Teacher Mindsets Help Explain Where a Growth Mindset Intervention Does and Doesn't Work

In press at *Psychological Science*

Authors:
David S. Yeager[1]*, Jamie M. Carroll[1], Jenny Buontempo[1], Andrei Cimpian[2], Spencer Woody[1], Robert Crosnoe[1], Chandra Muller[1], Jared S. Murray[1], Pratik Mhatre[1], Nicole B. Kersting[3], Christopher S. Hulleman[4], Molly Kudym[1], Mary C. Murphy[5], Angela L. Duckworth[6], Gregory M. Walton[7], & Carol S. Dweck[7]

Affiliations:
[1] University of Texas at Austin, Austin, TX, [2] New York University, New York, NY,
[3] University of Arizona, Tucson, AZ, [4] University of Virginia, Charlottesville, VA,
[5] Indiana University, Bloomington, IN, [6] University of Pennsylvania, Philadelphia, PA,
[7] Stanford University, Stanford, CA

* Address correspondence to David S. Yeager (dyeager@utexas.edu).

Acknowledgments:
This paper uses data from the *National Study of Learning Mindsets* (PI: D. Yeager; Co-Is: R. Crosnoe, C. Dweck, C. Muller, B. Schneider, & G. Walton; doi.org/10.3886/ICPSR37353.v1), which was made possible through methods and data systems created by the Project for Education Research That Scales (PERTS, PI: Dave Paunesku), data collection carried out by ICF (Project directors: Kate Flint and Alice Roberts), meetings hosted by the Mindset Scholars Network at the Center for Advanced Study in the Behavioral Sciences, assistance from M. Levi, M. Shankar, T. Brock, C. Romero, C. Macrander, T. Wilson, E. Konar, E. Horng, H. Bloom, and M. Weiss, and funding from the Raikes Foundation, the William T. Grant Foundation, the Spencer Foundation, the Bezos Family Foundation, the Character Lab, the Houston Endowment, and the President and Dean of Humanities and Social Sciences at Stanford University. Writing of this paper was supported by the National Institutes of Health under award number R01HD084772, the National Science Foundation under grant numbers 1761179 and 2004831, the William T. Grant Foundation under grant 189706, the William and Melinda Gates Foundation under grant numbers OPP1197627 and INV-004519, the UBS Optimus Foundation under grant number 47515, and an Advanced Research Fellowship from the Jacobs Foundation to David Yeager. This research was also supported by grant P2CHD042849, Population Research Center, awarded to the Population Research Center at The University of Texas at Austin by the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the National Science Foundation, and other funders.

Abstract

A growth mindset intervention teaches the belief that intellectual abilities can be developed. Where does the intervention work best? A prior paper examined school-level moderators using data from the National Study of Learning Mindsets (NSLM), which delivered a short growth mindset intervention during the first year of high school. This paper uses the NSLM to examine moderation by teachers' mindsets and answers a new question: Can students independently implement their growth mindsets in virtually any classroom culture, or must students' growth mindsets be supported by their teacher's own growth mindsets (i.e., the *mindset + supportive context* hypothesis)? The present analysis (N = 9,167 student records matched with N = 223 math teachers) supported the latter hypothesis. This result stood up to potentially confounding teacher factors and to a conservative Bayesian analysis. Thus, sustaining growth mindset effects may require contextual supports that allow the proffered beliefs to take root and flourish.

Keywords: Wise interventions, Growth mindset, Motivation, Adolescence, Affordances, Implicit theories.

Teacher Mindsets Help Explain Where a Growth Mindset Intervention Does and Doesn't Work

Psychological interventions change the ways that people make sense of their experiences, and have led to improvement in a wide variety of domains of importance to society and to public policy (Harackiewicz & Priniski, 2018; Walton & Wilson, 2018). These interventions offer people new beliefs that encourage them to tackle rather than avoid a challenge or to persist rather than give up. To the extent that people put these beliefs into practice, the interventions can improve outcomes months or even years later (see Brady et al., 2020).

For instance, a *growth-mindset* of intelligence intervention conveys to students the malleability of intellectual abilities in response to hard work, effective strategies, and help from others. Short (<50-minute), online growth mindset interventions evaluated in randomized controlled trials—including two pre-registered replications—have improved the academic outcomes of lower-achieving high school students and first-year college students (e.g., Yeager et al., 2019; see Dweck & Yeager, 2019). These interventions seek to dispel a *fixed mindset*, the idea that intellectual abilities cannot be changed, which has been associated with more "helpless" responses to setbacks and lower achievement around the world (OECD, 2021).

Is successfully teaching students a growth mindset enough? A fundamental tension centers on the role of the educational context. Should psychological interventions be thought of as "giving" people adaptive beliefs that they can apply and reap the benefits from in almost any context, even ones that do not directly support its use? Or do interventions simply offer beliefs that must later be supported by the context if they are to bear fruit?

In a previous paper (Yeager et al., 2019), we examined the role of a school factor, namely, the peer norms in a school, and found that the student growth mindset intervention could not overcome the obstacle of a peer culture that did not share or support growth mindset behaviors, such as challenge seeking. Here we ask how the growth mindset intervention might fare in classrooms led by teachers who endorse more of a fixed mindset (a less supportive context for students' growth mindsets) versus classrooms led by teachers who endorse more of a growth mindset (a more supportive context).

Why Might a Growth Mindset Intervention Depend on Teacher Beliefs?

The present paper tests the viability of the "*mindset + supportive context"* hypothesis. In this hypothesis, a teacher's growth mindset acts as an "affordance" (Walton & Yeager, 2020; also see Gibson, 1977) that can draw out a student's nascent growth mindset and make it tenable and actionable in the classroom.[i] This hypothesis grows out of the recognition that as people try to implement a belief or behavior in a given context, they become aware of whether it is beneficial and legitimate in that context by attending cues in their environments.

According to the *mindset + supportive context* hypothesis, teachers with a growth mindset may convey how, in their class, mistakes are learning opportunities, not signs of low ability, and back up this view with assignments and evaluations that reward continual improvement (Canning, Muenks, Green, & Murphy, 2019; Muenks et al., 2020). This could encourage a student to continue acting on their growth mindsets. By contrast, teachers with more of a fixed mindset may implement practices that make a budding growth mindset inapplicable and locally invalid. For instance, they may convey that only some students have the talent to get an *A*, or say that not everyone is "a math person" (Rattan, Good, & Dweck, 2012; also see Muenks et al., 2020). These messages could make students think that their intelligence would be evaluated negatively if they had to work hard or if they asked a question that revealed their confusion, discouraging students from acting out key growth mindset behaviors. According to this hypothesis, the intervention is like planting a "seed," but one that will not take root and flourish unless the "soil" is fertile (a classroom with growth mindset affordances) (see Walton & Yeager, 2020).

Despite its intuitive appeal, the *mindset + supportive context* hypothesis was not a foregone conclusion. Perhaps students are more like independent agents who can achieve in any classroom context so long as they bring adaptive beliefs to the context and put forth effective effort. Therefore, teachers' mindsets could be irrelevant to the effectiveness of the intervention. Research could even find stronger effects in a classroom led by teachers espousing more of a *fixed* mindset. This would imply that the intervention fortifies students to find ways to achieve (for example, by being less daunted by difficult tasks, working harder, persisting longer) even in contexts that are not directly encouraging these behaviors (Canning et al., 2019; Leslie, Cimpian, Meyer, & Freeland, 2015; Muenks et al., 2020). In this view, a student's growth mindset could be like an asset that can compensate for something lacking in the environment. Because no study has examined classroom context moderators of the growth mindset intervention, a direct test of the *mindset + supportive context* hypothesis was needed.

