Psychological Bulletin

ISSN: 0033-2909

https://doi.org/10.1037/bul0000352

Do Growth Mindset Interventions Impact Students' Academic Achievement? A Systematic Review and Meta-Analysis With Recommendations for Best Practices

Brooke N. Macnamara[1] and Alexander P. Burgoyne[2]
[1] Department of Psychological Sciences, Case Western Reserve University
[2] School of Psychology, Georgia Institute of Technology

According to mindset theory, students who believe their personal characteristics can change—that is, those who hold a growth mindset—will achieve more than students who believe their characteristics are fixed. Proponents of the theory have developed interventions to influence students' mindsets, claiming that these interventions lead to large gains in academic achievement. Despite their popularity, the evidence for growth mindset intervention benefits has not been systematically evaluated considering both the quantity and quality of the evidence. Here, we provide such a review by (a) evaluating empirical studies' adherence to a set of best practices essential for drawing causal conclusions and (b) conducting three meta-analyses. When examining all studies (63 studies, $N = 97{,}672$), we found major shortcomings in study design, analysis, and reporting, and suggestions of researcher and publication bias: Authors with a financial incentive to report positive findings published significantly larger effects than authors without this incentive. Across all studies, we observed a small overall effect: $\bar{d}$ = 0.05, 95% CI = [0.02, 0.09], which was nonsignificant after correcting for potential publication bias. No theoretically meaningful moderators were significant. When examining only studies demonstrating the intervention influenced students' mindsets as intended (13 studies, $N = 18{,}355$), the effect was nonsignificant: $\bar{d}$ = 0.04, 95% CI = [−0.01, 0.10]. When examining the highest-quality evidence (6 studies, $N = 13{,}571$), the effect was nonsignificant: $\bar{d}$ = 0.02, 95% CI = [−0.06, 0.10]. We conclude that apparent effects of growth mindset interventions on academic achievement are likely attributable to inadequate study design, reporting flaws, and bias.

Public Significance Statement
This systematic review and meta-analysis suggest that, despite the popularity of growth mindset interventions in schools, positive results are rare and possibly spurious due to inadequately designed interventions, reporting flaws, and bias.

Keywords: mindset, implicit theories, educational interventions, best practices, academic achievement

Supplemental materials: https://doi.org/10.1037/bul0000352.supp

Do you believe that your intelligence is relatively stable, or do you believe that you can grow your intelligence? According to *mindset theory* (Dweck, 2006, 2016; i.e., implicit theories: Dweck, 2000; Dweck et al., 1995), these differing beliefs form the core of people's meaning systems (Dweck & Yeager, 2019b). As such, your mindset "profoundly affects the way you lead your life" (Dweck, 2016, p. 6). Indeed, people's mindsets are said to create "entire psychological worlds" (Dweck, 2009, p. 4; see also Dweck, 2007b, 2008a; Yeager & Dweck, 2012) that operate under different motivational and behavioral "rules" (Dweck, 2007a, 2009).

Holding a *fixed mindset* (i.e., entity theory) means believing intelligence or other characteristics are relatively stable. Proponents of mindset theory claim holding a fixed mindset is detrimental for a variety of real-world outcomes because people with fixed mindsets (a) seek to appear smart/talented at all costs, (b) avoid effort, and (c) refrain from challenges and conceal weaknesses (Dweck, 2007a, 2009).

Brooke N. Macnamara https://orcid.org/0000-0003-1056-4996
Alexander P. Burgoyne https://orcid.org/0000-0002-2651-3782

The authors thank Elizabeth McNeilly, James Fox, and Owen Lockwood for help with the search; Essence Leslie for help with coding; Essence Leslie, Victoria Sisk, Amanda Merner, Huangqi Jiang, and Kyle LaFollette for comments on an earlier draft; Deborah Kashy, Andrew R. A. Conway, and Christopher Nye for help with rating how understandable analyses were to readers (see Supplemental Materials); and Jaylene Vazquez, Shivali Vyas, and Elise Rolston for help evaluating whether authors were proponents, neutral, or skeptics of mindset theory.

This project is not funded. The authors have no conflicts of interest to declare. The meta-analysis was preregistered at https://osf.io/ga9jk. Data are publicly available at https://osf.io/ajhxv/.

Correspondence concerning this article should be addressed to Brooke N. Macnamara, Department of Psychological Sciences, Case Western Reserve University, 11220 Bellflower Road, Cleveland, OH 44106, United States. Email: bnm24@case.edu

In other words, people with fixed mindsets have the "one consuming goal of proving themselves" (Dweck, 2016, p. 6), and therefore avoid challenges (Dweck, 2016) and are "devastated by setbacks" (Dweck, 2008a, p. 1). In short, it has been claimed that the fixed mindset "world" is one of "threats and defenses" (Yeager & Dweck, 2012, p. 304).

In contrast, holding a *growth mindset* (i.e., incremental theory) means believing intelligence or other characteristics are malleable. Proponents of mindset theory claim holding a growth mindset is beneficial for a variety of real-world outcomes because people with growth mindsets (a) focus on learning, (b) believe effort is key, and (c) embrace challenges and mistakes (Dweck, 2007a, 2009). In other words, people with growth mindsets have a desire to learn, and therefore seek challenges and are resilient to setbacks (Dweck, 1986, 2006, 2009, 2016). In short, proponents suggest the growth mindset "world" is one of "opportunities to improve" (Yeager & Dweck, 2012, p. 304).

Mindset theorists have claimed growth mindsets lead to positive outcomes in domains ranging from weight loss (Burnette & Finkel, 2012) to business success (Dweck, 2006, 2016) to achieving peace in the Middle East (Dweck, 2012; Goldenberg et al., 2018; Halperin et al., 2011). In particular, mindset theory has been influential in the educational sphere, where its impact has been described as a "revolution that is reshaping education" (Boaler, 2013, p. 143). Mindset proponents encourage parents and teachers to promote growth mindsets in students because, "what students believe about their brains—whether they see their intelligence as something that's fixed or something that can grow and change—has profound effects on their motivation, learning, and school achievement" (Dweck, 2008a, p. 1). The promise of profound effects on learning and achievement led researchers to develop growth mindset interventions—treatments designed to teach students to have more of a growth mindset.

Growth mindset interventions have been popularized through multiple avenues. For example, Dweck (2006, 2016) wrote a popular press book highlighting growth mindset interventions. She and other mindset researchers also founded a for-profit company—Mindset Works—that sells growth mindset interventions to parents and schools. Further, mindset researchers have called for policymakers to advocate for implementing growth mindset interventions in schools and for growth mindset research to be a national funding priority (Rattan et al., 2015).

Millions of dollars in funding from private foundations (e.g., Raikes Foundation, Gates Foundation) and government agencies (e.g., National Science Foundation, U.S. Department of Education) have been awarded to researchers, nonprofit organizations, and for-profit companies for growth mindset intervention studies. As an example, the U.S. Department of Education's Institute of Education Sciences recently awarded Mindset Works a $3.5 million grant. The goal of this grant was to determine whether "Brainology"—Mindset Works' flagship growth mindset intervention product—is effective or not.[1]

For context, Mindset Works has been selling Brainology to schools for thousands of dollars for the past decade claiming that it benefits students.[2] This conflicting information raises the question of whether (a) Brainology is beneficial, as Mindset Works claims on its website, or (b) there was not enough evidence to make this claim, hence why the grant from Institute of Education Sciences was needed.

These conflicting claims by Mindset Works—that Brainology benefits students and that funding is needed to determine whether Brainology is effective—raise two larger questions. Do growth mindset interventions generally improve students' academic achievement? And more deeply, if there is such a benefit, is it through the assumed underlying mechanism—growth mindset?

Do Growth Mindset Interventions Improve Students' Academic Achievement?

The most highly cited growth mindset intervention article (cited over 4,000 times; Google Scholar, July 23rd, 2021) was published by Blackwell, Trzesniewski, and Dweck in 2007. In their intervention, Blackwell and colleagues assigned classes of students to receive either a growth mindset intervention (48 seventh-grade students) or control sessions (43 seventh-grade students). The grades of students who received the growth mindset intervention did not increase following the intervention; rather, the grades of students in the control sessions became worse following the intervention. The lack of decline in grades for the treatment group, along with students' grades from the prior school year suggesting a downward trend, was interpreted as evidence that the intervention successfully prevented further decline of grades, and therefore improved students' academic achievement. Largely based on Blackwell et al.'s (2007) results, more and more researchers, teachers, and entrepreneurs began implementing growth mindset interventions in educational contexts.

In 2018, Sisk and colleagues conducted the first meta-analysis examining the effects of growth mindset interventions on students' academic achievement. Across 38 independent samples ($N = 57{,}155$), they found that the studies' results were mixed. A handful of studies demonstrated that the intervention improved academic achievement relative to control. A sample from one study demonstrated that students who completed the intervention experienced *worse* academic outcomes relative to students in the control condition (Dommett et al., 2013). The vast majority of samples demonstrated a nonsignificant difference in academic achievement between students who received the intervention and the control group.

Overall, Sisk et al. (2018) found a small meta-analytic standardized mean difference between intervention and control groups' academic achievement, favoring the interventions: $\bar{d} = 0.08$, 95% CI [0.02, 0.14]. Despite the positive effect, Sisk and colleagues took a tentative approach to interpreting the results. They cautioned that claims that growth mindset interventions "lead to large gains in student achievement" and have "striking effects on educational achievement" (Yeager & Walton, 2011, pp. 267 and 268, respectively) are likely unwarranted given the small overall effect.

A second reason Sisk et al. (2018) cautioned against interpreting the results as strong evidence for intervention effectiveness was based on their examination of manipulation checks—pre- to post-intervention changes in mindset scores in the treatment groups. They found that many studies did not conduct (or report) a manipulation check. The effect on academic achievement was significant for

[1] https://ies.ed.gov/funding/grantsearch/details.asp?ID=1,728 (retrieved September 23, 2020).

[2] "Who benefits ... • Both high and low-achieving students • Students in a full range of educational settings" https://www.mindsetworks.com/programs/brainology-for-schools (retrieved September 19, 2020).

studies that *did not* report manipulation checks but nonsignificant for studies that did report manipulation checks. Perhaps of greatest concern, among the studies that reported manipulation checks, the meta-analytic effect on academic achievement was significant for studies where the manipulation checks *failed* but nonsignificant for studies where the manipulation checks succeeded.

Based on the observation that significant effects only emerged when manipulation checks were not included or when manipulation checks failed, Sisk et al. (2018) suggested that factors *other than* growth mindset may be the source of academic achievement differences. If so, growth mindset might not be the mechanism underlying effects of "growth mindset" interventions.

Is Growth Mindset the Key Mechanism of Growth Mindset Interventions?

The claim that growth mindset interventions are important for academic achievement is explained as follows: (1) Interventions can teach students to have more of a growth mindset. (2) A growth mindset leads students to adopt learning goals, to put forth effort to pursue challenges, and to be resilient following failure. (3) These traits and behaviors originating from a growth mindset lead to higher levels of achievement. (4) Therefore, teaching students to have a growth mindset will lead to higher levels of academic achievement. In the following subsections, we ask whether each part of this claim is substantiated by evidence.

Can Interventions Teach Students to Have More of a Growth Mindset?

