by pushing more oxygenated blood to the cardiac system (Brownley, Hurwitz, & Schneiderman, 2000), though prolonged exposure to these elevated pressures can take a toll on the heart (Ayada, Toru, & Korkut, 2015; Vrijkotte et al., 2000). There is also evidence that blood pressure can be lowered through intervention, including both policy changes (Adam, Collier Villaume, & Hittner, 2020) and programmatic interventions (Daniels & Couch, 2013; Rainforth et al., 2007).

In addition to being a measure of acute physiological stress, blood pressure is also directly related to health. Clinically elevated blood pressure, or hypertension, affects over 40% of American adults, including one in five adults under the age of 40 (Dorans, Mills, Liu, & He, 2018). It is not evenly distributed throughout the population: hypertension is more common among Black than white Americans (Dorans et al., 2018; Mujahid, Diez Roux, Cooper, Shea, & Williams, 2011); more than twice as common among young adults who are obese, compared to those who are not obese (Grebla, Rodriguez, Borrell, & Pickering, 2010); and more prevalent among those with lower levels of education and income (Dorans et al., 2018; Lippert et al., 2017; Wickrama et al., 2012). Moreover, chronically elevated blood pressure can take a toll on bodily systems even at levels that would not meet the threshold for diagnosing hypertension (Lewington
et al., 2002; Sprint Research Group et al., 2015). The physical consequences of high blood pressure can be detected in adolescence or early adulthood, including damage to organs throughout the body (Haley et al., 2022; McCarron, Smith, & Okasha, 2002; Raitakari et al., 2009; Sorof, Poffenbarger, Franco, Bernard, & Portman, 2002a; Vos et al., 2003). The consequences of elevated blood pressure can accumulate among teens and young adults, with implications for later health and disease (Elliott & Black, 2007; Falkner, Gidding, Portman, & Rosner, 2008; Messerli, Williams, & Ritz, 2007).

The current study

The stress sub-study that our team conducted was designed to build on existing research by collecting blood pressure data from students whose high school participated in a larger RCT evaluation of TM. Designed as a pragmatic clinical trial, it was intended to assess the effects of a meditation intervention in the “normal practice” of a high school setting, with outcomes selected for their direct relevance to adolescent stress and health (Zwarenstein et al., 2008).

We hypothesized that in-school meditation could help all students reduce perceived stress levels, reduce their biological response to acute and chronic stress, and lead to improvements in students’ behavioral and cognitive outcomes. We expected students whose classes were randomly assigned to the meditation condition would have lower blood pressure than those assigned to the control condition. Given the real-world setting in which this study took place, its results are poised to yield actionable information for practitioners and policymakers whose decisions affect the types of interventions offered to adolescents.

Materials and Methods

Experimental Design: the larger RCT
The present study was designed and conducted as a smaller sub-study added to a multi-year, multi-site randomized controlled trial (RCT). As these results are the first to be published from the overall RCT, we briefly describe the larger “parent” study before providing details of this stress sub-study. Designed to test the impact of meditation on students’ behavior, academic achievement, and wellbeing, the parent study was conducted in several low-income, urban high schools within one large school district. The larger RCT was registered with the American Economic Association’s RCT registry (Guryan & Ouss, 2019). The study received approval from the school district department that reviews research projects; all procedures were also approved by the institutional review boards from the investigators’ institutions. Importantly, the design of the sub-study was presented to several meetings of parents and community members, with the protocol updated in response to community feedback.

The decision to participate in the RCT was made at the school level, with each participating school adjusting its bell schedule for the academic year to include two fifteen-minute periods of quiet during each school day. This adjusted bell schedule was in place for the duration of the school year. Random assignment was conducted at the classroom level, based on each student’s assigned second-period class (as this time of day coincided with the first of two fifteen-minute periods of quiet during the day). As such, all students enrolled in the school at the beginning of the school year were randomly assigned to one of the two conditions. Block randomization was employed to ensure that the treatment and control conditions were balanced according to student age, with one block each for classes with a majority of ninth, tenth, eleventh, and twelfth grade students.

Following random assignment, students whose second period classes were assigned to the treatment condition were offered training in transcendental meditation (TM) and encouraged to
meditate during the 15-minute quiet periods each school day. Training began to be offered in October of 2017. It was provided to students in small groups, instructed by program staff who were available to students both during quiet periods and throughout the school day. During the quiet periods, students were asked to remain seated quietly in their classroom (regardless of whether they were engaged in meditation). The control classes were assigned to spend this time as a study hall: they were permitted to read, sleep, do homework, or listen to music.

**Experimental Design: stress sub-study**

The stress sub-study took place in one high school that participated in the larger RCT. Data collection occurred between April and June of 2018, such that students had the opportunity to engage with the treatment for six to eight months prior to data collection. The data collection procedure and measures utilized for the stress sub-study are described below.

**Participants**

The larger RCT was designed to evaluate a meditation intervention for students living in disadvantaged neighborhoods that are disproportionately exposed to poverty, violence, and other structural sources of stress (Guryan & Ouss, 2019). The research team worked with the school district to identify and recruit high schools with student populations coming from low-income and high-violence neighborhoods. The school in which the stress sub-study took place has a student body that is more than 90% low-income. The school’s population primarily identifies as Black and Hispanic (with less than 1% identifying as white, Asian, or another ethnoracial group).

**Procedure**

*Recruitment.* Students were recruited for the stress measurement sub-study primarily through brief in-class presentations about one week prior to their participation. Informal
recruitment also took place in the lunchroom, in hallways, and through students’ social networks. Participation was open to all students enrolled in the school. The informed consent process included providing an introduction to the study’s purpose, description of the procedures involved in each type of data collection, and distribution of consent forms for students to obtain parental permission or, if at least 18 years old, to give consent for participation. Participation was compensated with a $20 debit card. More than 30% of the overall student body consented to participate in the stress sub-study and provided blood pressure data.

**Protocol.** Over the course of one school week, each participating student was asked to (1) fill out a daily diary in the morning and evening to provide information on their mood, stress, and activities, including whether they meditated that day; (2) complete a survey (either on paper or online) that asked about their health, stress, and recent experiences; and (3) have their blood pressure measured by a trained member of the study team during the school day.²

**Sample size and power**

During the study design stage, we conducted a formal power analysis using the PowerUp! software, which allowed us to account for the nesting of students within classrooms (Dong & Maynard, 2013). With a power of 0.80 and two-tailed alpha of 0.05, we calculated that sampling 40 classrooms with an average take-up of 10 students per class (N = 400 students) would yield a minimum detectable effect size (MDES) of 0.204. This target sample size was not fully realized due to real-world constraints on data collection, including student interest and willingness to participate (we collected data from more than one third of students enrolled in the school) and a limited timeframe within which to collect data (the final eight weeks before summer break). Our actual MDES is therefore larger than the 0.204 that we had anticipated (approximately 0.27).