The Importance of Studying Treatment Effect Heterogeneity

Our attention to teachers' mindsets as a moderating agent continues an important development in psychological intervention research: a focus on treatment effect heterogeneity (Tipton, Yeager, Iachan, & Schneider, 2019). Psychologists have often viewed heterogeneous effects as a limitation, as meaning that the effects are unreliable, small, or applicable in too limited a way, and therefore not important (for a discussion see Miller, 2019).[ii] But this view is shifting. First, heterogeneity is now seen as the way things in the world actually are (Bryan, Tipton, & Yeager, in press; Gelman & Loken, 2014). Nothing, and particularly no psychological phenomenon, works the same way for all people in all contexts. This fact that has been pointed out for generations (Bronfenbrenner, 1977; Cronbach, 1957; Lewin, 1952), but it has only recently begun to be appreciated sufficiently. Second, systematically probing where an intervention does and does not work provides a unique opportunity to develop better theories and interventions (Bryan et al., in press; McShane, Tackett, Böckenholt, & Gelman, 2019), including by revealing mechanisms through which the intervention operates.

The Present Research

This study analyzed data from the National Study of Learning Mindsets (NSLM, Yeager, 2019), which was an intervention experiment conducted with a U.S. representative sample of 9th grade students (registration: osf.io/tn6g4). The NSLM focused on the start of high school because this is when academic standards often rise and when students establish a trajectory of higher or lower academic achievement with lifelong consequences (Easton, Johnson, & Sartain, 2017). The NSLM was designed primarily to study treatment effect heterogeneity. The first paper, as mentioned, focused on a school's peer norms as a moderator (Yeager et al., 2019). The second planned analysis, presented here, focuses on teacher factors. Teachers are important to students directly because they lead the classroom and establish its culture. For example, teachers create the norms for instruction, set the parameters for student participation, and control grading and assessments, and thereby influence student motivation and engagement (Jackson, 2018; Kraft, 2019).

The present focus on math grades (rather than overall GPA as in Yeager et al., 2019) is motivated by the fact that students tend to find math challenging and anxiety-inducing (Hembree, 1990) and therefore a growth mindset might help students confront those challenges productively. Further, our focus on math is relevant to policy. Success in 9th grade math is a gateway to a lifetime of advanced education, profitable careers, and even longevity (Carroll, Muller, Grodsky, & Warren, 2017).

In this study of heterogeneous effects, what kinds of effect sizes should be expected? Brief online growth mindset interventions have tended to improve the grades of lower-achieving high school students by about .10 grade points (or .11 *SD*) (Yeager et al., 2019, 2016). This may seem small relative to benchmarks from laboratory research, but that is not an appropriate comparison for understanding intervention effects obtained in field settings (Kraft, 2020). An entire year of learning in 9th grade math is worth .22 *SD* as assessed by achievement tests (Hill, Bloom, Black, & Lipsey, 2008), and having a high-quality math teacher for a year during adolescence, as compared to an average one, is worth .16 *SD* (Chetty, Friedman, & Rockoff, 2014). Expensive and comprehensive education reforms for adolescents show a median effect of .03 *SD*. The largest effects top out at around 0.20 *SD*, with effects this large representing striking outliers (Boulay et al., 2018). Thus, Kraft (2020) concluded that "effects of .15 or even .10 *SD* should be considered large and impressive" (pg. 248) especially if the intervention is scalable, rigorously evaluated, and assessed in terms of consequential, official outcomes (e.g. grades).

Method

Data

Data come from the NSLM, which as noted was a randomized trial conducted with nationally representative sample of 9th grade students during the 2015-2016 school year (Yeager, 2019). The NSLM was approved by the IRBs at Stanford University, the University of Texas at Austin, and ICF International. The current analysis, which focuses on math teachers, was central to the original design of the study, appeared in our grant proposals, and referenced as the next analysis in our previous pre-analysis plan (osf.io/afmb6/). The present study followed the Yeager at al. (2019) pre-registered analysis plan for every step that could be repeated from the first paper (e.g., data processing, stopping rule, covariates, and statistical model). Analysis steps that are unique to the present paper are outlined in detail in the SOM-R and previewed below. There was no additional pre-registration for the present paper. Instead, we used a combination of a conservative Bayesian analysis and a series of robustness tests to guard against false positives and portray statistical uncertainty more accurately for analysis steps not specified in the pre-registration. The two planned analyses (i.e., for the present paper and Yeager et al., 2019) were conducted sequentially. The present study's math teacher variables were not merged with the student data until after the Yeager et al., (2019) analyses were completed.

The analytic sample included students with a valid condition variable, a math grade, and their math teacher's self-reported mindset (see online supplement Table S6). This sample included 9,167 records (8,775 unique students, as some students had more than one math teacher) nested within 223 unique teachers. It comprises 76% of the overall NSLM sample of students with a math grade. Those who are missing data either could not be matched to a math teacher or their math teacher did not answer the mindset questions. Missing data did not differ by condition (see online supplement Table S7). We retained students who took two math courses with different teachers, each of whom completed the survey. Listwise-deletion of them produced the same results (see Table 2). In terms of math level, 7% of records were from a math class at a level below Algebra 1, 70% were in Algebra 1; 19% were in Geometry, and 3% were in Algebra II or above. Students were 50% female and racially diverse; 14% reported being Black/African-American, 21% Latinx, 6% Asian, 4% Native American or Middle Eastern and 55% white, and 37% reported mothers with a bachelor's degree. Teachers' characteristics were similar to population estimates: 58% were female, 86% were white, non-Latinx, and 51% reported having earned a master's degree; they had been teaching an average of 13.83 years (*SD* = 9.95).

The previous, between-school analysis (Yeager et al. 2019) examined grades in all subjects (math, science, English, and social studies). That analysis focused on the pre-registered group of lower-achieving students (whose pre-treatment grades were below the school median) because it would be harder to detect improvement among the already higher-achieving students and because a previous pre-registered study had shown the effects to be concentrated among the lower achievers (Yeager et al., 2016), which replicated prior work (Paunesku et al., 2015). The current focus on math teachers and math grades, however, required us to include students at all achievement levels, a decision we made before seeing the results. This is because classrooms are smaller units than schools, so excluding half the sample would have left us with too few students in many teachers' classrooms and could have made estimates too imprecise. In addition, math grades are on average substantially lower than in other subjects, probably because students in the U.S. are tracked into advanced math classes earlier than in other subject areas, which suggests that students overall tend to be in math classes that they find challenging. This means that fewer students were already earning As, and more students' grades could improve in response to an intervention, particularly one focused on helping students engage with and learn from challenges. (See Table 2 for supplementary analyses among low-achievers).

Procedure

The NSLM implemented a number of procedures that allowed it to be informative with respect to contextual sources of intervention effect heterogeneity (Tipton et al., 2019). First, students were randomly assigned on an individual basis (i.e., within classroom and school) to a growth mindset intervention or a control group, while math teachers (who were unaware of condition and study procedures) were surveyed to measure their mindsets. Thus, each teacher in the analytic sample had some students in the control group and some students in the treatment group. Consequently, we could estimate a treatment effect for each teacher and examine variation in effects across teachers. The study procedures appear in Figure 1 and are described in more detail next. Additional information is reported in the technical documentation available from ICPSR (Yeager, 2019) and in the supplemental material in Yeager et al. (2019).