Sisk et al. (2018) found that about $^1/_3$ of growth mindset intervention studies did not report whether the growth mindset intervention influenced students to have more of a growth mindset. Of those that did, around half failed to demonstrate the intervention changed students' mindsets as intended. This result suggests that teaching students to have more of a growth mindset may be difficult to accomplish.

Further, studies that appeared to demonstrate that the intervention changed students' mindsets may have been influenced by demand characteristics. As Burgoyne et al. (2018) note, the wording of growth mindset measures administered following an intervention can closely match the wording of materials used in the growth mindset intervention. If students can guess the premise of the intervention, then they may respond in ways favorable to the research objective on the postintervention measure (see Orne, 1962). This problem might be especially likely in educational environments where students are frequently tested and wish to be graded favorably.

An alternative interpretation is that measures of mindset do not accurately reflect students' mindsets, and therefore failed manipulation checks might reflect measurement issues. The Implicit Theories of Intelligence Scale (Dweck, 2000) is the most frequently used measure of mindset (Limeri et al., 2020). It was originally developed for use with primary-school children but was subsequently used with students of all ages without rigorous empirical validation (Limeri et al., 2020). Though the measure demonstrates strong interitem reliability (see, e.g., Burgoyne & Macnamara, 2021; Dai & Cromley, 2014; Flanigan et al., 2017), recent research has called into question the measure's response process validity—that is, whether students engage in a common process to respond to the items (Limeri et al., 2020).

In particular, Limeri et al. (2020) found that undergraduate students interpret the term "intelligence," which is found in every item of the Theories of Intelligence Scale, in different ways and that their interpretation of this term corresponds to their responses on the scale. Students who interpreted "intelligence" to mean *knowledge* (around $^1/_3$ of students in the sample) agreed with growth mindset items and disagreed with fixed mindset items—presumably because it is logically obvious that one's knowledge can increase (Limeri et al., 2020). Students who interpreted "intelligence" as an *ability* (around ½ of students in the sample), such as the ability to learn, problem solve, or think critically, used the whole range of the scale, with some students agreeing more to fixed items and others agreeing more to growth items. Many students thought of intelligence as multifaceted and indicated that it was difficult to respond to the Theories of Intelligence Scale without knowing what definition of intelligence they were supposed to use.

In addition to variability in how students interpret the word "intelligence," students' response processes may differ because of the mindset intervention. For example, a student might initially interpret "intelligence" as knowledge; following the intervention, she might interpret "intelligence" as ability to learn. Functionally, she would be responding as though the item changed. Likewise, a student might endorse a statement that one's intelligence can grow; following the intervention he might realize his initial belief was flawed and that he did not agree with the statement as much as he thought he did and adjust his response. If measures of mindset do not accurately or reliably reflect a person's beliefs about the malleability of intelligence, then it is difficult to determine whether an intervention influenced students' mindsets.

Does Holding a Growth Mindset Lead Students to Adopt Learning Goals, Increase Effort to Pursue Challenges, and Be Resilient Following Failure?

According to mindset theory, the primary goal for people with growth mindsets is to learn, whereas the primary goal for people with fixed mindsets is to appear talented (Dweck, 2009). Mindset theory presumes downstream effects from these learning versus performance goals: People with growth mindsets are hypothesized to put forth more effort, pursue challenges, and be resilient following failure (e.g., Dweck, 2008a; Rattan et al., 2015; Yeager & Dweck, 2012). In contrast, people with fixed mindsets are hypothesized to refrain from putting forth effort, avoid challenges, and be devastated by failure (e.g., Dweck, 2008a; Rattan et al., 2015; Yeager & Dweck, 2012).

Yet, the overall empirical evidence does not support this claim. For example, despite the assumed link between mindset and goal orientation, Payne et al. (2007) conducted a meta-analysis of the relationship and concluded: "Contrary to Dweck's (1986) perspective, the effect sizes were very small, providing little evidence for Dweck's (1986) view that implicit theories are the primary underlying antecedent of [goal orientation]" (p. 140). In another meta-analysis, Burnette et al. (2013) independently tested multiple relationships in a hypothesized path model of mindset predicting goal orientations. They found that mindset weakly predicted learning goals ($\bar{r} = .19$) and performance goals ($\bar{r} = -.15$).

In a direct test of mindset theory's underlying premises, we (Burgoyne et al., 2020) found little evidence supporting the theory's claims. Students' mindsets accounted for only 1% of the variance in their proclivity for holding learning goals and 0%–1% in their proclivity for holding performance goals. We found no evidence that mindset had a bearing on one's likelihood to persist when facing challenges. The largest relationship (though still small: $r = -.12$) was between mindset and resilience following failure. Surprisingly, the effect was in the opposite direction, suggesting that holding a growth mindset was detrimental to resilience.

Do Traits and Behaviors Originating From a Growth Mindset Lead to Higher Levels of Achievement?

According to mindset theory, "students with a fixed mindset, no matter how bright, often develop values and habits that stand in the way of developing their abilities and doing well in school. In contrast, students with a growth mindset embrace learning, mistakes, and effort in a way that promotes their achievement" (Dweck, 2008b, p. 56). But if growth mindset is only weakly (at most) related to learning goal orientation, persistence, and resilience, the presumed effects further downstream—such as those on academic achievement—may be negligible.

The available evidence suggests this is the case. For example, though Burnette et al. (2013) found a weak association between growth mindset and learning goals, the association between learning goals and goal achievement was nonsignificant. As another example, Blackwell et al. (2007) hypothesized that the impact of their intervention would be strongest for students who initially had more of a fixed mindset. They reasoned that these students would have the most room to shift toward a growth mindset, and therefore the most room for change in motivation and effort, leading to higher gains in achievement. The evidence failed to support this hypothesis: The effect of the experimental condition × initial mindset interaction on change in achievement was nonsignificant.

If growth mindsets lead to learning goals, increased effort, and challenge seeking in ways that promote achievement (Dweck, 2008b), we should observe that individuals with growth mindsets are more likely to attain higher education levels. The available evidence does not support this premise. Two studies examined the relationship between mindset and the highest level of education attained. In a sample of 163 participants, Macnamara and Rupani (2017) found no significant association between mindset and educational attainment. In a sample of 450 participants, Yan et al. (2014) found a significantly negative association such that having more of a growth mindset was associated with *lower* levels of educational attainment.

In Sisk et al.'s (2018) Study 1, a meta-analysis of associations between students' mindsets and academic achievement across 129 studies (N = 365,915), they found a weak overall relationship: Mindset accounted for 1% of the variance in academic achievement. Though mindset theory would predict that the relationship would be stronger for students facing challenges (e.g., transitioning to a new school), level of challenge was not a significant moderator. Subsequent large-sample studies (ns = 211, 222, 246, 586) have failed to observe a significant relationship between mindset and grades (Li & Bates, 2019, 2020). The null result persisted regardless of whether students were facing difficult transitions or other academic challenges (Li & Bates, 2020). Overall, the relationship between students' mindsets and academic achievement appears to be, at most, weak, and not always in the hypothesized direction.

Does Teaching Students to Have a Growth Mindset Lead to Higher Levels of Academic Achievement?

Despite limited empirical evidence that holding a growth mindset leads to higher academic achievement, growth mindset interventions are widely popular and conducted in classrooms around the world (Moreau et al., 2019). In Sisk et al.'s (2018) Study 2, a meta-analysis of the effect of growth mindset interventions on academic achievement, they found a small effect: $\bar{d} = 0.08$. Several moderator results led them to question whether the effect was due to growth mindset; in particular, the finding that interventions where the manipulation check succeeded had no significant effect on academic achievement.

If not growth mindset, what else might be driving observed effects in growth mindset interventions? One factor that might be driving effects is effort encouragement (Li & Bates, 2019). Students given growth mindset interventions are typically taught that intelligence can change *and* are also encouraged to work harder, while students in control conditions neither receive the mindset treatment nor this extra effort encouragement. Multiple differences between treatment and control protocols make it unclear whether (a) differences between groups are due to changing students' mindsets, (b) mindset training must be augmented with encouragement to work harder to be effective, or (c) differences are simply due to encouragement to work harder. Indeed, working harder should produce higher achievement (Gneezy et al., 2019) regardless of one's beliefs about intelligence.

To disentangle these factors, Li and Bates (2019) separately manipulated fixed mindset and effort encouragement messages. They found that students given a fixed mindset message (which should hinder performance according to mindset theory) *and* effort encouragement performed as well as students given only effort encouragement. These results suggest that at least some of the observed differences between treatment and control groups found in the mindset intervention literature might be driven by encouraging students to work harder.

Effort encouragement is not the only common difference between treatment and control groups in growth mindset interventions. Many growth mindset interventions additionally encourage students in the treatment group—but not the control group—to practice, study, pursue challenges, persevere, find optimal learning strategies, and/or seek help (e.g., Blackwell et al., 2007; Burnette et al., 2019; Paunesku et al., 2015; Yeager et al., 2018). Interventions often additionally teach students in the treatment group, but not the control group, strategies for learning course content and overcoming setbacks (e.g., Boaler et al., 2018; Yeager et al., 2018) and include role models or inspirational stories (e.g., Burnette et al., 2019; Foliano et al., 2019). Finally, growth mindset interventions might additionally help students in the treatment group, but not the control group, by encouraging them to normalize mistakes, set goals, and/or create individualized study plans (e.g., Blackwell et al., 2007; Gauthreaux, 2015; Zonnefeld, 2015).

Very few growth mindset intervention studies have isolated the critical ingredient—teaching that intelligence or another characteristic is malleable—as the only difference between treatment and control groups. One example of such a study was conducted by Polley (2018) who attempted to isolate the mechanism by specifically testing whether the success of growth mindset interventions is due to teaching students about the malleability of intelligence or due to other factors that often covary with the experimental manipulation.

Polley (2018) created a typical growth mindset intervention using materials from mindsetworks.com and mindsetkit.com. This intervention taught students that the brain grows stronger when we learn (i.e., growth mindset). The intervention also encouraged students to practice, challenge themselves, and focus on learning deeply; it taught students that challenge meant learning and that mistakes and failure lead to success; and it taught students to set learning rather than performance goals and that instead of thinking that they can't do something that they should think "I can't do something yet!"

In the active control group, students were given the same encouragement, lessons, tips, and strategies, but received no information that the brain grows stronger with learning. Thus, the treatment and control groups were identical except for the critical ingredient of teaching a growth mindset. After controlling for prior achievement, the treatment and active control groups did not significantly differ on the study's main measure of academic achievement. Like Li and Bates' (2019) set of studies, Polley's (2018) study suggests that some of the effects found in the mindset intervention literature—where growth mindset is often not the only difference between treatment and control groups—may be due to factors other than teaching a growth mindset.

One argument for covarying multiple factors with treatment groups is that encouraging students to put forth more effort and practice, to embrace challenges, and to develop learning strategies is part of an effective growth mindset intervention (Yeager & Dweck, 2020). However, this type of design leaves opens the possibility that encouraging students to put forth more effort and practice, to embrace challenges, and to develop effective learning strategies *without* teaching growth mindset (i.e., that the brain or a characteristic is malleable) would be equally effective (see Polley, 2018). That is, one or more factors *other than* growth mindset may be the critical ingredients in "growth mindset" interventions.