---

² Participants were also asked to wear an actigraph device for objective measurement of their sleep and activity levels, though these data are not the subject of this paper.
Measures

Blood pressure data. The study’s protocol for measuring blood pressure in a school setting was designed to mirror recommendations for clinical measurement as much as possible (Pickering et al., 2005). Research assistants were trained to use a consistent protocol that involved asking students to remove clothing that covered or constricted the upper arm; sit comfortably with legs uncrossed and feet flat on the floor; and to relax quietly both before and during the measurement (Pickering et al., 2005). Oscillometric (automated) measurements were taken using the Omron 705 CP, which has been validated for use in research with adolescents (Furusawa, Ruiz, Saito, & Koch, 2005; Vera-Cala, Orostegui, Valencia-Angel, López, & Bautista, 2011). BP measurement took place in a quiet, dimmed classroom, where portable barriers were set up to create several “stations” that allowed for the simultaneous participation of multiple students, each in a relatively private space with minimal distraction. When students came into the room, they were asked to take a seat quietly at one of the measurement stations. A trained member of the research team explained the protocol for BP measurement and ascertained the student’s comfort before proceeding. They measured the circumference of the student’s upper arm to determine the appropriate cuff size and, after attaching the cuff, asked the student to sit quietly for two minutes.3 Three blood pressure readings were taken, with one minute of rest between each. Given evidence that an initial measurement can be elevated in relation to subsequent readings and normal daytime levels, the first reading was discarded and the second and third averaged4 to obtain an average value for each participant’s Systolic Blood Pressure

---

3 This allowed for a total of five minutes of seated quiet in the measurement room prior to the first reading that was utilized in analyses.
4 This procedure was in accordance with recommendations for the oscillometric measurement of blood pressure in children and adolescents.
(SBP), Diastolic Blood Pressure (DBP) (Flynn et al., 2017; Negroni-Balasquide, Bell, Samuel, & Samuels, 2016). At least 2 readings were available for all but one student.

For children and adolescents, blood pressure is interpreted in relation to the population distribution for a given sex, age, and height, with levels below the 90th percentile considered “normal”; those at or above the 90th percentile classified as “elevated”; and those at or above the 95th “hypertensive” (Flynn et al., 2017).

**Survey data.** Students were able to choose whether to complete surveys on paper or online through a link sent by email or text message. Measures included in the present study are described below.

*BMI.* Students’ self-reported height and weight were used to calculate BMI. Participant BMI was categorized as BMI as normal, overweight (≥85th percentile for sex and age), or obese (≥ 95th percentile for sex and age).

*Pubertal Development Scale.* Puberty stage was assessed using the Pubertal Development Scale, which has been validated for use among adolescents (Petersen, Crockett, Richards, & Boxer, 1988). Respondents are asked about aspects of physical development associated with puberty, with each item scored on a four-point scale (1 = “not yet begun” and 4 = “seems completed”). Some items (e.g., “growth in height,” “body hair”) pertain to both males and females, whereas others are specific to the respondent’s sex (e.g., deepening of voice for males, menarche for females). Responses were averaged, with the resulting pubertal stage score included as a covariate.

*Participation in meditation.* The questions, “Have you been trained in meditation?” and “Do you meditate?” were used to probe the extent of participation in the intervention. Those who indicated that they do meditate were also asked, “How often do you meditate?,” which
corresponded to a four-item scale for meditation frequency (1 = less than once per week; 2 = 1-2 times per week; 3 = 3-4 times per week; 4 = 5 or more times per week). These items were used to create three indicator variables that correspond to whether a participant reported that they received training in meditation; reported that they meditate at all; and reported that they meditate three or more times per week.

**Results**

**Descriptive statistics: the larger RCT**

Randomization was successful, with classroom-level random assignment carried out as intended. In the school where the stress sub-study took place, the treatment and control conditions were balanced at the individual level in terms of student age and sex for the \( N = 779 \) students who were enrolled at the time of random assignment (see Table S3).

**Descriptive statistics: stress sub-study**

Outcome data is available for a sample of \( N = 254 \) participants (50.78% female; \( M_{\text{age}} = 16.90; SD_{\text{age}} = 1.25; \) 61.81% Black; 38.19% Hispanic; \( M_{\text{BMI}} = 25.27; SD_{\text{BMI}} = 6.34; \) full descriptive statistics appear in Table S1 of the Supplemental Digital Content file). Twenty-two percent of the sample was classified as obese, with BMI at or above the 95th percentile for their age and sex (Adolescents, 2004).

**Blood Pressure**

In our sample, average levels for SBP were 115.28 mm Hg (\( SD = 11.18 \)) for males and 107.61 mm Hg (\( SD = 9.38 \)) for females (\( b_{\text{diff}} = 7.67, p < 0.001 \)). Average levels for DBP were 64.88 mm Hg (\( SD = 8.05 \)) for males and 66.17 mm Hg (\( SD = 7.75 \)) for females (\( b_{\text{diff}} = 1.29, p = 0.20 \)). One third of males and over 10% of females had blood pressure at or above the 90th percentile for their age, sex, and height (Table S1).
Balance tests

Participation in the stress sub-study was entirely voluntary; to ensure that compensation for study participation was offered to all students, participation was open to all students in the school. This resulted in response rates that differed between treatment and control conditions (as the sample was not stratified or capped in any way); 44% of our sample came from the control condition and 56% from the treatment condition. Those who provided blood pressure from the treatment condition were older ($b_{diff} = 0.53, p < 0.01$) and more often female ($b_{diff} = 0.13, p < 0.05$) than those who provided blood pressure from the control condition; in turn, they reported more advanced pubertal development than their peers in the control condition ($b_{diff} = 0.16, p = 0.05$) (Marshall & Tanner, 1970). The treatment and control groups did not differ significantly on BMI or race/ethnicity (Table S1). Note that since students had been randomly assigned to the two conditions, there are no differences in observable characteristics between the two conditions in the larger RCT; the differences arose from voluntary participation in the stress sub-study. We conduct several tests, detailed below, to assess how the observed selection bias could affect our ability to estimate the effect of the treatment.

Intent to Treat

Initial analyses estimated whether the intervention was associated with average differences in blood pressure between those assigned to the treatment and control conditions, regardless of actual meditation practice. To account for classroom-based random assignment and the possibility of a shared experience within a given treatment or control classroom, we used mixed models that nest students within the classroom that was the basis for random assignment, with a random effect for classroom and standard errors clustered at the classroom level.

---

5 This is consistent with evidence that, on average, adolescent females enter pubertal stages earlier than males (Petersen et al., 1988).
Covariates included sex, age, BMI, race/ethnicity, and pubertal stage. Results of Intent to Treat (ITT) models are presented in Table 1, with separate panels for models estimating impacts on SBP and DBP. Each column represents a separate regression, with impacts estimated for the full sample and then for several subgroups of interest (by sex, race/ethnicity, and obesity status).

For the full sample, we did not detect a significant difference between the treatment and the control condition, though significant impacts on SBP were observed for some subgroups. The ITT estimate for SBP was -2.37 mm Hg ($SE = 1.30, p = .07$), an effect of 0.21 SD units (Table 1). The ITT estimate for DBP was -0.14 SD ($p = .35$). Estimated impacts on SBP were comparable for male and female participants (-2.75 mm Hg and -2.36 mm Hg, respectively). However, ITT estimates diverged in models estimated separately by race and ethnicity: the treatment was associated with SBP 4.06 mm Hg ($SE = 1.66, p = .02; 0.36 SD$ units) lower for Black students, with no significant difference detected for Hispanic students. Neither group saw a significant treatment impact on DBP. For students with obese BMI, the treatment was associated with SBP that was 7.03 mm Hg ($SE = 2.64, p < 0.01$) lower and DBP that was 5.23 mm Hg ($SE = 1.72, p < 0.01$) lower, with effect sizes of 0.63 and 0.64 SD units, respectively. Among students with non-obese BMI, estimated effects did not significantly differ from zero.

*Treatment on the Treated*

The above results estimated the average effect of the treatment based on a student’s treatment assignment. They did not take into account variations in whether students who were assigned to the treatment group chose to be trained in meditation or engaged with the practice during the quiet periods in the school day, or possible spillover effects that may have reached students in the control condition. We then moved to a treatment on the treated (TOT) analysis (Angrist & Pischke, 2008), with random assignment used as an instrument to predict a student’s
reported level of treatment. To measure this, we used survey questions that asked students (1) whether they had been trained in meditation; (2) if they meditated; and, if so, (3) how many times each week they meditated (shown in Figure 1 and Table S1). Two-stage least squares regressions were used to estimate the effect of participation in the meditation program, using each of the three measures of participation described above.