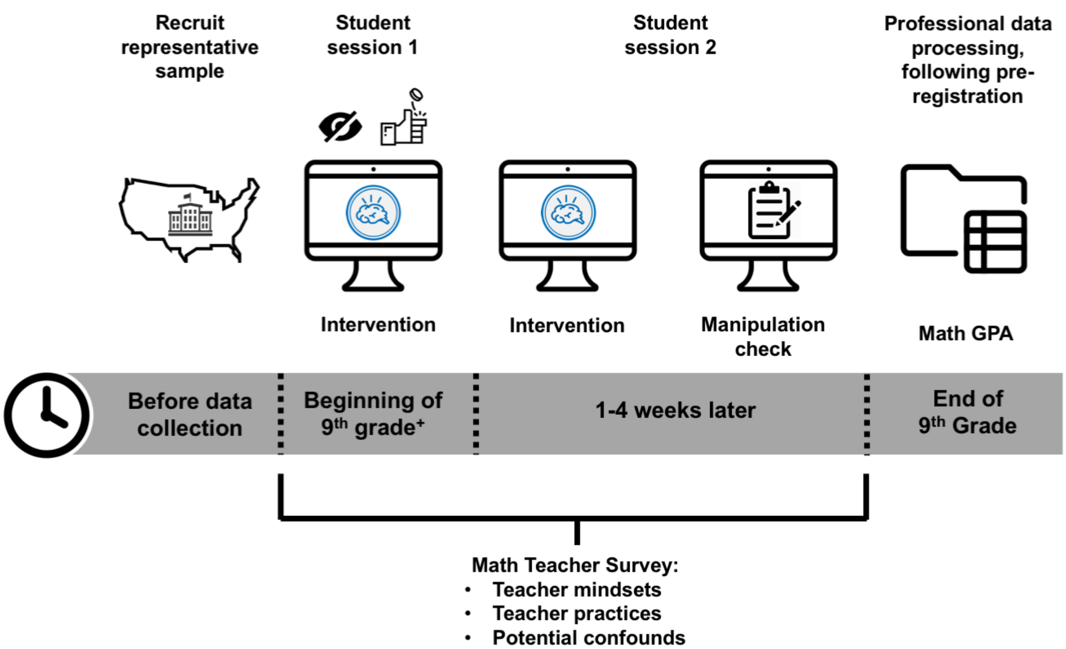

Figure 1. The student and teacher data collection procedure in the National Study of Learning Mindsets. Note: The non-seeing eye icon represents masking of condition assignments from teachers, students, and researchers. The coin flip icon represents the moment of random assignment. +About 85% of students took session 1 during the fall semester before the Thanksgiving break, as planned; the rest took it in January or early February, to accommodate school schedules. The pre-analysis plan specifying data processing rules can be found here: osf.io/afmb6/. The data processing was carried out by MDRC, an independent research firm, while unaware of condition assignment or results.

Data collection and processing. To reduce bias in the research process, three professional research firms were contracted to form the sample, administer the intervention, and collect all the data. ICF International selected and recruited a nationally-representative sample of public schools in the U.S. during the 2015-2016 academic year. Students within those schools completed online surveys hosted by the firm PERTS, during which they were randomly assigned to a growth mindset intervention or a control group. The final student response rates were high (median student response rate across schools: 98%), and the recruited sample of schools closely matched the population of interest (Gopalan & Tipton, 2018).

Random assignment to condition was conducted by the survey software at the student level, with 50/50 probability, when students logged on to the survey for the first time. To prevent expectancy effects, condition information was masked from involved parties, in that students did not know there were two conditions (i.e. a "treatment" and a "control") while teachers in the school were not allowed to "take" the treatment, were not told the hypotheses of the study, and were not told that students were randomly assigned to alternative conditions. The treatment and control conditions looked remarkably similar, to reduce the likelihood that teachers saw a difference. The intervention sessions generally occurred during electives (like health or PE), and schools were discouraged from conducting sessions in math classes. Math teachers were not used as proctors (usually, non-teaching staff coordinated data collection) so as to keep math teachers as unaware of the study as possible. The intervention involved two ~25-minute sessions, generally 1 to 4 weeks apart, and under 50 minutes in total for nearly all students. Immediately after the second intervention session, students completed self-reports of mindsets (which served as a manipulation check).

Prior to data collection, schools provided the research firm with a list of all instructors who taught a math class that academic year with more than two 9th grade students—the definition of a "9th grade math teacher" used here. This sample restriction was necessary because each teacher would need both treated and control students to provide a within-teacher treatment effect estimate. All such teachers were invited to complete an approximately one-hour online survey in return for a $40 incentive, and a large majority of teachers (86.8%) did so. This high response rate reduced the likelihood that biased non-response could have affected the distribution or validity of the teacher mindset measure.

The independent research firm ICF International obtained student survey data from the technology vendor PERTS and administrative data (e.g., grades) from the schools and readied both for final processing. MDRC, another independent research firm, then processed these data following a registered pre-analysis plan. They were all unaware of students' condition assignments. Only then did our research team access the data and execute the planned analyses. (In parallel, MDRC developed an independent evaluation report that reproduced the overall intervention impacts and between-school heterogeneity results, Zhu, Garcia, & Alonzo, 2019).

Growth mindset intervention. The growth mindset intervention presented students with information about how the brain learns and develops using this metaphor: *The brain is like a muscle that grows stronger (and smarter) when it learns from difficult challenges* (Aronson, Fried, & Good, 2002). Then, the intervention unpacked the meaning of this metaphor for experiences in school, namely that struggles in school are not signs that one lacks ability but instead that one is on the path to developing one's abilities. Trusted sources—scientists, slightly older students, prominent individuals in society—provided and supported these ideas. Students were then asked to generate their own suggestions for putting a growth mindset into practice; for example, by persisting in the face of difficulty, seeking out more challenging work, asking teachers for appropriate help, and revising one's learning strategies when needed, among others.

The intervention involved a number of other exercises designed to help students articulate the growth mindset, how they could use it in their lives, and how other students like them might use it. It was deliberately not a lecture or an "exhortation," so as to avoid the impression that the intervention was telling young people what to think, since we know that for adolescents an autonomy-threatening framing could be ineffective or even backfire. Instead, the intervention treated young people as collaborators in the improvement of the intervention, sharing their own unique expertise on what it is like to be a high school student. Additional detail on the intervention (and control) groups appears in the supplement to Yeager et al., (2019) (also see the SOM-R).

Control group. The control group was provided with interesting information about brain functioning and its relation to memory and learning, but the program did not mention the malleability of the brain or intellectual abilities. As in the growth mindset condition, trusted sources—scientists, older peers, and prominent individuals in society—provided this information and students were asked for their opinions and treated as having their own unique expertise. The graphic art, headlines, and overall visual layout was very similar to the treatment, to help students and teachers remain masked and to discourage comparison of materials. Because most students were taking biology at the time, the neuroscience taught in the control group would have added content above and beyond what students were learning in class and could even have increased interest in science and in school. Indeed, students have sometimes found the control material if anything more interesting than the treatment material (Yeager, Romero, et al., 2016). In sum, the active control condition was designed to provide a rather rigorous test of the effectiveness of the growth mindset intervention.