Another problem with encouraging students to put forth more effort and seek challenges while teaching that a characteristic or the brain is malleable is that increased effort and challenge seeking are theorized outcomes of holding a growth mindset. When these factors are combined with the growth mindset intervention, changes in effort and challenge seeking following the intervention cannot be attributed to holding more of a growth mindset—it may be from directly encouraging students to engage in these behaviors. This combination makes it difficult to determine whether growth mindset is the mechanism underlying growth mindset intervention effects or whether growth mindset is unnecessary for influencing motivations and behaviors that impact academic achievement.

The Present Study

The goal of the present study was to answer our main questions: (a) Do growth mindset interventions generally improve students' academic achievement? (b) If there is such a benefit, is it through the assumed underlying mechanism—growth mindset—or are apparent effects due to inadequate study designs, reporting flaws, and/or bias? We focus exclusively on growth mindset treatments aimed at improving students' academic achievement because, in educational settings, this is often the ultimate outcome assumed to occur from holding a growth mindset (see, e.g., Dweck et al., 2014; https://www.mindsetworks.com/Science/Impact [retrieved July 29, 2021]; Rattan et al., 2015; Yeager et al., 2019).

Sisk et al. (2018) previously meta-analyzed the effect of growth mindset interventions on academic achievement. However, the quality of the evidence was not systematically evaluated. Thus, to best answer our two questions, we conducted the first systematic and comprehensive review of the growth mindset intervention on academic achievement literature that examines both the quantity and the quality of the evidence according to a well defined set of best practices.

In the next section, we describe and justify the set of study design and reporting characteristics that are critical for evaluating mechanisms of growth mindset interventions. Following the General Method, we review the state of the growth mindset intervention literature and describe patterns observed across growth mindset intervention studies.

We then present the results of three meta-analyses. The first meta-analysis addresses our first question of whether growth mindset interventions generally improve students' academic achievement. This model used the same approach as Sisk et al.'s (2018): It included all studies we could find that met the inclusion criteria, regardless of study quality or interpretability of the mechanism. Thus, the first meta-analysis provides the estimated effect of growth mindset interventions on academic achievement when quality standards are lenient.

For this meta-analysis, we included the studies from Sisk et al.'s (2018) meta-analytic literature search and updated the search with studies that became available after their search stop date. Though these meta-analyses are only a few years apart, the popularity of growth mindset interventions has continued to increase, resulting in many studies entering the literature in the intervening years. We systematically searched for all growth mindset intervention studies that compared treated students to control students on a measure of academic achievement (student grades or standardized test scores). We focused on treatment-versus-control studies because they provide better evidence for an effect of a treatment than single-group studies.

Additionally, we conducted mixed-effect moderator analyses. We tested theoretically meaningful moderator variables, such as whether the effect size differed depending on students' socioeconomic status or level of challenge. We also tested multiple potential methodological moderators such as the intervention delivery mode and number of sessions. We completed this meta-analysis by conducting publication bias analyses to assess the extent to which such biases may be operating within the growth mindset intervention literature.

The second and third meta-analyses attempt to answer our second question of whether growth mindset is the underlying mechanism of growth mindset interventions. The second model follows up on Sisk et al.'s (2018) moderator analyses of manipulation checks. In this model, we only included studies that demonstrated a significant change in the mindsets of students who received the growth mindset intervention. Thus, the goal of the second meta-analysis is to evaluate treatment effects of studies that provide a minimal standard of evidence that growth mindset is the underlying mechanism. We note, however, that this model rests on the assumption that measures of mindset are valid and reliable, and that other factors are not the key mechanisms driving effects.

We conducted the same mixed-effects moderator analyses as in the first meta-analysis when possible. Relatively few studies demonstrated that the intervention influenced students' mindsets. Therefore, in some cases, not enough effects were available to conduct moderator analyses.

The third model presents the best available evidence—growth mindset intervention studies with the highest-quality study design and evidence that growth mindset could be one of the underlying mechanisms. As Yeager and Dweck (2020) point out, examining the average effect size from a meta-analysis that combines all studies—regardless of quality—is suboptimal. Instead, examining the best available evidence will lead to a better estimate of the true effect.

Our goal for the third meta-analysis was to include studies that demonstrated the intervention influenced students' mindsets and met all best practices criteria. No studies met all best practices criteria. In this case, our plan was to relax the standard for the number of best practices a study met until at least five studies could be included. This process resulted in six studies meeting at least 60% of the best practices criteria. There were not enough studies included in the third meta-analysis to conduct moderator analyses.

Study Characteristics Critical for Accurately Interpreting the Treatment Mechanism

The best practices criteria we evaluate belong to a larger group of study characteristics critical for evaluating evidence. The characteristics we describe apply to all psychological intervention studies, though we focus on their implementation in growth mindset interventions in educational settings. These characteristics are needed to draw clear conclusions from the research, particularly in interpreting the treatment mechanism.

Best Practices in Intervention Design

The Intervention Is Compared to an Active Control Condition

The control condition should be identical to the treatment condition in every way except for the critical ingredient of the treatment (Simons et al., 2016), including matching participant expectations (Boot et al., 2013). Passive comparison groups (no-contact, waitlist, and teaching-as-usual controls) differ from the treatment group in students' and teachers' expectations, attention, and engagement. Differences between treatment and passive control groups could account for perceived treatment effects, rather than growth mindset messages.

A more rigorous methodology compares an intervention to an active control group (sometimes called a treated control or a placebo control). In an active control group, participants engage in similar activities to those in the intervention, minus the critical ingredient (i.e., growth mindset). An active control group is necessary but not sufficient to attribute differences in achievement to the intervention (Simons et al., 2016).

A fixed mindset intervention comparison group is a type of active control. However, any differences between the groups cannot be attributed to the benefit of a growth mindset intervention over a detriment of a fixed mindset intervention, if there is one. Thus, choosing an active (non-fixed-mindset) control group is the best type of comparison group for isolating the critical ingredient of a growth mindset.

Aside From Attribute Malleability, No Other Differences Between the Treatment and Control Group Should Be Introduced

In addition to comparing the treatment to an active control, the two groups' activities, perceptions, and experiences should be identical except for the key manipulation: influencing one group to believe intelligence or another attribute can change—that is, teaching a growth mindset. Unless also applied to the comparison group, growth mindset interventions should not additionally encourage treatment group students to work harder, suggest strategies when facing challenges, help students set goals, or include any other treatment aspect. When the treatment and control groups differ in multiple ways this precludes interpretation of the mechanism driving any effects.

It may be that additional encouragement, strategies, and tips, such as providing concrete actions for students, are necessary for a mindset intervention to impact achievement (Yeager & Dweck, 2020). To determine if the intervention's impact is from a change in growth mindset augmented with encouragement, strategies, and/or tips—rather than only due to the encouragement, strategies, and/or tips—the control group needs to receive the same information as the treatment group except for information about attribute malleability (growth mindset). Without a control protocol that is otherwise identical to the treatment, we cannot determine whether teaching a growth mindset is necessary. Thus, in the presence of multiple differences between treatment and control groups, any effects cannot be clearly attributed to growth mindset.

A Priori Power Analysis

Adequate sample sizes are necessary for appropriate hypothesis testing. A priori power analyses help researchers determine the minimum sample size needed to appropriately test for an effect of a given size. Adequate sample sizes are necessary to have confidence in the precision of sample estimates, as small samples lead to high uncertainty in the results. Thus, researchers should conduct an a priori power analysis to determine the minimum sample size needed to reliably detect the smallest effect size that would be of theoretical and/or practical importance.

A priori power analyses, or sample size justification more broadly, have become increasingly common in the past decade. They have been part of the American Psychological Association's Journal Article Reporting Standards since their inception in 2008 (APA Publications & Communications Board Working Group on Journal Article Reporting Standards, 2008). Reporting a power analysis does not ensure the study is adequately powered. However, reporting an a priori power analysis requires researchers to determine the expected effect size or smallest effect size of interest and plan the study accordingly. If reported and followed, conducting a priori power analyses also protects against questionable research practices such as p-hacking and data peeking.

Random Assignment to Condition at the Individual Level

Along with adequate sample sizes, random assignment to conditions at the individual level helps ensure that the treatment and control groups are comparable. That is, if participants are randomly assigned to condition, there is an equally likely chance that any given individual with all their characteristics will be assigned to either condition. With a large enough sample, on average, the groups should be roughly equivalent regarding extraneous factors (e.g., ability, personality, motivation) that could affect the outcome of the intervention. As the literature in an area grows, if it contains many studies with large-sample sizes where participants were randomly assigned to condition, the average effect of an intervention should not be influenced by such extraneous factors.

Blinding

Blinding participants to condition assignment is important in treatment-control designs where awareness of one's condition assignment (or that there are multiple conditions) could alter beliefs, motivation, or otherwise influence behavior. Without blinding, subject-expectancy effects might occur, especially in cases where participants are aware or intuit that their assigned condition is designed to be beneficial (the treatment condition) or is not designed to be beneficial (the control condition; Boot et al., 2013). Participants not blinded to condition may also behave according to their beliefs about the study administrators' expectations (i.e., demand characteristics). Unfortunately, blinding to condition assignment does not ensure that expectations are equated: The intervention itself can influence students' expectations about improvement. In addition to blinding to condition, researchers should explicitly test for students' expectations for the effectiveness of the intervention.

Blinding study administrators to condition assignment may reduce demand characteristics if administrators do not have informed expectations for participants to intuit. Blinding study administrators to condition assignment may also help reduce observer-expectancy effects—when study administrators consciously or unconsciously influence participants (e.g., Rosenthal, 1976; Rosenthal & Rubin, 1978). Depending on the amount and type of administrator involvement in the intervention, researchers should explicitly test administrator's expectations for the effectiveness of the intervention.

Finally, blinding teachers to students' conditions reduces the chance that teachers' beliefs about condition effectiveness and knowledge of student assignment will influence their behavior toward the student or their evaluation of student performance (e.g., Ainsworth et al., 2015; Rosenthal & Jacobson, 1968). Researchers should explicitly test teachers' expectations about the effectiveness of the intervention.

Growth mindset interventions should be designed and administered to equate participant, administrator, and teacher expectations and to reduce biases. Otherwise, readers cannot evaluate whether differences between the treatment and control groups are due to the treatment or these extraneous variables.

Testing Whether the Intervention Influenced Treatment Students' Mindsets

Assuming construct measures are valid and reliable, manipulation checks are critical for drawing accurate conclusions about the effect of the independent variable on the dependent variable when the independent variable can only be manipulated indirectly (Hoewe, 2017), as is the case with growth mindset interventions. If manipulation checks are not included or the manipulation check fails, there is insufficient evidence to attribute intervention effects to the hypothesized mechanism. Only when the manipulation check succeeds is there evidence that an effect may be attributeable to the hypothesized mechanism.

Researchers must also keep in mind that manipulation check results are only as valid as their measures. If no measures accurately reflect the underlying construct (e.g., due to demand characteristics, variable response processes), then the responsible mechanism cannot be determined until a valid and reliable measure has been developed. If measures of mindset are valid and reliable, researchers should test whether the intervention successfully influenced students' mindsets as intended and interpret treatment results in light of this manipulation check.