Full results from the TOT analyses are presented in Table 2, with each cell representing the estimated treatment effect for a given frequency of meditation, categorized as (1) having received training in meditation; (2) reporting that they meditate at all; and (3) reporting that they meditate at least three times per week. Importantly, these estimates do not test whether a given frequency of meditation is more effective than another; they estimate the effect of the treatment on the treated at each level of participation (assuming that any amount below this threshold is akin to not receiving the treatment).

As demonstrated in Figure 2, TOT estimates are consistent with significant impacts on several subgroups of interest, with SBP impacts for Black students of -4.28 mm Hg ($SE = 1.77$, $p = .02$) among those who were trained in meditation; -5.75 mm Hg ($SE = 2.19$, $p = .01$) for those who reported meditating at all; and -6.87 mm Hg ($SE = 2.52$, $p = .01$) for those who meditated at least three times per week (effects of 0.38 to 0.61 SD). Among obese adolescents, these estimates were -7.80, -12.29, and -20.24 mm Hg, respectively (.69 to 1.79 SD, all $p <.05$).

Robustness tests

Finally, we considered whether observed effects are robust to selection bias, given that the students who provided blood pressure data reflect a convenience sample drawn from students who participated in the larger RCT. As mentioned above, imbalance across treatment and control groups in the sample that provided blood pressure data was observed on sex and age. First, we
confirmed that these were not correlated with the odds of random assignment to the treatment condition, given that random assignment occurred for the larger RCT and was therefore not subject to voluntary participation (Table S3).

We then used administrative school data to see whether accounting for other observable factors or a student’s school performance or behavior during the prior academic year impacted the estimated treatment effect. The addition of these covariates changed ITT estimates by less than 5%, as shown in Table S4. The fact that controlling for a rich set of observable socioeconomic and behavioral measures did not affect the estimated treatment effect is reassuring regarding concerns of selection into study participation. However, students could still have selected into the study based on unobservable characteristics. While we cannot directly test for selection on factors that are unobserved in our data, we follow the approach developed by Oster to estimate how much selection on unobservable factors there would have to be, relative to the selection we see on observable factors, for our estimated treatment effects to be zero (Oster, 2019). This approach relates changes in coefficients to changes in the R-squared; in our context, we find that selection on unobservable characteristics would have to be 7 times higher than selection on observables to produce a null effect of meditation (Oster, 2019).

Lastly, classroom-based random assignment was stratified according to the most common grade present in each class, resulting in four randomization blocks that differed significantly in average age (15.49 to 18.20 years, all \(p < 0.001\)). Given the high correlation between age and random assignment block \((r = 0.84, p < 0.001)\), we considered alternate model specifications that controlled for randomization block instead of age or alongside a series of indicator variables for age in years. These models estimated an average impact on SBP of 2.15 mm Hg \((SE = 1.42, p = .13)\) and 2.27 mm Hg \((SE = 1.39, p = .10)\), respectively (Table S5). These alternate specifications
changed our ITT estimate by less than 10%, consistent with an estimated treatment effect that is stable, though not entirely robust, to the addition of this control. Overall, we find consistent evidence of the primary treatment effect observed, an impact on SBP, for certain subgroups; and, for these groups, we find that it is robust to multiple model specifications.

**Discussion**

In this paper, we presented empirical results from the first in-school RCT of meditation that randomly assigned all enrolled students to either a meditation or control condition. Biological stress outcomes were measured at the end of the school year in which the intervention took place. While impacts for the full sample were modest in size and not statistically significant, we observed an impact on SBP for certain subgroups that is robust to a variety of controls and model specifications. Specifically, this impact was concentrated among subgroups of adolescents at elevated risk for early onset of hypertension, including Black students and students with an obese BMI, which represented 22% of our sample. These findings are consistent with previous studies (Anderson, Liu, & Kryscio, 2008; Barnes et al., 2008; Barnes et al., 2004).

We consider two potential explanations for the patterns observed. First, it is possible that subgroups at risk of having higher blood pressure\(^6\) stood to experience larger benefits from an intervention that lowers blood pressure. Indeed, our results are consistent with population data that find adolescent males are at twice the risk of developing systolic hypertension as their female peers (Grebla et al., 2010); that Black Americans are the racial/ethnic group in the United States at highest risk for hypertension (Din-Dzietham, Liu, Bielo, & Shamsa, 2007; Muntner, He, Cutler, Wildman, & Whelton, 2004); and that obese adolescents are three times as likely as their

---

\(^6\) Recall that we do not have baseline blood pressure data for our sample; we attempt, therefore, to draw inferences about baseline risk through a combination of population data on blood pressure among adolescents in these subgroups and by considering the blood pressure of control group students in the same subgroup.
nonobese peers to have BP that exceeds hypertensive levels (Flynn & Alderman, 2005; Sorof & Daniels, 2002; Sorof, Paffenbarger, Franco, Bernard, & Portman, 2002b). In addition, existing literature reports that risk compounds for youth who belong to multiple risk categories (Flynn & Alderman, 2005; Grebla et al., 2010). Put another way, there could be what we describe as a “floor effect” on blood pressure for lower-risk subgroups—that is, there could be less room for blood pressure to decrease among students who have, on average, blood pressure that falls lower in the population distribution (Loucks et al., 2015).

It is more difficult to explain the pattern observed between Black and Hispanic students. In our sample, Black and Hispanic adolescents in the control condition have similar blood pressures, but only Black students’ blood pressure is decreased with the intervention. The similarity in blood pressures in the control condition suggests that Black students in the treatment condition did not have higher baseline blood pressure than their Hispanic peers. Additionally, both Black and Hispanic students in the control condition have high BP levels; it is therefore unlikely that there would be a “floor effect” for Hispanic students relative to Black students. We have considered several other ways that the intervention could have been more effective for some groups of students than others. These include (1) differences in take up or dosage of the intervention and (2) a larger effect of meditation realized at a given dosage.

Considering the first explanation, if the intervention were more acceptable or appealing to some groups of students, we might expect to see differences in students’ reports of having been trained in meditation or of the frequency with which they meditate; however, no differences emerge by race or ethnicity on these measures of participation (see Table S1), though we note that our data would be unable to detect differences between groups in the quality of participation. In contrast, our TOT estimates found that Black students who were trained in meditation saw an
impact on SBP of -4.28 mm Hg and those who reported meditating at least three times per week saw an impact of -6.87 mm Hg, effects that range from .38 to .61 $SD$ units. For Hispanic students, estimates were approximately zero across all reported frequencies of meditation.

While we can empirically rule out some potential explanations with our current research design and data, it is possible that engagement was different among Black than Hispanic students for reasons that are unobservable to researchers. As an example, there could be different peer social dynamics or different relationships with meditation coaches, both interpersonal mechanisms that might be context dependent. Future studies could work to understand whether and how interpersonal relationships or cultural factors might affect engagement with a meditation intervention.

Thinking about scaling up this program in the light of our results warrants some attention to the time intensity of the overall intervention. It involved thirty minutes of dedicated time within each school day, for a total of 90 hours across a standard 180-day school year. For those assigned to the treatment group, having access to both meditation training and dedicated time to meditate during the school day may have eliminated an important barrier to habit formation, increasing the likelihood of establishing and maintaining a health-promoting practice (Irving et al., 2012; Lynch, Gander, Kohls, Kudielka, & Walach, 2011). Given that many adolescents report the demands of the school day as a major source of stress in their lives, this change to the school schedule may have brought students a measure of calm and reprieve regardless of their random assignment condition. That is, it is possible that the adjusted bell schedule, which brought periods of reduced noise and activity to each school day, represented a departure from the typical school day for all students, including those assigned to the control condition.
It is also possible that spillover effects arose, especially as afternoon quiet periods included a mix of students in the treatment and control conditions in each class. If no group truly experienced a school-day-as-usual condition, it is possible average differences between the two conditions may understate the true effect of the treatment. However, the control “activity” to which students were assigned (sitting quietly, as though in study hall) is not something that is associated with improved blood pressure or cardiometabolic health; therefore, these findings should be interpreted in comparison to an inactive state rather than as a trial of meditation in comparison to another lifestyle intervention.