Measures

Primary outcome: Math grades. The primary dependent variable was students' post-treatment grades in their math course, which were generally recorded 7 or 8 months after the intervention. All math grades were obtained from schools' official records. Grades ranged from 0 (an *F*) to 4.3 (an *A+*). The mean math GPA was 2.44 leaving considerable room for improvement for many students.

Grades are the dependent variable of interest, not test scores, for three reasons. First, grades are typically better predictors of college enrollment and lifespan outcomes than test scores, and the signaling power of grades is apparent even though schools and teachers could potentially inflate their grading scales (Pattison, Grodsky, & Muller, 2013). Thus, grades are relevant for policy and for understanding trajectories of development. Second, grades represent the accumulation of many different assignments (homework, quizzes, tests) and therefore signal the kind of dedicated persistence that a growth mindset is designed to instill. Third, test scores were not an option in this study because 9th grade is not always a grade in which state achievement tests are administered, and most students did not have a math test score.

Primary moderator: Teacher mindset. Math teachers rated two fixed mindset statements: "People have a certain amount of intelligence and they really can't do much to change it" and "Being a top math student requires a special talent that just can't be taught" (1=*Strongly agree*, 6=*Strongly disagree*, M = 4.74, SD = 0.76). The first is a general fixed mindset item intended to capture beliefs that might lead to mindset practices that are not specific to math, such as not allowing students to revise and resubmit their work or discouraging low-achievers' questions. The second item captures a belief that could lead to more math-specific mindset practices (see Leslie, Cimpian, Meyer, & Freeland, 2015). The two items were correlated (r = .48, p < .001) and were averaged. We scored them so that higher values corresponded to more growth mindset beliefs. We note that respondent time on this national math teacher survey was limited to encourage participation and survey completion, so every construct, even teacher mindset, was limited to a small number of items.

The two mindset items used for the composite had not been administered to large samples of high school math teachers before, so we assessed their concurrent validity by administering them to a large, pilot sample of high school teachers along with items that assessed teacher practices ($N = 368$ teachers). (The details of the sample and the exact item wordings are reported in the SOM-R.) In the pilot, we found that teachers' mindsets in fact predicted their endorsement of practices expected to follow from teachers' mindsets, based on theory and past research (Canning et al., 2019; Haimovitz & Dweck, 2017; Leslie et al., 2015; Muenks et al., 2020). Specifically, teachers' endorsement of a growth mindset was positively associated with *learning-focused* practices, $r = .30$, $p<.001$ (e.g., saying to a hypothetical struggling student, "Let's see what you don't understand and I'll explain it differently," and *not* agreeing that, "It slows my class down to encourage lower achievers to ask questions"). Further, teacher mindsets were negatively associated with *ability-focused* practices (emphasizing raw ability and implying that high effort was a negative sign about ability), $r = .28$, $p<.001$ (e.g., comforting a hypothetical struggling student with "Don't worry, it's okay to not be a math person," *a la* Rattan, Good, & Dweck, 2012, and praising a succeeding student with "You're lucky that you're a math person" or "It's great that it's so easy for you"). This is by no means an exhaustive list of potential mindset teacher practices, and this is certainly not the only way to measure teacher practices. But this validation study suggests that the teacher mindset measure captures differences in teachers that extend to classroom practices—practices that the student growth mindset treatment could either overcome or that could afford the opportunity for it to work.

Potential confounds for teacher mindsets. Because only the student mindset intervention was randomly assigned, and not teachers' mindsets, other characteristics of teachers could be correlated with their mindsets and with the magnitude of the intervention effect. For instance, perhaps teachers' growth mindsets are simply a proxy for competent and fair instructional practices in general. To account for this possibility, we measured several potential confounds for teacher mindsets: a video-based assessment of pedagogical content knowledge, a fluid intelligence test for teachers, teachers' masters-level preparation in education or math, and an assessment of implicit racial bias. We call these "potential" confounds because, during the design of the study, these were raised by at least one advisor to the study as something that could interfere with the interpretation of teacher mindsets (although, in the end, these factors showed rather weak associations with teacher mindsets; see Table S10). To this list of *a priori,* theoretically-motivated teacher confounds, we added teacher race, gender, years teaching, and whether they had heard of growth mindset before. We describe the potential confounds in the supplement because their inclusion or exclusion does not change the sign, significance, or magnitude of the key moderation results. To these potential teacher-level moderators we can also add the pre-registered school-level moderators (challenge-seeking norms among students/peers, school achievement level, and school percent racial/ethnic minority; see Yeager et al. 2019). Adding these school factors in interaction with the treatment did not change the teacher mindset interaction (see Table 2), suggesting that these factors examined previously (Yeager et al., 2019) and the classroom-level factors examined here account for independent sources of moderation. Last, in a *post-hoc* analysis we examined three student perceptions of the classroom climate that could be confounded with teacher mindset: the level of cognitive challenge in the course, how interesting the course was, and how much students thought the teacher was "good at teaching."

None of these factors were moderators and none altered the teacher mindset interaction (see Table S11 in the SOM-R).

Manipulation check and moderator: Students' mindset beliefs. At pre-test and again at immediate post-test participants indicated their level of agreement with the three fixed-mindset statements used as a manipulation check by Yeager et al. (2019) (e.g. "You have a certain amount of intelligence, and you really can't do much to change it.", 1 = *Strongly agree,* 6 = *Strongly disagree*). We averaged responses (Pre-test, $M = 2.95$, $SD = 1.14$; $\alpha = .72$; Post-test, $M = 2.70$, $SD = 1.19$; $\alpha = .78$), and higher values corresponded to more of a fixed mindset. An extensive discussion of the validity of this three-item mindset measure and its relation to the growth mindset "meaning system" appears in Yeager and Dweck (2020). The scale at pre-test was used in exploratory moderation analyses. The scale at post-test was used as a planned manipulation check.

Student-level covariates. Student-level control variables related to achievement included: the pre-treatment measure of low-achieving student status specified in the overall NSLM pre-analysis plan (osf.io/afmb6/), which indicates that the student received an 8th grade GPA below the median of other incoming 9th graders in the school; students' expectations of how well they would perform in math class ("Thinking about your skills and the difficulty of your classes, how well do you think you'll do in math in high school?"; 1=*Extremely poorly* to 7=*Extremely well*); students' racial minority status, gender, and whether their mother had earned a bachelor's or above. These covariates were specified in the NSLM pre-analysis plan because each could be related to achievement, and so a chance imbalance with respect to any of these within a teacher's classroom could bias treatment effect estimates. Controlling for these factors reduces the influence of chance imbalances. Covariates were school-mean-centered.

Analysis Plan

Estimands. The primary analysis focused on the sign and significance of the student growth mindset intervention × teacher growth mindset interaction. If the interaction was *positive* and significant it would be more consistent with the *mindset + supportive context* hypothesis.

The primary estimands of interest (i.e., values we wished to estimate) were the simple effects listed in Table 1. Row 1 assumes that teacher mindsets are unassociated with other teacher factors, but this is not sufficiently conservative so it is not our primary analysis of interest. Row 2 in Table 1 accounts for potential confounding in the interpretation of teachers' mindsets by fixing the levels of potentially-confounding moderators to their population averages (denoted by *c* in Table 1) and looking at the moderated effects of teacher mindsets (see row 2 of Table 1). Thus, later when we present the key results in the paper in Table 2, those estimates correspond to the estimands in row 2 of Table 1.