Best Practices in Documentation, Analyses, and Reporting

Following Detailed Preregistered Hypotheses, Design, and Analysis Plans

Preregistration is required for all government-funded medical clinical trials in the United States (Food & Drug Administration, 2007; see also Kaplan & Irvin, 2015). Preregistering in psychology is a newer practice—for instance, the badge identifying preregistered studies was introduced by *Psychological Science* in 2013 (Eich, 2014)—and is not mandated.

Preregistrations vary in their thoroughness, but the idea behind preregistration is that researchers record their hypotheses, planned sample, data collection stopping rule, measures, conditions, procedure, and analysis plan before the data have been examined (and ideally before the start of data collection). These public documents are timestamped and uneditable, and can be embargoed (with access given to reviewers) while authors work toward publication.

When a detailed preregistration has been created, the flexibility to engage in certain "questionable research practices" (John et al., 2012, p. 524) using "researcher degrees of freedom" (Simmons et al., 2011, p. 1359) is curtailed. In particular, preregistration may deter the following practices:

1. selectively excluding participants' data based on the results;
2. selectively reporting dependent measures based on the results;
3. selectively reporting, comparing, or combining conditions based on the results;
4. examining the data and deciding to collect more data if the results are not significant; and
5. examining the data and deciding to stop data collection if the results are significant (John et al., 2012).

Some might argue that insisting on preregistrations will suppress discoveries. This need not be the case. Discoveries can continue to be made via exploratory studies or exploratory analyses. To confirm those discoveries and test new hypotheses, planned confirmatory analyses and hypotheses should be registered a priori. Preregistered studies are not necessarily well-designed studies. Nonetheless, more trust can be placed in the veracity of results from studies that adhere to detailed preregistrations than similar studies that do not.

Reporting the Results of the Participants Who Participated

In large-scale randomized control trials where missing data, noncompliance, deviations from protocols, and failure to start cases are likely to occur, researchers must decide whether to conduct *per-protocol analyses*, *intent-to-treat analyses*, or both. *Per-protocol analyses* refer to analyses that only include the participants who received the treatment (or control) as intended. *Intent-to-treat analyses* refer to analyses that include all participants who were assigned to condition regardless of whether they received the treatment or violated protocols.

Per-protocol analyses are sometimes described as "proof of principle" or method effectiveness (Porta et al., 2007; Sheiner, 2002) and are most appropriate when the goal of the study is to evaluate the potential benefit for those who receive the treatment as planned (Hollis & Campbell, 1999). However, per-protocol analyses can introduce systematic bias by removing those who do not comply. Noncompliers might have certain characteristics that differ from compliers that would impact the effect of treatment (Shaya & Gu, 2007) and noncompliance might be higher in the treatment condition than in the control condition, violating the goal of random assignment.

Intent-to-treat analyses are sometimes described as "proof of practice" or use effectiveness (Porta et al., 2007; Sheiner, 2002) and are most appropriate when the goal of the study is to evaluate the estimated benefit of a change in treatment policy (Hollis & Campbell, 1999) because in the real world not all people will adhere to treatment. Though intent-to-treat analyses are often viewed as a solution to the bias that can come with per-protocol analyses, flexibility in intent-to-treat analysis decisions can also bias results. In an analysis of randomized controlled trials reported as using intent-to-treat analyses, Hollis and Campbell (1999) note that there is no standard definition of intent-to-treat and researchers vary widely in their interpretation. For example, researchers differed in whether they included participants who never started the treatment, whether they excluded participants after randomization and under which circumstances, and how they handled missing data on the outcome variable. They conclude "The intention to treat approach is often inadequately described and inadequately applied . . . . Readers should critically assess the validity of reported intention to treat analyses" (p. 670).

Likewise, Porta et al. (2007) conducted a systematic review of randomized two-armed clinical trials that reported both per-protocol and intent-to-treat analyses. They found poor agreement between the two approaches due to the uncertainty that protocol deviations and missing values introduce. This variability may be due to differences in exposure to the treatment mechanism as well as differences in the method used to impute missing outcome values in intent-to-treat analyses. Porta et al. concluded that using a single statistical approach of either per-protocol or intent-to-treat analyses does not protect against bias.

Intent-to-treat analyses should be more conservative than per-protocol analyses because they will include individuals who did not adhere to the treatment regimen, reducing the effect of treatment. We should, therefore, be concerned when failure to comply with treatment *increases* the reported effect of treatment. For example, Outes and colleagues conducted a large-scale growth mindset intervention in Peru. Of the 400 schools intended to administer the intervention, 161 schools did not receive the materials or did not administer the intervention. Per-protocol analyses—analyses that only included the 239 schools that administered the treatment as intended—yielded ds = 0.03 and 0.01 for the two academic achievement measures (see Sisk et al., 2018). Outes et al. (2017) reported intent-to-treat analyses—analyses where all 400 schools were included in the treatment group. Despite the assumption that intent-to-treat analyses should yield more conservative estimates than per-protocol analyses, the reported intent-to-treat effects were ds = 0.11 and 0.08: Nearly four times and eight times the size of the per-protocol results. Thus, *failing* to implement the treatment substantially increased the apparent effect of the treatment.

Outes et al. (2017) then additionally applied an adjustment for schools' noncompliance. Rather than bringing the effect size estimates closer to the effects for schools that complied (ds = 0.03 and 0.01), this adjustment further increased the effect sizes to ds = 0.25 and 0.18, respectively. Thus, an adjustment ostensibly intended to better reflect effects had students received the treatment, yielded effect sizes eight to eighteen times the size of the effects for students who received the treatment. The true explanatory effect of an intervention cannot be determined when researchers have such flexibility in applying estimates and adjustments.

Further, the type of analyses used should depend on the goal of the study. Intent-to-treat analyses should be used when the goal is to determine the pragmatic effects of a policy change rather than an estimate of the potential effect when the treatment is received as planned (Hollis & Campbell, 1999). We argue that the potential benefit of treatment when received as planned, who benefits, and how (by what mechanism), should be more firmly established in the growth mindset intervention literature before investigating the pragmatic effects of policy change.

That said, in the presence of protocol deviations, conclusions regarding the effect of treatment cannot rest on either per-protocol or intent-to-treat analyses alone (Porta et al., 2007). Thus, in cases of large-scale studies where missing data, noncompliance, deviations from protocols, and failures to start are likely to occur, both per-protocol and intent-to-treat analytical results should be reported.

Reporting the Results of All Subsamples

Like any type of underreporting, selectively excluding the results of subsamples provides an incomplete picture of the treatment effects. For example, Broda et al. (2018) administered a growth mindset intervention to incoming students at a large university and grouped students into one of six subsamples (African American, Asian, International, Latino/a, Multiracial, and White), but only reported the results of *three* of these six subsamples, along with the results of the "full sample." The "full sample" only included five of the six subgroups, without an explanation for why one group (approximately 15% of the whole sample) was excluded from all analyses. Readers cannot evaluate the full set of results when subgroups are excluded.

As another example, in a recent large-scale growth mindset intervention experiment, Yeager et al. (2019) excluded 50% of the participants and only reported the results of students who were "relatively lower achieving" (p. 366). If one or more subgroups are important enough to be separated instead of reported as part of the whole sample, all subgroups' results should be reported for comparison.

Best Practices in Avoiding Financial Conflicts of Interest

Many mindset researchers have a financial interest in demonstrating the benefits of growth mindsets. To be clear, a researcher can become an expert on a topic and have income tied to their expertise (e.g., raises at an institution, honoraria for scientific talks) without having a perceived conflict of interest. A perceived conflict of interest is introduced when one's income is tied to promoting a particular outcome, in this case, positive effects of growth mindset. Financial incentives to report particular outcomes may influence study design, analyses, how findings are interpreted, and which results are reported and published (Roseman et al., 2011).

Some mindset researchers profit from selling self-help books that tout the benefits of growth mindsets. With millions of books in print, author profits likely exceed $1 million (Peterson, 2019). When writing and promoting a growth mindset self-help book with promises to explain "how we can learn to fulfill our potential" (cover, Dweck, 2006, 2016) or build "confidence, courage, and grit" (cover, Coates, 2020), there is a financial incentive to describe evidence in favor of these claims and omit contradictory evidence.

The financial incentive to describe evidence aligned with promoted content continues after writing the book. That is, people may be more likely to buy a book on mindset after hearing about a research study describing growth mindset's benefits and less likely to buy the book if they hear about conflicting evidence. Thus, book authors subsequently conducting research on growth mindsets may be, perhaps without awareness, more likely to look for evidence supporting their book's claims (confirmation bias) and less likely to publish contradictory findings (contributing to publication bias).

Multiple growth mindset researchers are registered with speakers' bureaus as motivational, inspirational, and keynote speakers, where they charge $10,000–$50,000 per talk to speak about growth mindset.[3] This income is not often acknowledged in scientific publications (Chivers, 2019). Although psychology has no clear financial conflict of interest standards about this type of income, the medical field would consider not disclosing this revenue stream in publications a violation of their ethical guidelines: Readers should be alerted to financial interests they might perceive as potentially influencing findings (Chivers, 2019). That is, presumably, companies hire mindset researchers as motivational speakers to hear about the benefits of growth mindsets; knowledge of weak, mixed, null, or counter-evidence to the benefits of growth mindset reduces that appeal, which would reduce the number of invitations for these lucrative speaking engagements.

Additionally, several growth mindset researchers have cofounded, are employed by, or serve as paid consultants to organizations that sell or promote growth mindset interventions or services. In the case of for-profit companies, more growth mindset products sold/consulting contracts gained leads to more profit, and greater potential financial compensation for cofounders, executives, and employees. For nonprofit organizations, more clients seeking growth mindset services leads to a stronger organization and greater potential financial compensation for executive directors, employees, and consultants. Mindset researchers earning income from either type of organization are incentivized to report positive effects of growth mindset in their research.

These biases may be unintentional. For instance, when a growth mindset intervention has a null effect, growth mindset proponents with a financial incentive may be likely to question the timing, the training of the administers, contextual factors, or other nuances and assume one of these issues suppressed the result. Critically, if a positive effect is found, this same level of criticism may not be applied. As Simons et al. (2016) state:

> Although researchers no doubt view their own work as objective and untainted by corporate influences, evidence from fields like medicine raise doubts about the ability to remain neutral when financial incentives are aligned with one outcome (e.g., see Bekelman et al., 2003; Garg et al., 2005; Perlis et al., 2005). (p. 168)

Best Practices That Were Inclusion Criteria

Here, we mention best practices in intervention design and in documentation, analyses, and reporting that were part of our inclusion criteria. As such, all studies included in the subsequent systematic review and meta-analyses met these criteria and we do not count them when examining the number of best practices criteria studies met. We introduce these study and report characteristics here to explain their importance for accurately interpreting the treatment mechanism when evaluating growth mindset intervention studies.

The Intervention of Interest Is Isolated From Other Treatments

To test the efficacy of a particular intervention, there must be a condition in which participants only receive the intervention of interest for comparison with a control group. If an intervention is combined with another treatment, either by design or during analysis, readers cannot evaluate whether the intervention is effective or whether effects are due to the other treatment or their combination. For example, Tillis (2019) randomly assigned students to a treatment group or a control group. The treatment group students received a growth mindset intervention, a stereotype threat intervention, and a relevance and cognitive dissonance intervention. By designing the study to combine multiple interventions, the effects of the growth mindset intervention cannot be isolated experimentally. We only included studies where the growth mindset intervention was administered independent of other intervention protocols.