As the study’s impacts were concentrated among several subgroups of students, it is not clear whether other groups simply experienced smaller effects or whether there are adolescents for whom meditation has no impact on blood pressure. We have raised the possibility of a threshold (or floor) below which blood pressure is unlikely to decrease. It is also important to note that our study is underpowered to detect a small effect of meditation, in this case a decrease below 2 mm Hg. Nonetheless, an impact of this magnitude, observed within a single school year, may be nontrivial for adolescents as it could prevent progression beyond recommended levels. Given that effects may vary based on sex, race/ethnicity, and BMI, future research would benefit from a larger sample size that accounts for the potential for heterogeneous effects.

This study has several key strengths in relation to the existing body of literature. It is the first to our knowledge to offer experimental evidence that a school-wide meditation program can lead to clinically meaningful impacts on blood pressure, in particular for adolescents at elevated risk for hypertension. We extend previous literature that has primarily studied meditation in smaller samples or older populations. The American Heart Association identifies a “lack of randomization and/or appropriate control group” and small number of research groups
responsible for many studies of meditation among the “limitations [that] prevent definitive conclusions regarding efficacy of meditation on cardiovascular risk reduction” (Levine et al., 2017). As all students enrolled in the school went through random assignment, we avoid the type of selection bias that plagues studies in which participants seek out meditation due to an expectation of benefit. This study design also allows us to distinguish between the average effects of the treatment and its impacts on those who engaged with the meditation intervention. Finally, we study blood pressure in a school context where most students are low-income and African American or Hispanic, groups that can face heightened risk for early onset of hypertension.

In light of these strengths, we move next to providing some context for the magnitude of impacts observed. The average treatment effect on SBP is modest and non-significant, an impact of -2.37 mm Hg or .21 SD. This amount is comparable to the average annual increase in SBP observed in mid-to-late adolescence (Muntner et al., 2004). If a significant effect of this magnitude were replicated in future research, achieving this impact within one school year would be sufficient to mitigate further increases among adolescents who already have elevated blood pressure. Additionally, it is estimated that non-Hispanic Black males have SBP that is 2.9 mm Hg higher than that of non-Hispanic whites (Muntner et al., 2004). The ITT estimate for Black adolescents, an impact of 4.06 mm Hg, would be sufficient to eliminate this racial disparity. Lastly, obese students saw more than 0.6 SD impacts on both SBP and DBP (-7.03 and -5.23 mm Hg, respectively). This group was overrepresented in our sample, with 22% of students having BMI that exceeds the 95th percentile for their age and sex. The potential to reduce BP to healthy or high-normal levels for this group of adolescents has clinical implications, as it may slow the progression towards hypertension or the need for treatment with medication. In fact, the impacts
observed for obese adolescents are among the largest impacts of meditation that have been published, including among older adults or those selected on the basis of CVD risk (Schneider et al., 2012). Our observed impacts on the SBP of obese adolescents are comparable to the impacts that a fifteen-pound (7 kilogram) weight loss (Daniels & Couch, 2013; Neter, Stam, Kok, Grobbee, & Gelseijnse, 2003) or the initiation of a medication such as an ACE inhibitor or β-blocker (Morgan, Anderson, & MacInnis, 2001) could be expected to have on blood pressure for a hypertensive individual. A recent meta-analysis estimates that each 10 mm Hg reduction in SBP is associated with significant reduction of risk of major cardiovascular events and all-cause mortality (Ettehad et al., 2016). If school-based meditation can achieve blood pressure reductions approaching this magnitude among obese adolescents, even with more modest impacts for their lower-risk peers, this intervention is poised to meaningfully reduce cardiovascular risk.

**Taking a multidisciplinary perspective**

If this were a typical empirical paper, the paragraph above would roughly mark the transition from Discussion into Conclusion. But accomplishing our dual purpose in this article calls us to interpret this paper’s findings through the lens of our multidisciplinary collaboration. We do so in the pages that remain.

We summarize the findings presented above by saying that the intervention appeared to work for some—but not for everyone. Further, the research design that was realized (as opposed to what might have been intended) makes this set of findings a bit less conclusive than we may have intended or expected. The first observation has been unpacked in detail earlier in the Discussion section. Our only additional note on it is that heterogeneous effects may be especially important when complex studies are carried out in real-world contexts (specifically, a behavioral intervention, delivered in a school setting, to a sample of adolescents who may not mirror the
The subgroup impacts we identified, both for Black students and for those with obese BMI, would be quite promising if they are shown to replicate in another sample. This evidence can contribute to the evidence base by spurring future work, both in more traditional disciplinary (i.e., psychology) research and in other applied settings like this one.

We turn now to the second observation: there are limitations to what we can conclude from this type of research. In an ideal world, participants in a sub-study are quite similar to—statistically indistinguishable from—the participants in the larger RCT. Additionally, both in an ideal world and in our initial study design, participants in the sub-study provide baseline data that allows the researchers to confirm that there was no difference between those in the treatment and control conditions when the intervention began. In our case, neither of these “ideals” materialized. Participants in the sub-study did not perfectly resemble the full student body that had been randomly assigned. And we were not able to obtain baseline stress biology data. Simply put, there’s messiness in the data collected.

We rely on the tools of economics to help us make sense of this messiness. Its tools are well suited to this task, given the emphasis in econometrics on identifying causal effects using quasi-experiments instead of randomized experiments. For instance, there are tools to quantifying how differences in observable characteristics, like the imbalances we observed, may impact estimates of program outcomes (e.g., Oster, 2019). But methodology papers may not be in wide circulation outside of their respective disciplines. If an approach is not broadly understood by a multidisciplinary audience, readers or reviewers from other disciplines may be less able to interpret, or less willing to accept, results that appear to be based on imperfect random assignment. One insight we offer for future multidisciplinary work is that the analytic
tools of one discipline may not translate effectively to others. In our case, subsequent analyses of this data may incorporate techniques more common in psychology as a complement to the econometric tools employed here. For others seeking to conduct multidisciplinary work, this consideration could inform analysis plans that are developed prior to collecting data.

Community engagement and power sharing. We return to the experience of community engagement with several reflections. It was important that we came to recognize the importance of the different perspectives university researchers and the community in which the study was taking place brought to the study. An awareness of our positionality allowed us to change course upon learning of parent concerns and to establish a plan for including the school community in the decision-making process. Looking back, we feel confident that this decision was the right choice. We also acknowledge that holding space for the community requires allocating personnel resources and time for these important conversations. In our case, one result of this time commitment, and a change from our originally conceived outcome variable, was that data collection was limited to the final eight weeks of the school year. This meant we could not collect baseline sleep or blood pressure data. It may also have limited the study’s sample size, as neither recruitment nor data collection could continue into summer break.

Other literature has described relevant constraints and tradeoffs in the context of co-designed research. These include time constraints (Goodyear-Smith et al., 2015) and elements of research design that may lead studies not to meet the “gold standard” for academic research if decisions are made to prioritize community needs or preferences (Minkler & Baden, 2008). In other words, power sharing may come at a cost when it comes time to disseminate the research findings. We believe this is a cost worth paying. At the same time, we see a need for more

7 This point is also relevant to other aspects of research design or study operations
conversation in adolescent research about these tradeoffs, including their implications for spurring future research and for impacts on policy and practice.