Table 1. Estimands of Interest: Conditional Average Treatment Effects (CATEs).

	Teachers reporting fixed mindsets	Teachers reporting growth mindsets (i.e. *mindset + supportive context*)
Assuming no confounding of the moderator	$\text{CATE}_{S=\text{Fixed}} = \mathbb{E}[Y_{ij} \mid T_{ij} = Mindset, S_j = Fixed] - \mathbb{E}[Y_{ij} \mid T_{ij} = Control, S_j = Fixed]$	$\text{CATE}_{S=\text{Growth}} = \mathbb{E}[Y_{ij} \mid T_{ij} = Mindset, S_j = Growth] - \mathbb{E}[Y_{ij} \mid T_{ij} = Control, S_j = Growth]$
E Adjusting for potential confounding (*primary estimand of interest*)	$\text{CATE}_{S=\text{Fixed}, C=c} = \mathbb{E}[Y_{ij} \mid T_{ij} = Mindset, S_j = Fixed, C_j = c] - \mathbb{E}[Y_{ij} \mid T_{ij} = Control, S_j = Fixed, C_j = c]$	$\text{CATE}_{S=\text{Growth}, C=c} = \mathbb{E}[Y_{ij} \mid T_{ij} = Mindset, S_j = Growth, C_j = c] - \mathbb{E}[Y_{ij} \mid T_{ij} = Control, S_j = Growth, C_j = c]$

Note: CATE = Conditional average treatment effect, or the treatment effect within a subgroup. *i* indexes students, *j* indexes teachers, *Y* = math grades, *T* (for treatment) = treatment status, *S* = teacher mindset, *C* (for confounds) = vector of teacher mindset confounds, *c* = population average for potential teacher or school confounds. See proofs and justifications in Yamamoto and Yeager (2019).

Primary statistical model: Linear mixed effects analysis. The primary analysis examined the cross-level interaction using a typical multilevel, linear mixed effects model, with a random treatment effect that varied across teachers and was predicted by teacher-level factors, but with one twist: fixed teacher intercepts. Such a model has become the standard approach for multi-site trial heterogeneity analyses (Bloom, Raudenbush, Weiss, & Porter, 2017) because the fixed intercept for each group prevents biases from chance imbalances in the random assignment to treatment within small groups. This "hybrid" (fixed intercept, random slope) approach can make a big difference in the present analysis, since some teachers may have small numbers of students and, due to random sampling error, be more likely to have chance imbalances.[iii] This is why the fixed intercept, random slope model was specified in the NSLM pre-analysis plan (Yeager et al., 2019). As in all standard multilevel models, the random slope allows different teachers' students to have different treatment effects, but uses corrections to avoid overstating the heterogeneity (called an empirical Bayesian shrinkage estimator). Specifically, the model we estimate appears in Eq. 1,

$$y_{ij} = \alpha_j + \sum_{h=1}^{p} \beta_h x_{ijh} + \left[\tau_1 s_j + \sum_{l=2}^{k} \tau_l c_{jl} + \gamma_j\right] t_{ij} + \epsilon_{ij} \quad (1)$$

where y_{ij} is the math grade for student *i* in teacher *j*'s classroom. At the student level, x is a vector of *k-2* student-level covariates (prior achievement, prior expectations for success, race/ethnicity, gender, and parents' education, all school-centered). At the teacher level, α_j is a fixed intercept for each teacher. The large section in brackets represents the multi-level moderation portion of the model, our main interest. The student-level treatment status, t_{ij}, is interacted with the continuous measure of teachers' mindset beliefs (s_j) with controls for potential confounds of teacher mindset beliefs (c_j, a vector that includes implicit bias, pedagogical content knowledge, fluid intelligence, and teacher master's certification). The teacher-level random error is γ_j and the student-level error term is ϵ_{ij}.

The primary hypothesis test concerns the regression coefficient τ_1, which is the cross-level student treatment × teacher mindset interaction. When τ_1 is positive and significant, it means that treatment effects are higher when teachers' growth mindset scores are higher. The case for a stronger interpretation of τ_1 is bolstered if the coefficient's sign and significance persists even when accounting for the potential confounds indexed by c_j. The model in Eq. 1 allows s_j (teachers' mindsets, the primary moderator) to remain a continuous variable. We estimated the CATEs in Table 1 by implementing a standard approach in psychology: calculating the treatment simple effect at -1 *SD* (teachers reporting relatively more of a fixed mindset) and +1 *SD* (teachers reporting relatively more of a growth mindset), while holding confounding moderators constant. We used the *margins* post-estimation command in Stata SE to do so. We call the former teachers "relatively" more fixed mindset because their position on the scale suggests they are in an intermediate group, not clearly growth mindset, but, on the whole, not extremely fixed.

Secondary statistical model: Bayesian analysis. The primary model had at least one major limitation: it presumed that all student and teacher-level variables had linear effects and did not interact. The pre-analysis plan for the NSLM therefore stated that we would follow-up the primary analysis by using a multi-level application of a flexible but conservative approach called Bayesian Causal Forest (BCF), which relaxes the assumptions of nonlinearity and of no higher-order interactions. BCF has been found, in multiple open competitions and simulation studies, to detect true sources of complex treatment effect heterogeneity while not lending much credence to noise (Hahn, Murray, & Carvalho, 2020). See Eq.2:

$$y_{ij} = \alpha_j + \beta(\boldsymbol{x}_{ij}) + [\tau(\boldsymbol{s}_j, \boldsymbol{c}_j) + \gamma_j]t_{ij} + \epsilon_{ij} \qquad (2)$$

The BCF model in Eq. 2 retained the key features of the primary statistical model in Eq. 1: teacher-specific intercepts, student-level covariates, random variation in the treatment effect across teachers (unexplained by covariates), and potential confounds for teacher mindset beliefs (collected in the vector $\boldsymbol{c}_{ij}$). The most notable change is that BCF replaces the additive linear functions from the primary model with the nonlinear functions $\beta(\boldsymbol{x})$ and $\tau(\boldsymbol{s}, \boldsymbol{c})$. These nonlinear functions have "sum-of-trees" representations that can flexibly represent interactions and other non-linearities (thus avoiding the researcher degree of freedom of specifying a functional form), and that can allow the data to determine how and whether a given covariate contributes to the model predictions (thus avoiding the researcher degree of freedom of covariate selection). The nonlinear functions are estimated using machine-learning techniques. Bayesian Additive Regression Trees (BART) prior distributions that shrink the functions toward simpler structures (like additive or nearly additive functions) while allowing the data to speak. See the SOM-R for more detail about the priors used for BCF.

From the BCF output, there is no single regression coefficient to interpret, as there would be in a typical linear regression model, because the output of the BCF model is a richer posterior distribution of treatment effect estimates for each of the 9,167 teacher mindset/student grade records in the sample. This means that we do not have to set the moderator to + or -1 *SD*. Instead, we can summarize the subgroup treatment effects for each level of teacher mindsets, while holding all of the potential confounds constant at their population means (see Figure 2 for the plot). We note that conducting subgroup comparisons or hypothesis tests does not entail changes to the model fit or prior specifications. The data were used exactly one time, to move from the prior distribution over treatment effects to the posterior distribution. This facilitates honest Bayesian inference concerning subgroup effects and subgroup differences, and eliminates concerns with multiple hypothesis testing that can threaten the validity of a frequentist p-value (Woody, Carvalho, & Murray, 2020).