It can be similarly problematic when researchers design a study where the effect of the growth mindset intervention can be isolated, but then combine multiple interventions' effects when reporting the results. For example, Paunesku et al. (2015) conducted a study where students were randomly assigned to a control condition, a growth mindset intervention, a sense-of-purpose intervention, or a growth-mindset-and-sense-of-purpose combined intervention. This design is excellent because it allows the researchers to determine (a) the effect of the growth mindset intervention, (b) the effect of the sense-of-purpose intervention, and (c) the effect of the combined growth mindset and sense-of-purpose intervention relative to the control.

[3] See for example, https://www.aaespeakers.com/speakers/carol-dweck, https://www.speakerbookingagency.com/talent/carol-dweck. See also for example, https://www.jla.co.uk/conference-speakers/carol-dweck, https://www.celebrityspeakersbureau.com/talent/carol-dweck/, https://www.nopactalent.com/speaker/carol-dweck.php (all preceding retrieved August, 22nd, 2020). See https://www.allamericanspeakers.com/celebritytalentbios/Lisa+Blackwell,+PhD/399102 (retrieved September 16th, 2020). See also https://www.thelavinagency.com/speakers/david-yeager (retrieved March 28th, 2021).

However, Paunesku and colleagues combined all three intervention groups into one (the mindset intervention, the sense-of-purpose intervention, and the combined mindset and sense-of-purpose intervention) and compared this single combined intervention group to the control group for one of their two outcomes. Combining all three intervention conditions to compare with the control group obfuscates the effects of the individual treatment conditions as well as the effects of the combined mindset and sense-of-purpose condition. We only included studies where the effect of the growth mindset intervention could be evaluated independently of other interventions' effects.

Measuring Direct Outcomes Rather Than Indirect Outcomes

If the goal is to measure the effect of a treatment on an outcome, researchers should measure the outcome of interest as directly as possible. In the case of growth mindset interventions in education, much of the time, the outcome of interest is students' academic achievement. There are several potential direct measures of academic achievement, such as grades and standardized tests scores. All studies included in our systematic review and meta-analyses examined the effect of growth mindset interventions on grades (exam grades, course grades, grade averages) and standardized test scores.

Though the measurement of indirect effects is useful for testing mediation, conclusions about an unmeasured outcome should not be drawn from indirect effects. In the case of mindset, the effect of growth mindset interventions on academic achievement is assumed to occur via motivational processes, such as motivation to learn and seek challenges. However, given the weak links between mindset and motivational processes (e.g., Burgoyne et al., 2020; Burnette et al., 2013; Payne et al., 2007) and between these motivational processes and achievement (e.g., Burnette et al., 2013), one cannot infer the effect of mindset on academic achievement based on mindset's relationship to a motivational process. For example, if a study were to test the impact of a growth mindset intervention on motivation to seek challenges, this outcome measure should not be used to draw conclusions about mindset interventions' effects on academic achievement—academic achievement should be measured directly if attempting to draw conclusions about it.

As another example, if the outcome of interest is on-time college graduation, on-time college graduation should be the measured outcome. Measuring full-time status as a freshman, which in turn predicts on-time college graduation, should not be the outcome variable used to draw conclusions about growth mindset intervention effects on on-time graduation (see Yeager, Walton, et al., 2016). Examining whether a growth mindset intervention predicts a predictor of academic achievement does not provide strong evidence that the intervention directly predicts that outcome.

Using Continuous Variables Over Artificially Dichotomized Variables

Artificially dichotomizing continuous measures is rarely conceptually or statistically defensible, and frequently yields misleading results (MacCallum et al., 2002). Although artificial dichotomization is most often performed on the predictor variable in correlational psychological research (e.g., a median split), some researchers will artificially dichotomize the outcome variable (MacCallum et al., 2002). An example is artificially dichotomizing grade point averages (GPA) into satisfactory grades (As, Bs, and Cs) and unsatisfactory grades (Ds and Fs) when examining the effect of a mindset intervention. This practice reduces the information available to answer the question of what effect an intervention has on achievement. For instance, it reduces variance such that a C− grade is treated the same as an A+. No studies included in our systematic review and meta-analyses artificially dichotomized continuous measures.

Interpretation Considerations

The following characteristics are critical for interpreting treatment mechanisms accurately, but we do not code them as best practices criteria. We explain why in each case. We introduce these study and report characteristics here to explain their importance for evaluating growth mindset intervention studies.

Equivalent Control Group at Baseline

In addition to using an appropriate control condition, the control group should be comparable to the experimental group on key outcome variables before treatment (in the present case, academic achievement). Otherwise, the results may be difficult to interpret (Redick & Webster, 2014). Statistically controlling for preintervention differences does not necessarily solve this problem. For example, suppose the treatment group's average baseline score was higher than the control group's average baseline score, but neither group showed any improvement from pre- to postintervention. An ANCOVA controlling for preintervention differences would indicate a benefit of the intervention for the treatment group despite no improvement from the intervention (i.e., Lord's paradox, Lord, 1967; see also Wright, 2006).

Additionally, baseline scores or factors associated with baseline scores could interact with the effect of the intervention—for example, some mindset proponents have suggested that growth mindset interventions may be more effective for lower-achieving students than higher-achieving students (e.g., Yeager et al., 2019). If the treatment and control groups differ on the outcome variable measure at baseline, the effect of the intervention may be ambiguous.

An equivalent control group at baseline is critical for interpreting the treatment mechanism. However, we do not consider having an equivalent control group at baseline a best practice because it may not be a failure of the study design. That is, despite random assignment to condition, the treatment and control groups could still differ on preintervention academic achievement due to chance.

Controlling for the Familywise Error Rate

If multiple tests are conducted for the same hypothesis, the familywise error rate should be adjusted. Otherwise, the chance of erroneously observing a significant effect increases beyond the α level. For example, suppose a growth mindset intervention measured academic achievement at three time points for each of seven classes, any of which could support the hypothesis that the growth mindset intervention impacted achievement. With 21 significance tests and an α level of .05, we should expect that one outcome will be significant by chance even if the null hypothesis is true because the typical α level (.05) permits a false positive rate of 5%. If only one significance test is conducted per hypothesis, then there is only a small chance that it will reveal a significant effect when none truly exists. Thus, the α level for multiple tests of the same hypothesis should be controlled (e.g., applying Bonferroni's correction).

Though controlling for the familywise error rate is a best practice in reporting, we do not code it as such because it does not impact the results of meta-analyses, as meta-analyses evaluate the size of the effect rather than its significance. That said, as we systematically reviewed the literature, we found multiple instances of practices that inflate the Type I error rate as well as inappropriate interpretations of significance levels. Readers should consider whether the familywise error rates need to be adjusted when evaluating growth mindset intervention studies.

Testing for Differences Between Groups When Claiming Group Differences

When authors make claims about the importance of the treatment for a given subgroup compared with other subgroups, they should test for those claimed differences. For example, Broda et al. (2018) stated their mindset intervention was "designed to impact underrepresented student groups" (p. 322). Yet, they do not report a test of whether the treatment impacted students differently depending on their identified group membership.

As another example, Yeager et al. (2019) only reported the results of the 50% of students performing below their school's median. They stated this was in part because lower-achieving students may have more academic difficulties and therefore may benefit more from a growth mindset intervention than higher-achieving students. Yet, they did not report a test of whether the treatment effect for lower-achieving students differed from the treatment effect for higher-achieving students.

Testing for group differences when claiming or implying group differences is a best practice in analyses. As with controlling for the familywise error rate, we do not code this characteristic as a best practice criterion here because it does not impact the results of meta-analyses. That said, when evaluating growth mindset intervention studies, readers should consider whether claims of subgroup differences are warranted by proper statistical tests.

General Method

Transparency and Openness

Hypotheses, methods, and planned analyses were preregistered at https://osf.io/ga9jk. Deviations and decisions not explicit in the preregistration are reported in the Appendix. We designed the meta-analyses and report the results in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009) and the American Psychological Association's Quantitative Meta-Analysis Article Reporting Standards (Appelbaum et al., 2018). The meta-analytic data are openly available at https://osf.io/ajhxv/.

Inclusion Criteria and Literature Search

The criteria for including a study were as follows:

- A growth mindset treatment not combined with any other intervention, was administered directly to students, where the primary goal was to increase students' belief that one or more human characteristics (e.g., intelligence, personality) are malleable.
- A relevant comparison group (active, passive, or fixed-mindset condition), henceforth control, was included.
- A measure of academic achievement—course exam grade (e.g., midterm exam), single course grade, GPA, or standardized test performance—was reported.
- An effect size reflecting the difference between the growth mindset intervention group and the control group on one or more measures of academic achievement after the intervention was reported, or enough information was provided to compute this effect size.
- The methods and results were in English.

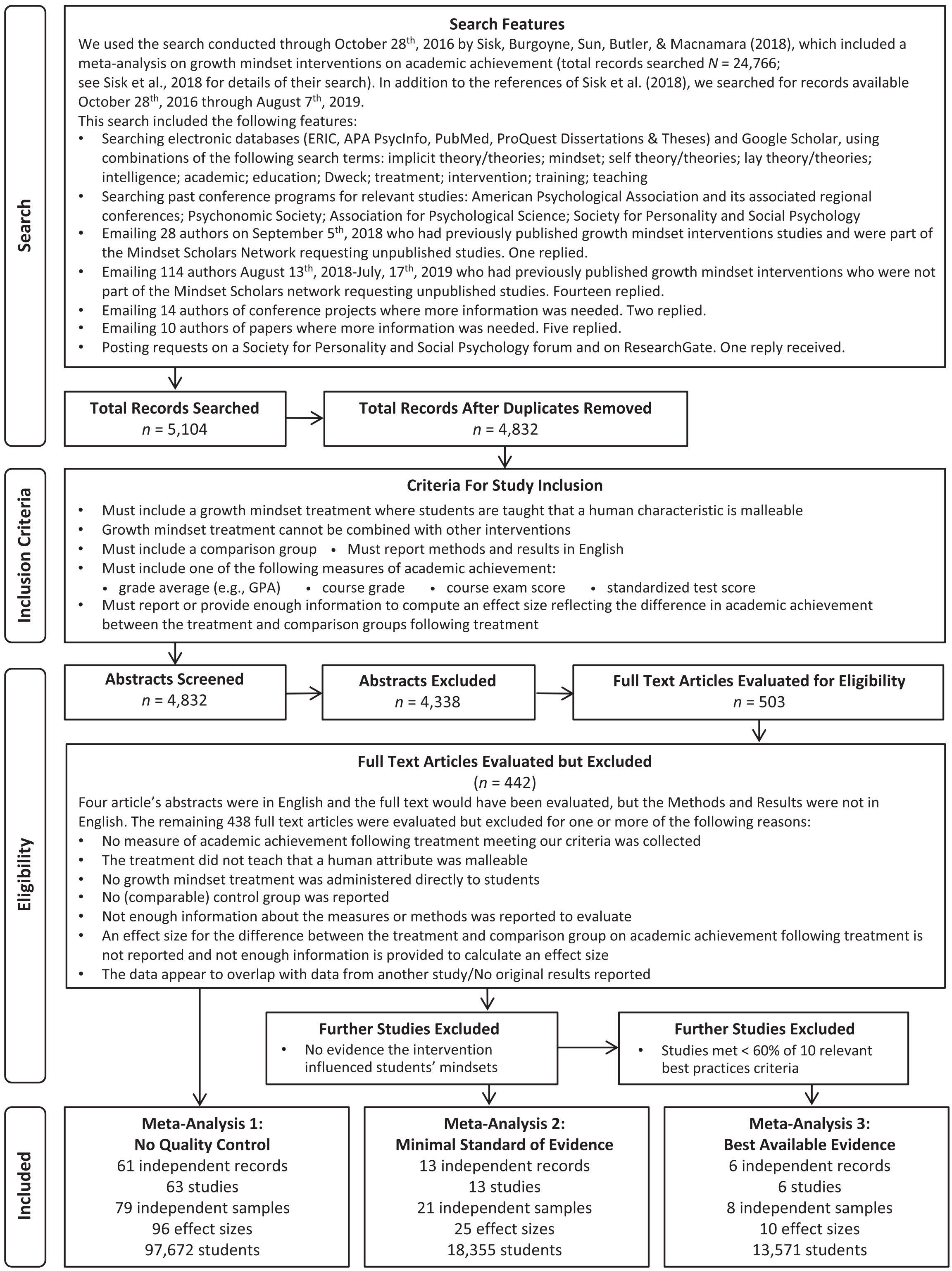
These inclusion criteria are identical to Sisk et al.'s (2018) inclusion criteria, with two exceptions. The first is that we included studies that only reported intent-to-treat analyses when we could not obtain the effects for students who completed the intervention (i.e., per-protocol analyses). The second is that we only included interventions that directly taught that a human characteristic could change. This criterion excluded studies where the apparent goal was to induce growth mindsets, but the manipulation only involved praising effort or attributing success to effort.