Publication and dissemination. After a multidisciplinary team successfully collaborates on research design, data collection, and data analysis, one might expect publishing the study’s findings to come naturally at the end of this process. But, even here, communication across disciplinary lines can prove challenging. Our primary insight for future research is that disseminating the findings of a multidisciplinary study is itself a multidisciplinary effort. There may be uncertainty around the audience(s) to which the presentation of results should be tailored, whether certain journals are comparably beneficial for each author’s professional development, or whether disciplinary journals’ editors or reviewers will be able to fairly review a multidisciplinary paper.

Our team chose to prioritize the professional needs and interests of the study’s more junior investigators, including the first author, with the understanding that each additional publication has a greater impact on the profile of an early-career scholar than for a study PI. (In economics parlance, more established researchers receive less marginal benefit from each additional publication.) However, this approach did not make it straightforward to “sell” a multidisciplinary paper to a journal. While some journals, including the *Journal of Research on Adolescence*, take pride in publishing research from a variety of disciplines, it remains the case that most editors and reviewers have a disciplinary rather than multidisciplinary orientation. It seems that responsibility falls to the research team to carefully call attention to and explain the details of a multidisciplinary study (e.g., in the cover letter that accompanies a journal submission) or to ask the editor to include a reviewer from the discipline that would typically be underrepresented in that journal’s review process.
There are things that can be done to mitigate these challenges. We see value in journals for multidisciplinary studies or interdisciplinary social sciences. But our enthusiasm is tempered by an awareness that such outlets are few and, even among them, it remains challenging to identify reviewers with the correct array of skills to evaluate multidisciplinary studies. Perhaps reviewer profiles could better account for multidisciplinary training or areas of methodological expertise that complement a prospective reviewer’s disciplinary orientation. To the extent that conducting multidisciplinary research might slow the time to publication (i.e., if the existing review processes is not well-suited to evaluating multidisciplinary products), some open science practices may be conducive to the dissemination of findings. For instance, preregistration and results-free peer review are recommended practices (Axford, Berry, Lloyd, Hobbs, & Wyatt, 2022), as is the posting of manuscripts to pre-print servers. Posting preprints online or presenting early versions of multidisciplinary papers at conferences may also help the authors of multidisciplinary papers to solicit feedback from a range of disciplinary perspectives that, in turn, inform their publication-related decisions. Finally, the Abdul Latif Jameel Poverty Action Lab (J-PAL) has suggested that registry metadata be used to identify (or quantify) studies that have been initiated but not yet published (source), whether due to publication bias or other hurdles. We note that this recommendation could be especially informative in the context of multidisciplinary research, both for individual researchers embarking on multidisciplinary work and for the field to build an understanding of whether multidisciplinary work faces notable obstacles in getting from registration to publication.

**Concluding recommendations for multidisciplinary adolescent research**

In a 2017 paper, Sigfusdottir and her colleagues called for an interdisciplinary approach to the study of adolescent stress, arguing it would be necessary to elucidate links from stress
physiology to behavior (Sigfusdottir, Kristjansson, Thorlindsson, & Allegrante, 2017). We agree that bringing multiple disciplinary perspectives is an important step forward for adolescent stress research and see our study as one example of a response to this call. Through the lens of our own multidisciplinary collaboration, we have offered several insights and recommendations for future multidisciplinary teams. They include that, when multidisciplinary research takes on complex, real-world topics, its findings stand to inform bench science as well as policy and practice. But there can be messiness that arises from taking on research that transcends disciplinary boundaries—and making sense of this messiness will require the research team to assemble a strong methodological toolkit. The dissemination and communication of findings is itself a multidisciplinary endeavor in this type of work—and it may be one that operates on a slower timeline than traditional disciplinary research. With this paper, we have also called attention to the need for more communication about multidisciplinary research, including as part of the journal review process. One of our contributions to this objective includes the development of a vocabulary list to help bridge some of the common language that may be needed for collaboration among developmental psychologists, economists, and others interested in multidisciplinary research about adolescent stress. We hope this conversation will continue among the broader community that is interested in what multidisciplinary intervention research can tell us about adolescent stress, from elucidating its causes to identifying promising solutions.
References


individual data for one million adults in 61 prospective studies. *Lancet, 360*(9349), 1903-1913. doi:10.1016/S0140-6736(02)11911-8


Minkler, M., & Baden, A. C. (2008). Impacts of CBPR on academic researchers, research quality and methodology, and power relations. In M. Minkler & N. Wallerstein (Eds.), *Community-based participatory research for health: From process to outcomes* (pp. 243-262): John Wiley & Sons.


Figures and Tables

Figure 1. Reported frequency of meditation by random assignment condition. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Full results are detailed in Table S1.
Figure 2. TOT Impacts on SBP, by subgroup. + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. Error bars reflect standard errors for each estimate. Full results are detailed in Table 2.
### Table 1. Intent to Treat (ITT) Estimates

<table>
<thead>
<tr>
<th></th>
<th>Full sample</th>
<th>Females</th>
<th>Males</th>
<th>Hispanic</th>
<th>Black</th>
<th>Obese</th>
<th>Not obese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 129</td>
<td>N = 125</td>
<td>N = 97</td>
<td>N = 157</td>
<td>N = 57</td>
<td>N = 197</td>
<td></td>
</tr>
<tr>
<td><strong>Systolic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment effect</td>
<td>-2.365</td>
<td>-2.357</td>
<td>-2.746</td>
<td>0.025</td>
<td>-0.056</td>
<td>-7.029</td>
<td>-1.392</td>
</tr>
<tr>
<td></td>
<td>(1.299)</td>
<td>(1.461)</td>
<td>(1.540)</td>
<td>(1.962)</td>
<td>(1.663)</td>
<td>(2.635)</td>
<td>(1.626)</td>
</tr>
<tr>
<td>Female</td>
<td>-8.675</td>
<td>1.439</td>
<td>0.370</td>
<td>-1.026</td>
<td>2.399</td>
<td>-14.797</td>
<td>-7.064</td>
</tr>
<tr>
<td></td>
<td>(1.140)</td>
<td>(0.690)</td>
<td>(0.849)</td>
<td>(1.210)</td>
<td>(0.685)</td>
<td>(3.379)</td>
<td>(1.003)</td>
</tr>
<tr>
<td>Age</td>
<td>1.144</td>
<td>1.979</td>
<td>5.405</td>
<td>4.000</td>
<td>3.682</td>
<td>2.247</td>
<td>0.858</td>
</tr>
<tr>
<td></td>
<td>(0.714)</td>
<td>(0.666)</td>
<td>(1.113)</td>
<td>(0.965)</td>
<td>(0.0865)</td>
<td>(1.050)</td>
<td>(0.819)</td>
</tr>
<tr>
<td>BMI</td>
<td>3.639</td>
<td>-2.561</td>
<td>0.406</td>
<td>0.808</td>
<td>1.786</td>
<td>-4.328</td>
<td>-0.749</td>
</tr>
<tr>
<td></td>
<td>(0.705)</td>
<td>(1.565)</td>
<td>(2.256)</td>
<td>(1.070)</td>
<td>(0.877)</td>
<td>(2.195)</td>
<td>(1.550)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.601</td>
<td>2.026</td>
<td>0.962</td>
<td>0.808</td>
<td>1.786</td>
<td>0.197</td>
<td>1.795</td>
</tr>
<tr>
<td></td>
<td>(0.639)</td>
<td>(0.727)</td>
<td>(0.995)</td>
<td>(0.965)</td>
<td>(0.0865)</td>
<td>(0.997)</td>
<td>(1.373)</td>
</tr>
<tr>
<td></td>
<td>(0.720)</td>
<td>(1.154)</td>
<td>(1.767)</td>
<td>(1.944)</td>
<td>(1.876)</td>
<td>(2.550)</td>
<td>(1.777)</td>
</tr>
<tr>
<td><strong>Diastolic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment effect</td>
<td>-1.174</td>
<td>-2.507</td>
<td>0.176</td>
<td>-0.353</td>
<td>-1.920</td>
<td>-5.232</td>
<td>-0.767</td>
</tr>
<tr>
<td></td>
<td>(0.874)</td>
<td>(1.627)</td>
<td>(0.169)</td>
<td>(1.371)</td>
<td>(1.542)</td>
<td>(1.716)</td>
<td>(1.495)</td>
</tr>
<tr>
<td>Female</td>
<td>0.704</td>
<td>1.123</td>
<td>0.818</td>
<td>0.788</td>
<td>1.420</td>
<td>3.264</td>
<td>1.795</td>
</tr>
<tr>
<td></td>
<td>(0.420)</td>
<td>(0.575)</td>
<td>(0.687)</td>
<td>(2.016)</td>
<td>(1.233)</td>
<td>(2.854)</td>
<td>(0.624)</td>
</tr>
<tr>
<td>Age</td>
<td>1.137</td>
<td>1.157</td>
<td>2.974</td>
<td>0.922</td>
<td>2.931</td>
<td>1.269</td>
<td>1.294</td>
</tr>
<tr>
<td></td>
<td>(0.874)</td>
<td>(0.524)</td>
<td>(0.697)</td>
<td>(0.626)</td>
<td>(0.471)</td>
<td>(1.400)</td>
<td>(0.654)</td>
</tr>
<tr>
<td>BMI</td>
<td>2.064</td>
<td>1.500</td>
<td>0.070</td>
<td>0.608</td>
<td>1.133</td>
<td>-6.477</td>
<td>0.506</td>
</tr>
<tr>
<td></td>
<td>(0.454)</td>
<td>(1.518)</td>
<td>(1.537)</td>
<td>(0.881)</td>
<td>(0.648)</td>
<td>(1.524)</td>
<td>(1.085)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.675</td>
<td>-1.504</td>
<td>-0.741</td>
<td>0.608</td>
<td>1.133</td>
<td>-1.626</td>
<td>1.601</td>
</tr>
<tr>
<td></td>
<td>(0.958)</td>
<td>(1.518)</td>
<td>(1.537)</td>
<td>(0.881)</td>
<td>(0.648)</td>
<td>(2.262)</td>
<td>(0.636)</td>
</tr>
<tr>
<td>Puberty stage</td>
<td>0.824</td>
<td>66.022</td>
<td>65.138</td>
<td>65.421</td>
<td>66.691</td>
<td>73.833</td>
<td>64.633</td>
</tr>
<tr>
<td></td>
<td>(0.128)</td>
<td>(1.092)</td>
<td>(1.240)</td>
<td>(1.240)</td>
<td>(1.296)</td>
<td>(2.291)</td>
<td>(1.244)</td>
</tr>
</tbody>
</table>