The BCF analysis had another advantage: it could accommodate the fact that there were researcher degrees of freedom about *which* aspect of math classrooms might moderate the treatment effect—teacher mindsets, the other teacher variables, or qualities of the schools in which teachers were embedded. BCF allowed all of these teacher and school factors to have the same possibility of moderating the treatment effect, and gave them equal likelihood in the prior distribution. In other words, BCF built uncertainty into the model output, which helped to guard against spurious findings (see the SOM-R).

Results

Preliminary Analyses

Effectiveness of random assignment. The intervention and control groups did not differ in terms of pre-random-assignment characteristics (see Table S5 and see Yeager et al. 2019).

Average effect on the manipulation check. The manipulation check was successful on average. The growth mindset intervention led students to report lower fixed mindset beliefs relative to the control group, (Control M = 2.91, SD = 1.17; Growth mindset M = 2.48, SD = 1.16), t=16.82, p<.001, d = .37.

Homogeneity of the manipulation check. The immediate treatment effect on student mindsets (the beliefs students reported on the post-treatment manipulation check) was not significantly moderated by teachers' mindsets, B = .04 [95% CI: -.031, .102], t = 1.04, p = .297. Further, there was very little cross-teacher variability in effects on the manipulation checks to explain. According to the BCF model's posterior distribution, the standard deviation of the intervention effect across teachers was just 5% of the average intervention effect, which means that the posterior prediction interval ranged from 90% to 110% of the average intervention effect, a very narrow range. Here is what this means: treated students, regardless of their math teacher mindsets, ended the intervention session with similarly strong growth mindsets that could be tried out. If we later found heterogeneous effects on math grades, measured months into the future, it could reflect differences in the affordances that allowed students to act on their mindsets in class.

Preliminary analyses of effect on math grades. A previous paper (Yeager et al., 2019, Extended Data Table 1) and an independent impact evaluation (Zhu et al., 2019) reported the significant main effect of the growth mindset treatment on math grades for the sample overall ($p = .001$). Next, the present study's BCF model found that there was about as much heterogeneity in treatment effects across teachers (47% of the variation) as there was across schools (49%, with the remaining 4% of variation coming from covariation between the two). Combined, these analyses mean that the present paper was justified in focusing on heterogeneity in the treatment effect on math grades independently from the school factors reported by Yeager et al., (2019).

Primary Analyses: Moderation by Teachers' Mindsets

Linear mixed effects model. Teachers' mindsets positively interacted with the intervention effect on math grades: Student intervention × Teacher mindset interaction $B = .09$ [95% CI: .026, .150], $t = 2.79$, $p = .005$ (see Eq. 1). This result was robust to changes to the model, including consideration of the school-level moderators previously reported by Yeager et al. (2019), and changes in the sub-sample of participating students (see Table 2).

Thus the data were consistent with the mindset + supportive context hypothesis: the intervention could alter students' mindsets, but a growth-affording context was necessary for students' grades to be improved. Students whose teachers did not clearly endorse growth mindset beliefs showed a significant manipulation check effect immediately after the treatment, but their math grades did not improve.

Effect sizes. The CATEs (conditional average treatment effects) for students with more fixed versus more growth mindset teachers are presented in Table 2. The effect for students in classrooms with growth mindset teachers was 0.11 grade points and was significant at $p<.001$, and there was no significant effect in classrooms of teachers reporting more of a fixed mindset (compare columns 2 and 3). Notably, our primary analyses did not exclude students whose grades could not have been lifted any further. If we limit our sample to the three-fourths of students who were not already making straight *A*s across all of their core classes before the study, and who therefore had room to improve their grades, the estimated effect among students in classrooms with growth mindset teachers becomes slightly larger, .14 grade points (see row 5, Table 2).

The present analysis included a representative sample and used "intent-to-treat" analyses. This means that we included students who could not speak or read English, who had visual or physical impairments, who had attentional problems, whose computers malfunctioned, and more. Thus, there were many students in the data who could not possibly have shown treatment effects. This study therefore estimates effects that could be anticipated under naturalistic circumstances.

Table 2. Effect of Growth Mindset Intervention on Math Grades in 9th Grade Among Students with Fixed Versus Growth Mindset Math Teachers, Estimated in Linear Mixed Effects Models.

Model specification	Teachers reporting more of a **fixed mindset**	Teachers reporting more of a **growth mindset**	Student intervention × Teacher mindset (continuous) interaction
	Primary Model Specification		
Teacher mindset as moderator + potential teacher confounds, (N = 9,167)	CATE = -.02 [-.074, .038], $t = -0.63$, $p = .531$	**CATE = .11 [.046, .167], $t = 3.46$, $p < .001$**	B = .09 [026, .150], $t = 2.79$, $p = .005$
	Robustness Test: Accounting for School-Level Moderators from Yeager et al. (2019)		
Plus school-level moderators, (N = 9,167)	CATE = -.02 [-.075, .039] $t = 0.61$, $p = .542$	**CATE = .11 [.045, .168] $t = 3.37$, $p < .001$**	B = .09 [.025, .151], $t = 2.76$, $p = .006$
	Robustness Tests: Alternative Sub-samples#		
Only students with only one math teacher, (N=8,383)	CATE = -.04 [-.108, .026] $t = -1.20$, $p = .230$	**CATE = .11 [.040, .170] $t = 3.18$, $p = .001$**	B = .09 [.028, .159], $t = 2.81$, $p = .005$
Only previously-lower-achieving (i.e. below-median pre-intervention GPA) students †, (N=4,811)	CATE = .02 [-.050, .097] $t = 0.63$, $p = .527$	**CATE = .13 [.067, .196] $t = 4.01$, $p < .001$**	B = .09 [.008, .165], $t = 2.17$, $p = .030$
Only students previously without straight *A*s, (N=6,958)	CATE = -.01 [-.062, .041] $t = -0.39$, $p = .696$	**CATE = .14 [.071, .203] $t = 4.07$, $p < .001$**	B = .11 [.040, .180], $t = 3.10$, $p = .002$

Note: CATE = Conditional average treatment effect, in GPA units (0 to 4.3 scale) estimated with the *margins* postestimation command in Stata SE, holding potentially-confounding moderators constant at their population means. All CATES estimated using teacher survey weights provided by ICF International to make the estimates generalizable to the nation as a whole. Teachers with more of a growth mindset in this analysis are those reporting mindset at +1 *SD* for the continuous teacher mindset measure, while teachers with more of a fixed mindset are at -1 *SD*. Numbers in brackets represent 95% confidence intervals. Regression model specified in Eq. 1. B = unstandardized regression coefficient (i.e. expected treatment effects on GPA). † This was the pre-registered subgroup in Yeager et al. (2019). # Models included all teacher-level moderators.

Bayesian machine-learning analysis. The BCF analyses yielded conclusions consistent with the primary linear mixed effects model. First, there was a positive zero-order correlation of $r(223) = .55$ between teachers' mindsets and the estimated magnitude of the classroom's treatment effect (i.e., the posterior mean for the CATE for each teacher), which mirrors the moderation results of the primary linear model. Figure 2, which depicts the posterior distribution for each level of teacher mindset, holding all other moderators constant at the population mean, shows no overlap between the interquartile range (IQR) for teachers with more of a growth mindset (5 or 5.5) and the IQR for teachers with more of a fixed mindset (4 or lower). This supports the conclusion of a positive interaction, again consistent with the mindset + supportive context hypothesis.