Praise-only and attribution-only interventions were described as instilling a growth mindset for many years (see, e.g., Dweck, 2007c, 2008b, 2010; Dweck et al., 2014; Haimovitz & Dweck, 2017; Levy & Dweck, 1999; Walton & Wilson, 2018; Yeager & Dweck, 2012). For example, Mueller and Dweck (1998) performed a now-classic praise-only study where Dweck (2008b) later described the manipulations as follows: "intelligence praise instilled more of a fixed mindset, making students believe that their intelligence was a fixed trait, whereas the effort praise instilled more of a growth mindset" (p. 57). Later on, after Li and Bates (2019) attempted near replications of Mueller and Dweck (1998) and concluded the results failed to replicate, Dweck and Yeager (2019a) argued that praise- and attribution-only studies are *not* mindset studies. They stated: "In mindset studies, participants receive explicit instruction about the malleability of ability" (p. 18).

The included studies came from two distinct searches. We used the output from the systematic search conducted by Sisk et al. (2018) through October 28, 2016, who used psychology-oriented, education-oriented, and multidiscipline databases (APA PsycINFO, ERIC, PubMed, ProQuest Dissertations & Theses), as well as Google Scholar, emailing authors, contacting organizations, and posting requests for data on a Society for Personality and Social Psychology forum (see Sisk et al., 2018, for details).

Following the methods used by Sisk et al. (2018), we systematically searched for relevant published and unpublished studies that became accessible between October 28th, 2016 and August 7th, 2019. See Figure 1. For records in 2016 where it was not clear which month they became available, we rescanned them in the current search. A team of five (the authors and three research assistants) searched the databases, Google Scholar search engine year by year, and past conference programs for references based on the search terms listed in Figure 1. The authors divided the references by year

Figure 1
Flowchart Describing the Literature Search



to screen the abstracts and evaluate the full-text articles. If we encountered nonidentical duplicates of the same study, we used the report that included more participants. If they included the same number of participants, we used the published version. We also searched for published versions of unpublished studies included in Sisk et al. (2018) that were also included in the present study. In the Supplemental Materials, we provide a table of studies that met many but not all of the inclusion criteria and note why they were excluded from the meta-analyses.

Including the records previously identified by Sisk et al. (2018), altogether we identified 61 independent records, which included 63 studies and 79 independent samples with a total sample size of 97,672 students. Thus, relative to Sisk et al. (2018), the present meta-analysis includes more than double the number of records and nearly double the total number of participants.

Effect Sizes and Associated Variances

For each study, we obtained the standardized mean difference (Cohen's *d*) between the treatment and control group to measure the magnitude of the effect of the growth mindset intervention on academic achievement. When possible, we accounted for baseline academic achievement. When multiple options were available for obtaining the effect size, we used the first option available from this list: (a) raw data; (b) mean post–pre difference score of the treatment group minus the mean post–pre difference score of the control group divided by the pooled standard deviation of the difference scores; (c) mean post–pre difference score of the treatment group minus the mean post–pre difference score of the control group divided by the pooled preintervention standard deviation (Morris, 2008); (d) ANCOVA or regression accounting for baseline achievement; (e) postintervention group means and standard deviations; (f) *t*-test or simple effects analysis; (g) reported Cohen's *d*, 8) another statistical test (e.g., analysis of variance [ANOVA]). We used the Campbell Collaboration effect size calculator (Lipsey & Wilson, 2001; https://campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD-main.php) except in the case of Option (c), which is not an option in the Campbell Collaboration effect size calculator and was manually calculated.

We also obtained the sampling error variance associated with each Cohen's *d*. We obtained the sampling error variance from the Campbell Collaboration effect size calculator. In the case of Option (c) in the previous paragraph, we manually calculated variance (see Coe, 2002).

For studies with multiple standardized tests (e.g., verbal SAT and quantitative SAT) or multiple measures of both grade and standardized test (e.g., an effect size for GPA and an effect size for performance on a standardized test from the same students), we adjusted the variance by applying Cheung and Chan's (2004, 2008) method, which combines dependent effects and adjusts the associated sample size. We used the weighted mean of all effect sizes for the weighted average component of their formula. When dependent effects were relevant to different levels of a moderator (e.g., an effect size for GPA and an effect size for performance on a standardized test from the same students when examining academic achievement type), the dependent effects were treated as independent for that moderator analysis so that each relevant effect could be included.

For studies where assignment to condition occurred at the classroom or school level rather than at the student level, we adjusted the variance associated with these effect sizes by applying Kish's (1965) design effect adjustment, which accounts for the average cluster size (e.g., average number of students per cluster) and the intraclass correlation. If a study did not report the intraclass correlation, we used the intraclass correlation default used by the What Works Clearinghouse (2020) for achievement measures ($\rho = .20$).

Coding

Studies were coded for reference information, sample sizes, sample descriptions, school level of the students involved, country of origin, type of mindset taught, publication status, whether the effect was significant, and whether the article reported any significant effect on academic achievement. When degrees of freedom did not correspond with the sample size reported, we calculated the sample size from the degrees of freedom of the relevant analysis. If the sample size per group was reported for baseline (preattrition), but only the total sample size was reported for analysis, we assumed attrition was even between groups and estimated the sample size per group based on the preintervention proportions. If preintervention proportions were also not available, we assumed equal sample sizes (if there was an odd number of participants, the additional subject was included in the treatment group).

We coded the information used to calculate or code the effect size (e.g., means and standard deviations, reported Cohen's *d*) and whether the effect size accounted for baseline academic achievement. We coded for student factors and methodological factors that served as preregistered potential moderators (described in the following subsections). We coded whether or not the study met study design and reporting characteristics critical for mechanistic interpretation, including best practices (described following the moderator descriptions) and other interpretation considerations such as whether the groups were equivalent at baseline on academic achievement, whether the familywise error rate was adjusted (if appropriate), and whether tests of subgroup differences were performed in cases where subgroup differences were claimed or implied.

A team of three coded the study features. First, each author coded half the studies independently. Second, the first author and a research assistant checked all studies and made any necessary coding adjustments, including a detailed description of the reason for the adjustment. Third, the second author checked all studies, including the rationale for any further coding adjustments. Disagreements were resolved through discussion.

Moderators

Theoretical Moderators

Developmental Stage. Some researchers suggest that holding a growth mindset is particularly important during the tumultuous period of adolescence (e.g., Blackwell et al., 2007). Student samples were coded as children if they were primary-school students; adolescents if they were middle school, junior high, or secondary school students; and adults if they were postsecondary school students.

Academic Challenge Status. Mindset theory holds that academically struggling students (e.g., Paunesku et al., 2015) and students facing situational challenges such as transitioning to a new school (e.g., Blackwell et al., 2007) are especially likely to benefit from a growth mindset, more so than students without these risk factors. Student samples were coded as high challenge if the majority of the sample had low grades or was otherwise at risk of failing; moderate challenge if the majority of the sample was facing a situational challenge: transitioning to a new school or under a stereotype threat manipulation; or low challenge if the majority of the sample had no indicators of risk or additional challenge.

Some studies reported the results of the whole sample and a high-challenge subsample, but not the remainder of the sample who were less challenged. For this reason, we conducted this moderation analysis twice, once with the full samples and once where we replaced these studies' full samples with subsamples of students with a higher challenge status.

Socioeconomic Status. Mindset researchers suggest that growth mindsets are especially beneficial for students coming from low socioeconomic status (SES) households (e.g., Claro et al., 2016). Student samples were coded as low SES if the majority of the sample came from financially poor households (e.g., qualified for free or reduced-price lunch) or not low SES if the majority of the sample came from middle-class or upper-class households. Studies without sample-level SES information were not included in this moderator analysis. The exception to this rule was if all or almost all students at the school (or in that grade) participated in the study. In that case, we accepted school-level SES information.

Some studies reported the results of the whole sample and a low-SES subsample, but not the remainder of the sample who were not low SES. For this reason, we conducted this moderation analysis twice, once with the full samples and once where we replaced these studies' full samples with low-SES subsamples.

Time Interval Between the Intervention and the Measure of Academic Achievement. According to mindset theory, growth mindset interventions interact with recursive processes (Yeager & Walton, 2011), producing enduring changes that compound benefits over time (Dweck et al., 2014). The time interval between the end of the intervention and the measure of academic achievement was coded as immediate (immediately following the final growth mindset intervention session); short (within 4 months [approximately a semester's time] of the mindset intervention); or long (longer than 4 months following the mindset intervention).

When studies included measures of academic achievement at multiple time points postintervention, we used the longest interval available within the same academic context (where students are taking the same classes) as the intervention (e.g., same academic year for elementary school students, same semester for college students). We repeated the moderation analysis with the longest interval available, regardless of context.

Methodological Moderators

Intervention Type. Interventions were coded as passive (i.e., students read a document or watched a video on how a human characteristic is malleable); feedback (i.e., students were given feedback on their performance in terms of growth mindset); or interactive (e.g., participants read materials and then wrote letters to future students about how intelligence can be developed).

Number of Sessions. We recorded the number of growth mindset intervention sessions delivered to the sample. Differing from Sisk et al. (2018), "filler" sessions that did not contain any growth mindset content were not counted.

Mode of Intervention. Modes were coded as a computerized program (e.g., Brainology computer program); reading materials (e.g., reading how intelligence can change); in-person training (structured discussion or lecture); or a combination of modes (e.g., computerized training and in-person training).

Intervention Administrator. Of those interventions that were administered in person (either solely or as part of a combination of modes), we coded whether the administrator was the teacher; researcher; the teacher who was also the researcher; both teachers and researchers; or other (another administrator).