Note. Each column presents results from a separate multilevel regression model that nests students within classrooms and clusters standard errors at the classroom level.
## Table 2. Treatment on the Treated (TOT) Estimates

<table>
<thead>
<tr>
<th></th>
<th>A. Full sample</th>
<th>B. Subgroups by sex</th>
<th>C. Subgroups by race/ethnicity</th>
<th>D. Subgroups by obesity status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Black</td>
</tr>
<tr>
<td></td>
<td>B(SE)</td>
<td>p-value</td>
<td>B(SE)</td>
<td>p-value</td>
</tr>
<tr>
<td><strong>Systolic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trained in meditation</td>
<td>-2.509</td>
<td>0.083</td>
<td>-3.218</td>
<td>0.078</td>
</tr>
<tr>
<td>(1.448)</td>
<td></td>
<td></td>
<td>(1.823)</td>
<td></td>
</tr>
<tr>
<td>Meditate at all</td>
<td>-3.448</td>
<td>0.061</td>
<td>-4.524</td>
<td>0.038</td>
</tr>
<tr>
<td>(1.838)</td>
<td></td>
<td></td>
<td>(2.182)</td>
<td></td>
</tr>
<tr>
<td>Meditate ≥ 3X per week</td>
<td>-4.096</td>
<td>0.066</td>
<td>-4.899</td>
<td>0.042</td>
</tr>
<tr>
<td>(2.244)</td>
<td></td>
<td></td>
<td>(2.404)</td>
<td></td>
</tr>
<tr>
<td><strong>Diastolic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trained in meditation</td>
<td>-1.174</td>
<td>0.408</td>
<td>0.403</td>
<td>0.755</td>
</tr>
<tr>
<td>(1.418)</td>
<td></td>
<td></td>
<td>(1.291)</td>
<td></td>
</tr>
<tr>
<td>Meditate at all</td>
<td>-1.538</td>
<td>0.406</td>
<td>0.557</td>
<td>0.747</td>
</tr>
<tr>
<td>(1.851)</td>
<td></td>
<td></td>
<td>(1.722)</td>
<td></td>
</tr>
<tr>
<td>Meditate ≥ 3X per week</td>
<td>-1.827</td>
<td>0.418</td>
<td>0.603</td>
<td>0.745</td>
</tr>
<tr>
<td>(2.257)</td>
<td></td>
<td></td>
<td>(1.851)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Each model presents results from a separate two-stage least squares regression that uses reported frequency of meditation as an instrument for treatment status, with standard errors clustered at the classroom level. Models include controls for sex, age, BMI, race/ethnicity, and pubertal stage.
Appendix 1 Vocabulary List and Recommended Reading

From economics (listed in order of appearance in the paper):

**Balance tests** analyze the similarity across treatment conditions on measures that were present prior to the treatment or that are expected to be unchanged by the intervention. For example, a school-based intervention might test whether participants assigned to treatment and control conditions were similar in age or school grade, baseline depressive symptoms, or GPA during the academic year prior to the intervention. Statistical tests may be used to test whether the groups are significantly different from one another on any of the selected measures, with results reported to either demonstrate the success of randomization or to identify areas of imbalance that may be addressed before (or alongside) statistical analyses. Balance tests are common in economics as well as in several other disciplines that conduct RCTs, including political science and medicine. See Athey & Imbens (2017) for more on the use of balance tests in economics; a perspective that is more critical of balance tests can be found in Mutz, Pemantle, & Pham (2019).

**Intent to treat (ITT) estimates** compare outcomes between those assigned and not assigned to the treatment. This estimate is the average effect of the experiment, regardless of how many people “took up” the treatment. It is an unbiased estimate of the effect of being assigned, though it is understood that not all who are assigned to, or offered, a treatment will choose to take it. This means the ITT estimate may substantially underestimate the true effect of the treatment, as the impact is averaged across all who were assigned. ITT estimates can only be obtained if outcome data are observed for all (or very nearly all) participants who went through randomization. This is common when outcome data are obtained through administrative data sources that are not affected by any participant attrition that occurs. For more on econometric methods in RCT evaluation, including ITT estimates, see Angrist & Pischke (2008).

**Effect of the treatment on the treated (TOT)** refers to the average effect of the treatment on those who received it. This estimate is understood to be affected by selection bias, since individuals choose whether to take up an intervention after being assigned to the treatment. This estimate is always larger than the ITT estimate, as it divides the ITT estimate in light of the proportion who took up the treatment (see Angrist & Pischke, 2008 for details).

**Instrumental variables (IV)** are an approach to estimating the effect of the treatment on compliers, those who are both offered the treatment and actually receive it. The method involves predicting exposure to the treatment $X$ from another variable $Z$, through which exposure operates. In the case of an RCT, the instrument may be random assignment, but IVs can also be non-experimental influences that are as good as random but not randomly assigned. In the case of our RCT, we used random assignment to the treatment as an instrument that increased the likelihood of participants’ reporting that they meditate each week. This method assumes that the instrument affects the outcome (here, blood pressure) through just one channel, the increased likelihood of taking up the treatment. For more, see Bloom (1984) and Wing & Clark (2017).