The model also shows that teachers who strongly endorse growth mindset beliefs show a positive average intervention effect greater than zero with approximately 100% certainty (see Figure 2), confirming the results of the simple effects analysis from the linear model. We note again that the BCF model is relatively conservative. It utilizes a prior distribution centered at a homogeneous treatment effect of zero. This should be taken as strong evidence of moderation and strong evidence that the intervention was effective for students of growth mindset teachers. The BCF analysis also yielded new evidence that extended the primary linear model's results. Figure 2 shows that teachers' growth mindsets were related to higher treatment effect sizes in a linear fashion for most of the distribution, but there was no marginal increase in treatment effects when teachers endorsed a growth mindset to an even greater extent once they were already high on the scale (see the rightmost groups of teachers in Figure 2). The non-linearity, discovered by the BCF analysis, should invite further investigation into whether teachers already endorsing a very high growth mindset are using practices that encourage all of their students (even those in the control group) to engage in growth mindset behaviors, potentially narrowing the contrast between treatment and control group students.

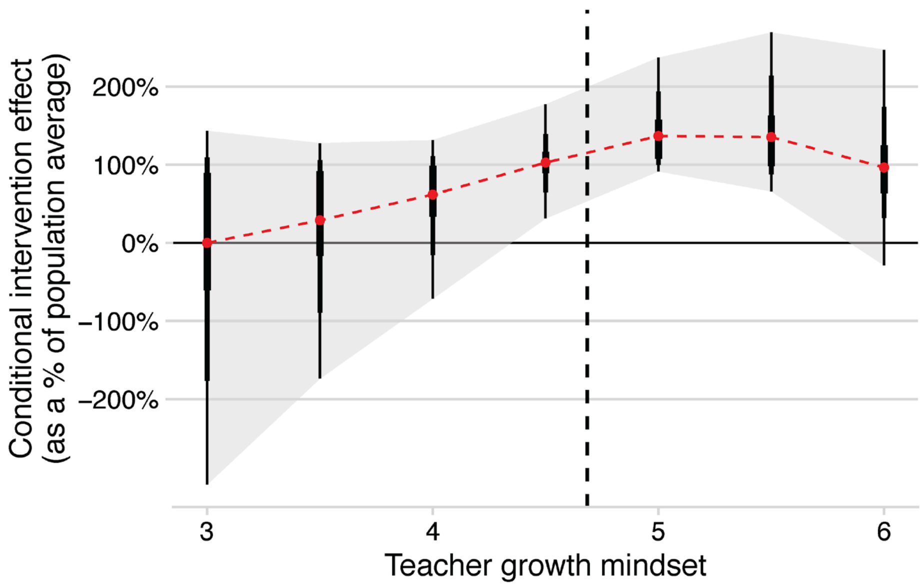

Figure 2. Evidence for the mindset + supportive context hypothesis regarding teacher mindsets and a student mindset intervention—up to a point—in a flexible Bayesian Causal Forest model. Note: Posterior distributions are of the conditional average treatment effect (CATE), as a proportion of the average treatment effect (ATE). Thus, 100% means the CATE is equal to the population ATE. Red dots represent the estimated intervention effect (posterior means) at each level of teacher mindset. The widths of the bars, from wide to narrow, represent the middle 50% (i.e., IQR), 80% and 90% of the posterior distribution, respectively. The teacher mindset measure ranges from 1 to 6. The dashed vertical line represents the population mean for teacher mindsets. However, the x-axis stops at 3 because only five teachers had a mindset score below this and the model cannot make precise predictions with so few teachers.

Exploratory Analyses of Baseline Student Mindsets

The brief, direct-to-student growth mindset intervention did not appear to overcome local *contextual* factors that can suppress achievement (e.g., a teacher with a fixed mindset). Could it address *individual* risk factors suppressing achievement, such as the student's own fixed mindset? A slight suggestion of this possibility appeared in one of the original growth-mindset intervention experiments (Blackwell, Trzesniewski, & Dweck, 2007); a student's prior growth mindset negatively interacted with the intervention effect, but the result was imprecise (p = .07). To revisit this question, we added students' baseline mindsets as a moderator in the present study's primary linear mixed effects model. We found a significant negative interaction with student baseline growth mindsets, $B = -.06$ [95% CI: .018, .098], $t = 2.85$, $p = .004$, suggesting stronger effects for students with more of a fixed mindset. Thus the (marginal) Blackwell et al. (2007) moderation finding was borne out. This interaction was additive with, but not interactive with, the teacher mindset interaction, which did not change in magnitude or significance by including the student mindset interaction (two-way still $p = .005$; three-way interaction $p > .20$). Exploring the CATEs, students reporting more fixed mindsets at baseline (- 1 *SD*), in classrooms with a teacher reporting more of a growth mindset (+1 *SD*), showed an intervention effect on their math grades of 0.16 grade points [0.079, 0.234], $t = 3.957$, $p<.001$. By contrast, and there was no significant effect among students who already reported a strong growth mindset in growth mindset classes, and, as noted, no effect overall in more fixed-mindset classes.

Discussion

In this nationally-representative, double-blind clinical trial, successfully teaching a growth mindset to students lifted math grades overall, but this was not enough for all students to reap the benefits of a growth mindset intervention. Supportive classroom contexts also mattered. Students who were in classrooms with teachers who espoused more of a fixed mindset did not show gains in their math grades over 9th grade compared to the control group, whereas students in classrooms with more growth mindset teachers showed meaningful gains. This finding suggests that students cannot simply carry their newly enhanced growth mindset to any environment and implement it there. Rather, the classroom environment needs to support, or at least permit, the mindset, by providing necessary affordances (see Walton & Yeager, 2020).

In addition, we discovered that students who formerly reported more of a fixed mindset and who went back into a classroom with a teacher who had more of a growth mindset showed larger gains in achievement than did students who began the study with more of a growth mindset. This finding supports the Walton and Yeager (2020) hypothesis that individuals at the intersection of vulnerability (prior fixed mindset) and opportunity (high affordances) are the most likely to benefit from psychological interventions.

The national sampling, and the use of an independent firm to administer the intervention, permits strong claims of generalizability to U.S. public high school math classrooms. Future studies could use or adapt a similar methodology to assess generalizability to other age groups, content areas, or cultural contexts. In general, materials may need to be adapted, sometimes extensively (see Yeager et al., 2016), to be appropriate to new settings.

A main limitation in our study is that teachers' mindsets were measured, not manipulated. The fact that teacher mindsets were moderators above and beyond other teacher confounders lends support to our hypotheses about the importance of classroom affordances. But more research is needed to determine whether teachers' mindset beliefs, or the practices that follow from them, play a direct, causal role. Thus, the mindset × context approach opens the window to a new, experimental program of research.

If a future experimental intervention targeted both students *and* teachers, what kinds of moderation patterns might be expected? There, we actually might see the largest effects for formerly fixed mindset teachers. That is, the benefits of planting a seed *and* fertilizing the soil should be greatest where soil was formerly inhospitable, and smaller where the soil was already adequate.