Intervention Context. We coded intervention context as integrated into classroom activities (e.g., teacher fosters discussion of mindset in class) or outside the classroom (e.g., researchers conduct the intervention in a lab).

Academic Achievement Measure. We coded academic achievement measures as course exam (e.g., final exam score); course grade (e.g., math course grade); GPA (cumulative or current GPA); or standardized test score (e.g., Iowa Test of Basic Skills, SAT). When studies reported multiple course performance measures, we used the most comprehensive one available (i.e., GPA over course grade, course grade over course exam, cumulative final exam over a midterm exam). When studies included multiple standardized test scores (e.g., verbal SAT, quantitative SAT, total SAT), we used the combined score when available.

Best Practices

We coded each sample for whether it met each best practice criterion. Studies without these characteristics increase the risk to internal validity in individual study results. While some best practices criteria are undoubtedly more critical than others, there is not yet a clear consensus on each characteristic's relative importance.

Best Practices in Intervention Design

The Intervention Is Compared to an Active Control Condition. The treatment must have been compared to an active (i.e., placebo) control group.

Aside From Manipulating Students' Beliefs About Attribute Malleability, There Are No Other Study Design Differences Between Treatment and Control Groups. There must not be differences between the treatment and the control conditions other than teaching that intelligence or another characteristic is malleable.

A Priori Power Analysis Conducted. The study report must indicate that an a priori power analysis was conducted.

Random Assignment to Condition at the Individual Student Level. The study report must indicate that students were randomly assigned to condition at the individual level.

Blinding. The study report must indicate that the students, teachers, and intervention administrators were unaware of students' assignment to condition.

Including Manipulation Checks. The study report must indicate that a pre- and postintervention measure of mindset was administered to the treatment group.

Best Practices in Documentation, Analyses, and Reporting

Study Hypothesis, Planned Sample (and Stopping Rule), Intervention Design, and Analysis Plan Were Preregistered and Followed. The preregistration must describe the hypotheses or aim of the study, planned sample, growth mindset intervention design, planned analyses, and the authors must have reported all measures and results and followed these methods and analyses, or were clear when there was a change from the preregistered plan. A preregistration must have been completed and posted prior to examining the data. Arguably, preregistrations were not clearly introduced to psychology researchers until November 27th, 2013 (Eich, 2014). However, 89% of the reports in the present study were produced after this date.

Reporting Results of Those Who Participated. The study report must provide the results of participants who received the treatment and control protocols.

Reporting the Relevant Results of the Whole Sample or All Subsamples. The study report must provide the critical results for the whole sample and/or all subsamples. For example, if focusing on a particular subsample, that subsample's results and the remainder of the sample should be reported (either together or separately) rather than excluding subsamples.

Best Practices in Avoiding Conflicts of Interest

Having No Financial Interest in a Particular Outcome. No authors of the study (a) founded, are employed by, or are a consultant for a for-profit or nonprofit entity that sells or promotes growth mindset interventions or services; (b) are registered with a speaker's bureau to talk about the merits of growth mindset for pay; or (c) receive book sale profits from a self-help book promoting growth mindset. The authors and research assistants searched the authors' webpages; within speakers' bureaus' websites; and authors' names ± "mindset" on Google, examining relevant links to determine financial interest status.

State of the Literature

Here, we describe the growth mindset intervention literature landscape. This section does not test hypotheses. Rather, it provides an overview of the state of the growth mindset intervention literature based on features for which we coded and patterns we observed during our search and screening processes.

First, we describe observed patterns of growth mindset interventions entering the literature. Second, we describe patterns of significant effects across the mindset intervention literature. Third, we provide descriptive statistics on adherence to best practices. Last, we provide descriptive statistics for studies in terms of the students involved and methodologies employed.

Interventions Entering the Literature

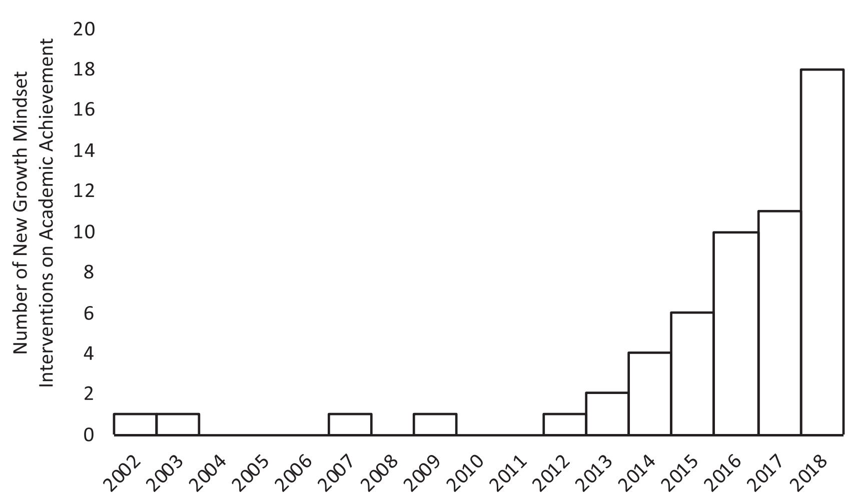
Rate of Growth Mindset Interventions Entering the Literature

The number of growth mindset intervention studies examining academic achievement appears to be increasing rapidly, see Figure 2. Of the 61 independent records (e.g., articles) included in the present study, 44 were from 2016 or later.

Publication Patterns

Thirty-eight (62%) of the included records were not published in journals or books. With the advent of internet-accessible repositories (e.g., PsyArXiv), an unpublished study is no longer necessarily an unknown study. Given the popularity of growth mindset interventions, unpublished studies can include summaries with press releases, which remain nonpeer reviewed, but are reported on by media outlets (e.g., Outes et al., 2017 in World Bank Blogs, 2017). Some studies are commissioned by large organizations with reports that are evaluated by experts but are not submitted to journals for publication (e.g., Churches & Education Development Trust, 2016). Thus, some

Figure 2
Histogram of Independent Records Included Through the Last Full Year of Our Search



unpublished studies enter the scientific literature and popular media with varying amounts of peer review and quality control.

Despite the increase in attention paid to some unpublished studies, many unpublished studies are largely unknown, and publication bias remains a major problem in psychological science (Friese & Frankenbach, 2020). Publication bias is bias in effect size estimates that occurs when the likelihood of publication depends on the direction or magnitude of the study's results (e.g., Dickerson, 2005); typically, null or small effects are systematically suppressed from publication (Rosenthal, 1979).

Publication suppression leads to a skewed representation of results in the available literature. Skewed representation can lead to misunderstandings regarding the strength or reliability of effects. Publication bias can be due to journals rejecting certain studies but is most often caused by authors not submitting null or small effects for publication (see Cooper et al., 1997; Franco et al., 2014). In the following subsections, we describe differences found between published and unpublished studies more formally using other procedures.

Excluding Participants Before Publication. For around 10% of the studies included in this systematic review, we discovered both published and unpublished versions. In five of the seven cases, participants' data included in the unpublished version were excluded from the published version. The number of excluded participants ranged from four (2% of the sample) to 6,222 (50% of the sample). Following our preregistration, we used the effect reported in the unpublished version when it contained a larger sample than the published version. See Table 1.

We cannot know how many published studies without accessible prepublication records removed participants' data prior to publication. In our search, the majority of published studies for which we found an unpublished version excluded participants from the published study. This observation raises the possibility that removing data prior to publication may be common in the growth mindset intervention literature. It could also be that these studies are not representative of the growth mindset intervention literature. Regardless, we recommend growth mindset intervention researchers preregister and follow clear exclusion criteria as well as design studies to minimize the number of cases removed.

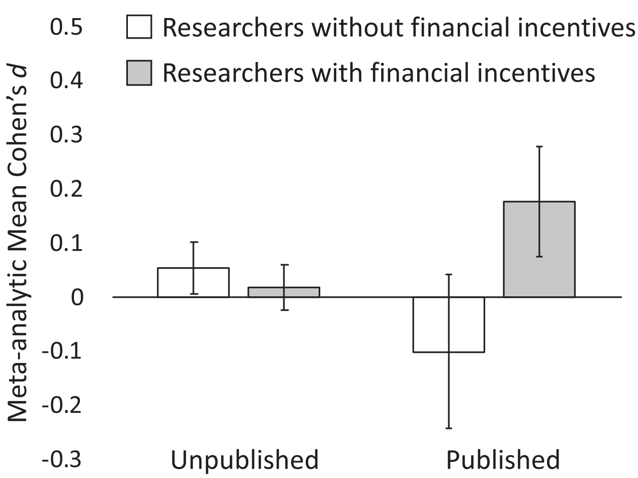
Published Versus Unpublished Effects. One way to examine potential publication bias is to compare the meta-analytic average effect from published documents (37 effects), $\bar{d} = 0.09$, 95% CI [−0.01, 0.18], $p = .073$, versus unpublished documents (59 effects), $\bar{d} = 0.04$, 95% CI [0.003, 0.07], $p = .032$. They were not significantly different from one another, $Q(1) = 0.96$, $p = .328$.

This null result appears to be due to different publishing patterns between researchers with a financial incentive to find positive effects versus researchers without such a financial incentive. For studies authored by one or more researchers with a financial incentive to report positive effects, the average effect from published studies (19 effects), $\bar{d} = 0.18$, 95% CI [0.07, 0.28], $p = .001$, is significantly larger than the average effect from unpublished studies (9 effects), $\bar{d} = 0.02$, 95% CI [−0.02, 0.06], $p = .375$; $Q(1) = 7.90$, $p = .005$. In contrast, for studies where no authors have a financial incentive to report positive effects, there is no significant difference between the average effect from published studies (18 effects), $\bar{d} = -0.10$, 95% CI [−0.24, 0.04], $p = .164$, and the average effect from unpublished studies (50 effects), $\bar{d} = 0.05$, 95% CI [0.01, 0.10], $p = .030$; $Q(1) = 4.04$, $p = .044$ (Bonferroni corrected $\alpha = .025$). The Publication Status × Financial Incentive Status interaction is significant, $F(1, 75) = 6.35$, $p = .014$. See Figure 3.

Table 1
Unpublished Versions Used Due to Excluded Participants in the Published Versions

References	Version used in our review and meta-analysis and rationale	Study authors' rationale for participant data exclusion
Bostwick (2015), Bostwick and Becker-Blease (2018)	We used the 2015 unpublished master's thesis because four participants were missing from the published version.	No rationale given. The dissertation reports the final exam results for 260 participants (across three conditions) whereas the published version (see their Supplemental Materials) reports the final exam results for 256.
Broda (2015), Broda et al. (2018)	We used the 2015 unpublished dissertation because 378 participants were missing from the published version.	No rationale given. These 378 students could be all or some of the international students who were explicitly excluded from analysis without rationale in the published version.
Burnette et al. (2016), Burnette et al. (2019)	We used the 2016 unpublished preprint because 13 participants were missing from the published version.	No rationale given. The Method describes the full sample size, but degrees of freedom suggest 13 participants were not included in the final grade analysis in the published version.
Yeager et al. (2018), Yeager et al. (2019)	We used the 2018 unpublished preprint because 6,222 participants were missing from the published version.	Authors state they only report the results of 50% of the sample following the preregistered analysis plan. However, according to this document, the first research question and planned analysis described estimating the treatment effect for the whole sample.
Zonnefeld (2015), Zonnefeld (2019)	We were able to obtain information about the full (combined) sample of 445 students from the author.	No rationale given. The 2015 unpublished dissertation included 411 students. The 2019 published book chapter added 34 treatment students but removed 270 control students.