From developmental psychology:

**Stress** has been defined as “a real or interpreted threat to the physiological or psychological integrity of an individual that results in physiological and/or behavioral responses”
Stress refers broadly to both environmental challenges that elicit a “stress response,” and to the processes within the individual that they set in motion. Most definitions of stress encompass at least two elements, (1) a demand, challenge, or perturbing force, and (2) a response, consisting of cognitive, emotional, behavioral and biological elements. As with most stress experts, we distinguish between stressors and the stress response, each defined below.

**Stressors** are experiences that bring about, or elicit, a response. Stressors can be real or perceived, current or anticipated, good or bad.

**Stress response** refer to a range of processes that can be set in motion by exposure to environmental stressors. Stress responses may be biological, cognitive, emotional, or behavioral.

For more on stress, including stressors and the stress response, we recommend: (Adam et al., 2023; Del Giudice, Ellis, & Shirtcliff, 2011; Grant et al., 2003; McEwen, 2010; Monroe, 2008)

**Biological stress systems** are the “first responders” of stress experiences that help us to cope with stressors at hand, for instance by mobilizing energy resources or helping to prevent/contain injury. We typically focus on primary stress mediators, which include the autonomic nervous system, the hypothalamic pituitary adrenal axis (HPA-axis), the immune and inflammatory systems, and the alertness/sleep system. These primary stress mediators regulate each other and have impacts on a variety of “downstream” neurological and biological systems that are important for long-term health and wellbeing. See Adam, Collier Villaume, & Hittner (2020) for an accessible overview and McEwen & Seeman (1999) for more detail on primary stress mediators.

**Stress disparities** are circumstances, including social or demographic factors, that lead a person or group to experience systematically different levels of stress exposure that, in turn, create differences in stress biology. Stress disparities can come from various types of stressors, happen at different times in the developmental lifespan, and can interact or accumulate over time. The concept of stress disparities is introduced in Adam, Collier Villaume, & Hittner (2020).

**Effectiveness studies** are RCTs that take place in “real-world” settings. In contrast to “efficacy studies,” which aim to carry out interventions under ideal circumstances, effectiveness studies tend to have more relaxed eligibility criteria that prioritize enrolling a sample that is representative of the population of interest. See work from Bauer & Kirschner (2020) and Loucks and his colleagues (2021) for more discussion of differences among study designs focused on external validity (including effectiveness trials and pragmatic studies). Solomon, Cavanaugh, & Draine (2009) provide a helpful introduction to community-based RCTs.

**Developmental timing** refers to the idea that the effect of a person’s environment, or an event they experience, may have dramatically different impacts depending on the developmental timing of the organism’s exposure to that event. For an accessible overview of adolescence as a developmental period, see Steinberg & Morris (2001). For a deeper dive into the history (and future) of developmental science, see Lerner, Lerner, & Buckingham (2023).

**Ecological validity** has to do with whether a study accurately reflects the day-to-day environments or contexts experienced by the population of interest. For more on developmental theory that foregrounds ecological systems, see Bronfenbrenner (1992) and a
discussion of ecologically valid measurement in Shiffman & Stone (1998). See Coppens and Coppinger for more recent work that addresses how ecological validity can help to advance developmental science that is “strengths-based” and “equity-oriented” (2023).

**Face validity** is a type of internal validity that refers to whether something *measures what it appears to measure* (Tanner, 2018). This means it is clear and comprehensible, which is especially important when research findings are intended to be useful or relevant to the community that participates in the research.

**Compounding risk** refers to the notion that disease risks can accumulate for a person who holds multiple identities that are related to higher odds of a particular health outcome. For instance, in the United States individuals who identify as Black or African American experience hypertension at higher rates, as do individuals who are overweight or obese. Those who hold both identities may be at further elevated risk due to these two identity categories. For more on hypertension risk in young people, see Flynn & Alderman (2005).

*From other disciplines or methods:*

**Community-Based Participatory Research** (CBPR) is research that involves active participation by those affected by the issue (i.e., participants are *involved in* the research process as collaborators) and that is oriented towards action or towards social change (Green et al., 1995; Minkler, 2005). CBPR is an orientation to research, not a method, and can take place in the context of quantitative or qualitative research. CBPR can be practiced in a range of social science disciplines, in multidisciplinary research, and by both academic and community-based researchers. For more on CBPR in the co-design of interventions, see work from Perez Jolles and colleagues (2022) and, for getting started on building community partnerships, see work from Green and colleagues (1995).

**Pragmatic clinical trials** are described in clinical medical research as being primarily focused on effectiveness (rather than efficacy) and delivered in the course of normal practice. They measure outcomes that have direct relevance to participants and communities. Most importantly, they are designed to meet the decision-making needs of real-world end users (Zwarenstein et al., 2008). In a set of guidelines on reporting in pragmatic clinical trials, Zwarenstein and his colleagues note that there is “a continuum rather than a dichotomy between explanatory and pragmatic trials” (2008, p. 2).
References


Minkler, M., & Baden, A. C. (2008). Impacts of CBPR on academic researchers, research quality and methodology, and power relations. In M. Minkler & N. Wallerstein (Eds.), *Community-based participatory research for health: From process to outcomes* (pp. 243-262): John Wiley & Sons.


### Table S 1. Descriptive statistics

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Full sample</th>
<th>Control</th>
<th>Treatment</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 254</td>
<td>N = 113</td>
<td>N = 141</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>.51</td>
<td>.43</td>
<td>.57</td>
<td>.03</td>
</tr>
<tr>
<td>Age (years)</td>
<td>16.90</td>
<td>16.61</td>
<td>17.13</td>
<td>&lt;.01</td>
</tr>
<tr>
<td></td>
<td>(1.25)</td>
<td>(1.21)</td>
<td>(1.24)</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>25.27</td>
<td>24.98</td>
<td>25.51</td>
<td>.51</td>
</tr>
<tr>
<td></td>
<td>(6.34)</td>
<td>(6.44)</td>
<td>(6.26)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>.62</td>
<td>.57</td>
<td>.66</td>
<td>.13</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.38</td>
<td>.43</td>
<td>.34</td>
<td>.13</td>
</tr>
<tr>
<td>Puberty stage</td>
<td>3.18</td>
<td>3.09</td>
<td>3.25</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>(.60)</td>
<td>(.61)</td>
<td>(.58)</td>
<td></td>
</tr>
<tr>
<td>Reported frequency of meditation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trained in meditation</td>
<td>.58</td>
<td>.11</td>
<td>.95</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Meditate at all</td>
<td>.67</td>
<td>.32</td>
<td>.94</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Meditate ≥ 3X per week</td>
<td>.43</td>
<td>.15</td>
<td>.66</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic</td>
<td>111.39</td>
<td>112.69</td>
<td>110.34</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>(10.98)</td>
<td>(11.51)</td>
<td>(10.45)</td>
<td></td>
</tr>
<tr>
<td>Diastolic</td>
<td>65.54</td>
<td>65.63</td>
<td>65.46</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>(7.91)</td>
<td>(8.27)</td>
<td>(7.64)</td>
<td></td>
</tr>
</tbody>
</table>

**Risk categories**

- Normal (<90th percentile): .77, .75, .79, .51
- High normal (≥ 90th - 95th percentile): .13, .12, .14, .68
- Hypertensive (≥ 95th percentile): .09, .12, .07, .15
### Table S2. Descriptive statistics by subgroup