In general, we view the testing and understanding of the causal effect of teacher mindsets as the next step for mindset science—followed, if successful, by the creation of programs to promote more growth-mindset-sustaining classroom practices. Such research will be challenging to carry out, however. For example, we do not think it will be enough to simply copy or adapt the student intervention and provide it to teachers. A new intervention for teachers will need to be carefully developed and tested. We do not yet know which teacher beliefs or practices (or combinations thereof) may be most important in which learning environments. Even if we did, there is much to be learned about how to best encourage and support key beliefs and practices in teachers. The current findings, along with other recent findings about the importance of instructors' mindsets in promoting achievement for all groups and reducing inequalities between groups (Canning et al., 2019; Leslie et al., 2015; Muenks et al., 2020), point to the urgency and value of this research.

References

Aronson, J. M., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38*(2), 113–125. doi: 10.1006/jesp.2001.1491

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (in press). Persistence and fadeout of educational intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246–263. doi: 10.1111/j.1467-8624.2007.00995.x

Bloom, H. S., Raudenbush, S. W., Weiss, M. J., & Porter, K. (2017). Using multisite experiments to study cross-site variation in treatment effects: A hybrid approach with fixed intercepts and a random treatment coefficient. *Journal of Research on Educational Effectiveness*, *10*(4), 817–842. doi: 10.1080/19345747.2016.1264518

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., … Sarna, M. (2018). *The investing in innovation fund: Summary of 67 evaluations* (No. NCEE 2018-4013). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Brady, S. T., Cohen, G. L., Jarvis, S. N., & Walton, G. M. (2020). A brief social-belonging intervention in college improves adult outcomes for black Americans. *Science Advances*, *6*(18), eaay3689. doi: 10.1126/sciadv.aay3689

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*(7), 513–531. doi: 10.1037/0003-066X.32.7.513

Bryan, C. J., Tipton, E., & Yeager, D. S. (in press). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*.

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, *5*(2), eaau4734. doi: 10.1126/sciadv.aau4734

Carroll, J. M., Muller, C., Grodsky, E., & Warren, J. R. (2017). Tracking health inequalities from high school to midlife. *Social Forces*, *96*(2), 591–628. doi: 10.1093/sf/sox065

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632. doi: 10.1257/aer.104.9.2593

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671.

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science*. doi: 10.1177/1745691618804166

Easton, J. Q., Johnson, E., & Sartain, L. (2017). *The predictive power of ninth-grade GPA*. Chicago, IL: University of Chicago Consortium on School Research. Retrieved from Chicago, IL: University of Chicago Consortium on School Research website: https://consortium.uchicago.edu/sites/default/files/publications/Predictive%20Power%20of%20Ninth-Grade-Sept%202017-Consortium.pdf

Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, 460–465. doi: 10.1511/2014.111.460

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting, and Knowing* (pp. 67–82). Hillsdale, NJ: Lawrence Erlbaum.

Gopalan, M., & Tipton, E. (2018). Is the National Study of Learning Mindsets nationally-representative? *Https://Psyarxiv.Com/Dvmr7/*. Retrieved from https://psyarxiv.com/dvmr7/

Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. *Bayesian Analysis*. doi: 10.1214/19-BA1195

Haimovitz, K., & Dweck, C. S. (2017). The origins of children's growth and fixed mindsets: New research and a new proposal. *Child Development*, *88*(6), 1849–1859. doi: 10.1111/cdev.12955

Harackiewicz, J. M., & Priniski, S. J. (2018). Improving student outcomes in higher education: The science of targeted intervention. *Annual Review of Psychology*, *69*, 409–435. doi: 10.1146/annurev-psych-122216-011725

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, 33–46. doi: 10.2307/749455

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177. doi: 10.1111/j.1750-8606.2008.00061.x

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241–253. doi: 10.3102/0013189X20912798

Lazarus, R. S. (1993). From psychological stress to the emotions: A history of changing outlooks. *Annual Review of Psychology*, *44*(1), 1–22. doi: 10.1146/annurev.ps.44.020193.000245

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265. doi: 10.1126/science.1261375

Lewin, K. (1952). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). London, England: Tavistock. Retrieved from http://trove.nla.gov.au/version/21157377

McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. *The American Statistician*, *73*(sup1), 99–105. doi: 10.1080/00031305.2018.1505655

Miller, D. I. (2019). When Do Growth Mindset Interventions Work? *Trends in Cognitive Sciences*, *23*(11), 910–912. doi: 10.1016/j.tics.2019.08.005

Muenks, K., Canning, E. A., LaCosse, J., Green, D. J., Zirkel, S., Garcia, J. A., & Murphy, M. C. (2020). Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset beliefs predict their psychological vulnerability, engagement, and performance in class. *Journal of Experimental Psychology: General*. doi: 10.1037/xge0000763

OECD. (2021). *Sky's the limit: Growth mindset, students, and schools in PISA*. Paris: PISA, OECD Publishing. Retrieved from PISA, OECD Publishing website: https://www.oecd.org/pisa/growth-mindset.pdf

Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher*, *42*(5), 259–265. doi: 10.3102/0013189X13481382

Rattan, A., Good, C., & Dweck, C. S. (2012). “It’s ok — not everyone can be good at math”: Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology*, *48*(3), 731–737. doi: 10.1016/j.jesp.2011.12.012

Tipton, E., Yeager, D. S., Iachan, R., & Schneider, B. (2019). Designing probability samples to study treatment effect heterogeneity. In P. J. Lavrakas (Ed.), *Experimental Methods in Survey Research: Techniques That Combine Random Sampling with Random Assignment* (pp. 435–456). New York, NY: Wiley. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119083771.ch22

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, *125*(5), 617–655. doi: 10.1037/rev0000115

Walton, G. M., & Yeager, D. S. (2020). Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science*, *29*(3), 219–226. doi: 10.1177/0963721420904453

Woody, S., Carvalho, C. M., & Murray, J. S. (2020). Model interpretation through lower-dimensional posterior summarization. *ArXiv:1905.07103 [Stat]*. Retrieved from http://arxiv.org/abs/1905.07103

Yamamoto, T., & Yeager, D. S. (2019). Causal mediation and effect modification: A unified framework. *Working Paper, MIT.*

Yeager, D. S. (2019). *The National Study of Learning Mindsets, [United States], 2015-2016*. Inter-university Consortium for Political and Social Research [distributor]. doi: 10.3886/ICPSR37353.v1

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., … Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369. doi: 10.1038/s41586-019-1466-y

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., … Dweck, C. S. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, *108*(3), 374–391. doi: 10.1037/edu0000098

Zhu, P., Garcia, I., & Alonzo, E. (2019). *An independent evaluation of growth mindset intervention*. New York, NY: MDRC. Retrieved from MDRC website: https://files.eric.ed.gov/fulltext/ED594493.pdf

[i] The mindset + supportive context, or "affordances," hypothesis is akin to what Bailey and colleagues (2020) call the "sustaining environments" hypothesis, which is the idea that intervention effects will fade out when people enter post-intervention environments that lack adequate resources for an intervention to continue paying dividends.
[ii] Lazarus (1993) summarized well the field's pejorative view of treatment effect heterogeneity: "psychology has long been ambivalent … opting for the view that its scientific task is to note invariances and develop general laws. Variations around such laws are apt to be considered errors of measurement" (pg. 3).
[iii] An exploratory analysis allowed the intercept and slope to vary randomly. It showed the same sign and significance of results and supported the same conclusions as the pre-registered fixed intercept, random slope model.