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>p-value</th>
<th>Hispanic</th>
<th>Black</th>
<th>p-value</th>
<th>Obese</th>
<th>Not obese</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 129</td>
<td>N = 125</td>
<td></td>
<td>N = 97</td>
<td>N = 157</td>
<td></td>
<td>N = 57</td>
<td>N = 197</td>
<td></td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>16.85</td>
<td>16.95</td>
<td>.50</td>
<td>16.85</td>
<td>16.93</td>
<td>.62</td>
<td>16.78</td>
<td>16.93</td>
<td>.40</td>
</tr>
<tr>
<td>(1.29)</td>
<td>(1.22)</td>
<td></td>
<td></td>
<td>(1.31)</td>
<td>(1.22)</td>
<td></td>
<td>(1.28)</td>
<td>(1.24)</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m^2)</td>
<td>25.53</td>
<td>25.01</td>
<td>.51</td>
<td>26.29</td>
<td>24.65</td>
<td>.05</td>
<td>34.83</td>
<td>22.51</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Black</td>
<td>.64</td>
<td>.60</td>
<td>.56</td>
<td>.53</td>
<td>.65</td>
<td>.11</td>
<td>.47</td>
<td>.36</td>
<td>.11</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.36</td>
<td>.40</td>
<td>.56</td>
<td>.47</td>
<td>.36</td>
<td>.11</td>
<td>.38</td>
<td>.45</td>
<td>.31</td>
</tr>
<tr>
<td>Puberty stage</td>
<td>3.43</td>
<td>2.91</td>
<td>&lt;.001</td>
<td>3.20</td>
<td>3.16</td>
<td>.61</td>
<td>3.21</td>
<td>3.168</td>
<td>.61</td>
</tr>
<tr>
<td>(.54)</td>
<td>(.54)</td>
<td></td>
<td></td>
<td>(.58)</td>
<td>(.61)</td>
<td></td>
<td>(.65)</td>
<td>(.58)</td>
<td></td>
</tr>
<tr>
<td><strong>Reported frequency of meditation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trained in meditation</td>
<td>.65</td>
<td>.51</td>
<td>.02</td>
<td>.52</td>
<td>.62</td>
<td>.11</td>
<td>.52</td>
<td>.60</td>
<td>.29</td>
</tr>
<tr>
<td>Meditate at all</td>
<td>.68</td>
<td>.65</td>
<td>.63</td>
<td>.61</td>
<td>.70</td>
<td>.13</td>
<td>.63</td>
<td>.68</td>
<td>.47</td>
</tr>
<tr>
<td>Meditate ≥ 3X per week</td>
<td>.45</td>
<td>.42</td>
<td>.54</td>
<td>.37</td>
<td>.47</td>
<td>.11</td>
<td>.38</td>
<td>.45</td>
<td>.31</td>
</tr>
<tr>
<td><strong>Blood Pressure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic</td>
<td>107.61</td>
<td>115.28</td>
<td>&lt;.001</td>
<td>111.89</td>
<td>111.07</td>
<td>.57</td>
<td>117.82</td>
<td>109.53</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(9.38)</td>
<td>(11.18)</td>
<td></td>
<td></td>
<td>(11.44)</td>
<td>(10.71)</td>
<td></td>
<td>(12.80)</td>
<td>(9.66)</td>
<td></td>
</tr>
<tr>
<td>Diastolic</td>
<td>66.17</td>
<td>64.88</td>
<td>.20</td>
<td>65.43</td>
<td>65.60</td>
<td>.87</td>
<td>69.14</td>
<td>64.49</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(7.75)</td>
<td>(8.05)</td>
<td></td>
<td></td>
<td>(7.37)</td>
<td>(8.25)</td>
<td></td>
<td>(8.41)</td>
<td>(7.46)</td>
<td></td>
</tr>
<tr>
<td><strong>Risk categories</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal (&lt;90th percentile)</td>
<td>.88</td>
<td>.66</td>
<td>&lt;.001</td>
<td>.79</td>
<td>.76</td>
<td>.51</td>
<td>.53</td>
<td>.84</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High normal (≥90th - 95th percentile)</td>
<td>.06</td>
<td>.21</td>
<td>&lt;.01</td>
<td>.11</td>
<td>.15</td>
<td>.45</td>
<td>.18</td>
<td>.12</td>
<td>.30</td>
</tr>
<tr>
<td>Hypertensive (≥ 95th percentile)</td>
<td>.06</td>
<td>.13</td>
<td>.07</td>
<td>.09</td>
<td>.10</td>
<td>.94</td>
<td>.30</td>
<td>.04</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Table S 3. Whole-school values for balance tests

<table>
<thead>
<tr>
<th></th>
<th>Treatment Group Mean</th>
<th>Control Group Mean</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>15.9</td>
<td>15.6</td>
<td>0.98</td>
</tr>
<tr>
<td>% male</td>
<td>55.4%</td>
<td>61.7%</td>
<td>0.26</td>
</tr>
<tr>
<td>Total</td>
<td>N = 779</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table S4. ITT with administrative controls

<table>
<thead>
<tr>
<th></th>
<th>Systolic</th>
<th></th>
<th>Diastolic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>p-value</td>
<td>B</td>
</tr>
<tr>
<td>Treatment</td>
<td>-2.47</td>
<td>1.12</td>
<td>.03</td>
<td>-1.49</td>
</tr>
<tr>
<td>Female</td>
<td>-7.60</td>
<td>.95</td>
<td>.00</td>
<td>1.44</td>
</tr>
<tr>
<td>Age</td>
<td>1.23</td>
<td>.64</td>
<td>.05</td>
<td>1.33</td>
</tr>
<tr>
<td>BMI</td>
<td>3.72</td>
<td>.71</td>
<td>.00</td>
<td>2.13</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-.38</td>
<td>1.49</td>
<td>.80</td>
<td>-.55</td>
</tr>
<tr>
<td>Reading score</td>
<td>-.22</td>
<td>1.04</td>
<td>.83</td>
<td>-.38</td>
</tr>
<tr>
<td>Math score</td>
<td>.49</td>
<td>1.01</td>
<td>.63</td>
<td>.88</td>
</tr>
<tr>
<td>GPA</td>
<td>-.38</td>
<td>1.55</td>
<td>.81</td>
<td>-.44</td>
</tr>
<tr>
<td>Course failures</td>
<td>-.61</td>
<td>.79</td>
<td>.44</td>
<td>-.74</td>
</tr>
<tr>
<td>Attendance</td>
<td>.01</td>
<td>.01</td>
<td>.45</td>
<td>.01</td>
</tr>
<tr>
<td>Free or reduced lunch</td>
<td>3.15</td>
<td>2.40</td>
<td>.19</td>
<td>.08</td>
</tr>
<tr>
<td>Learning disability</td>
<td>.73</td>
<td>1.64</td>
<td>.66</td>
<td>.21</td>
</tr>
<tr>
<td>Accommodation (504)</td>
<td>-8.89</td>
<td>5.89</td>
<td>.13</td>
<td>-2.65</td>
</tr>
<tr>
<td>Any misconduct</td>
<td>3.42</td>
<td>1.24</td>
<td>&lt;.01</td>
<td>1.10</td>
</tr>
<tr>
<td>Intercept</td>
<td>112.60</td>
<td>5.74</td>
<td>&lt;.001</td>
<td>65.67</td>
</tr>
</tbody>
</table>

Note. Each model presents results from a separate multilevel regression model that clusters standard errors at the classroom level.
Table S5. ITT estimates with models that include randomization block

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B(SE)</td>
<td>p</td>
<td>B(SE)</td>
</tr>
<tr>
<td>Treatment estimate</td>
<td>-2.37</td>
<td>.07</td>
<td>-2.15</td>
</tr>
<tr>
<td>estimate</td>
<td>(1.30)</td>
<td></td>
<td>(1.42)</td>
</tr>
</tbody>
</table>

Model fit statistics

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>BIC</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1873.07</td>
<td>1904.91</td>
<td>1879.32</td>
<td>1918.23</td>
<td>1880.93</td>
</tr>
</tbody>
</table>

Model includes:

- Age (continuous)  
- Randomization block
- Indicator variables for age

Note. Each model presents results from a separate multilevel regression model that clusters standard errors at the classroom level. Models include controls for sex, BMI, race/ethnicity, and pubertal stage.