Figure 3
Differences in Effects of Published Versus Unpublished Growth Mindset Interventions Depending on the Authors' Financial Incentives to Report Positive Effects



Note. Error bars represent 95% confidence intervals.

Patterns of Significant Effects

As we obtained effect sizes for the meta-analyses, we coded whether the effect was significant or not. In this section, we examine patterns of significance for these effects.

Financial Incentives

Articles with one or more authors with a financial incentive to report positive effects (30% of documents) were more than two and a half times as likely (56% vs. 21%) to find a significant difference between treatment and control groups on academic achievement in the hypothesized direction than articles where no authors had a financial incentive, $\chi^2(1, N = 61) = 7.09$, $p = .008$.

This result suggests that financial incentives may be the source of this difference. It is also possible that this difference is due to noise or a factor other than financial incentives. We considered the possibility that mindset theory proponents, regardless of financial incentives, might be more likely to find significant effects, whereas mindset theory skeptics might be more likely to find null or negative effects. We searched study reports and authors' social media for language suggestive of mindset theory support and skepticism in order to classify authors. We found no compelling evidence that mindset skeptics conducted any of the growth mindset interventions included in the current meta-analysis. All authors either appeared neutral or expectant of positive effects.[4] Mindset theory skeptics appear less likely to conduct growth mindset interventions than mindset theory proponents.

Type I Error

Less than one-third of the documents (19 of 61) reported a statistically significant difference between treatment and control groups on academic achievement in the hypothesized direction. For multiple studies reporting significant effects, the researchers assigned students to condition by group (e.g., by classroom) rather than on an individual basis and did so without correction. These studies violate the assumption of independence and underestimate sampling error, as students within a classroom or school will be more similar to each other and more different from other groups due to factors (e.g., shared teachers, shared schedule) other than the experimental condition. Consider the case where a researcher assigns one classroom to receive the treatment and another classroom to receive the control condition. Differences in grades cannot be solely attributed to the treatment manipulation: Students sharing a classroom have commonalities in their educational environment other than the treatment and therefore some degree of relatedness. Cluster sampling without adjustment inflates the between-subjects variance to within-subjects variance ratio, thereby increasing the rate of Type I error (see Chow & Liu, 2004, for review).

The design effect (*deff*; Kish, 1965) is the ratio of the operating sample variance to the variance expected with simple random sampling (for reviews see, e.g., Alimohamadi & Sepandi, 2019; Hox, 1998; McCoach & Adelson, 2010). The design effect accounts for the similarity of students within clusters (the intraclass correlation) and the average number of students per cluster (cluster size). *Deff* is multiplied by the observed variance (calculated according to simple random sampling) to provide the true sampling variance (Hox, 1998). Even when the design effect is small, not correcting for cluster sampling underestimates sampling variance and produces highly misleading significance test results (e.g., Hox, 1998; McCoach & Adelson, 2010).

Indeed, after adjusting for the design effect (using a default intraclass correlation of $\rho = .20$ when the intraclass correlation was not reported by the study authors), several studies' effects were no longer significant. Most notably, the effect in Blackwell et al.'s (2007) highly cited publication indicating that growth mindset interventions can improve grades—reported as $b = .53$, $t = 2.93$, $p < .05$ (see p. 257)—was no longer significant ($p = .119$) after correcting for the design effect.

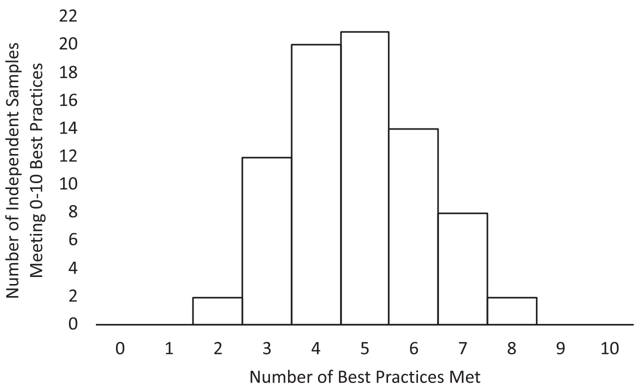
Significant Effect Interpretation Issues

About 10% of the null effects in this literature were described by the study authors as though they were significant. For example, Good et al. (2003) stated, "The incremental condition increased both boys' and girls' math performance" (p. 657). Yet, there was no significant increase in boys' math performance in their studies. Nonsignificant results described as significant likely causes confusion about the impact of growth mindset interventions.

Of the articles that reported a significant effect, about half attributed the effect to the intervention influencing students' mindsets without evidence the intervention had any effect on students' mindsets. For example, Outes et al. (2017) titled their project "Growth mindset at scale—Increasing school attainment by affecting the mindset of pupils and teachers." Yet, there is no indication that Outes et al. ever measured change in mindsets and so we cannot know whether the intervention affected the mindsets of pupils or teachers.

[4] According to Dweck and Yeager (2019a), Yue Li and Timothy Bates are skeptics of mindset effects. No studies conducted by Li or Bates were included in the meta-analysis. Li and Bates' (2019) failed replication and extension of Mueller and Dweck's (1998) study involved a manipulation of praise/effort attributions and fixed mindset and therefore did not meet our inclusion criteria for a growth mindset intervention.

Figure 4
Histogram of Independent Samples Meeting Best Practices Criteria



Relatedly, Yeager et al.'s (2014) Study 3 manipulation check failed: They found no measurable evidence for the impact of the growth mindset intervention on students' mindsets. Nevertheless, they concluded, "we found evidence for the relation between implicit theories of personality and overall adjustment in both [Study 2 and Study 3] .... Importantly, justification for causal inferences within these schools was clear" (p. 880). However, when the manipulation check fails, it is unclear whether the latent construct of growth mindset is responsible for the causal effect of the intervention. To attribute effects to changes in growth mindset, researchers need to provide evidence that the intervention changed students' mindsets.

Best Practices Criteria

Earlier, we described study characteristics that are critical for interpreting the treatment mechanism. In many cases, these study characteristics can be met by following best practices in study design, reporting, and avoiding conflicts of interest. Failure to implement these best practices does not indicate an intentional effort to deceive or bias the results. However, these best practices should be considered when evaluating a study or body of evidence and when planning future studies.

Table 2
Adherence to Best Practices

Best practice	Percent of samples (%)	Percent of total students (%)
Intervention design		
Tested whether the treatment group's mindset changed	59	33
Active control group comparison	58	33
Individual students randomly assigned to condition	51	33
Students, administrators, and teachers blind to condition	28	29
A priori power analysis conducted	25	26
Only the critical manipulation differed between groups	6	3
Analyses and reporting		
Reported results for the whole sample/all subsamples[a]	96	99
Reported effect on those treated	87	64
Preregistered hypotheses, methods, and planned analyses	3	7
Avoiding conflicts of interest		
No financial incentives to find a positive effect	68	76

[a] This includes cases where subsamples were not reported in the published version if we were able to obtain subsample results from an unpublished version of the study.

How well do studies in the growth mindset intervention literature adhere to best practices criteria? Adherence to best practices varied considerably, with the majority (70%) of the independent samples meeting 50% or fewer best practices criteria (see Figure 4). Table 2 provides information about the entire set of studies and reports the percentage of the 79 independent samples and of the 97,672 students associated with each sample that met each best practice criterion. See open data (https://osf.io/ajhxv/) for the best practices criteria and other interpretation considerations met by each sample.

Only two studies (Foliano et al., 2019; Polley, 2018) had accompanying preregistrations that included relevant research aims, methods, and planned analyses, and were registered prior to examining the data. Six other studies reported that they were preregistered, but we did not code them as such for reasons outlined in Table 3.

Table 3
Studies Described as Preregistered That We Did Not Code as Preregistered and Rationale Why

Study reference	Rationale for not coding as preregistered
Burnette et al. (2018)	No hypotheses or analysis plan about the growth mindset intervention were included in the preregistration
Ganimian (2018)	Preregistration did not include an analysis plan
Orosz et al. (2017)	Authors state the study was in accord with the Declaration of Helsinki, but we could find no record of a preregistration
Yeager, Romero, et al. (2016)	There is no preregistration associated with this study. There is a wiki statement with hypotheses but no methods. It was posted approximately 3 months after data were collected, 3 weeks before manuscript submission. Authors state the study was preregistered before researchers received the data from the third-party research firm, but employees of the third-party research firm were study authors
Yeager, Lee, et al. (2016)	The preregistration contains no hypotheses, methods, or planned analyses for the impact of a growth mindset intervention on academic achievement
Yeager et al. (2018)	The registration was created after examining the data and analyzing a portion (approximately 760 participants) to "inform" the preregistration (see p. 3 of document)

Note. Burnette et al. (2018) clearly state the preregistration does not pertain to the growth mindset intervention.

Table 4
Included Studies' Student Characteristics

Sample characteristic	Number of effect sizes	Number of students
Developmental stage		
Adults	34	13,017
Adolescents	47	75,507
Children	11	6,069
Mixed sample	3	2,967
Unclear	1	112
Academic challenge		
High	16 (19)	18,633 (18,949)
Moderate	31 (30)	18,327 (17,530)
Low	46 (44)	17,982 (17,801)
Unclear	3	42,730
Socioeconomic status		
Not low	23 (15)	28,865 (21,263)
Low	16 (24)	2,959 (5,504)
Not reported	57	65,848

Note. Unclear = Study reported limited sample information. Numbers in parentheses represent the numbers when replacing whole samples' effects with relevant subsamples' effects. For example, a reported high-challenge subsample replacing an overall low-challenge whole sample increases the high-challenge effect size count and the number of students while decreasing the low-challenge effect size count and the number of students.

Student, Methodological, and Other Study Characteristics

Tables 4, 5, and 6 provide the number of effect sizes and number of students associated with sample characteristics (Table 4), intervention methods and outcome measures (Table 5), and other characteristics associated with the study's article (Table 6) for all studies included in this systematic review and meta-analysis.

Meta-Analyses

Meta-Analysis 1: No Quality Control

Meta-Analysis 1 included all relevant studies from our comprehensive search, regardless of quality.

Method

Meta-Analytic Procedure. Following study coding (see the "Coding" section), including variance adjustments for dependent measures and the design effect (see the "Effect Sizes and Associated Variance" section), we searched for extreme values. We defined outliers as effect sizes whose residuals had z-scores ≥ 3. One outlier was identified ($d = 1.51$, residual z-score $= 4.24$, study $N = 46$). We Winsorized the effect such that the residual z-score was 2.99 ($d = 1.07$).

Analyses. To determine the average effect of growth mindset interventions on academic achievement for all studies regardless of study quality, we conducted a random-effects meta-analysis, which assumes meaningful differences across studies. The random-effects meta-analytic model estimates the meta-analytic mean effect size and heterogeneity of effect sizes. Effect sizes were weighted by the inverse of the variance (i.e., larger studies were weighted more heavily than smaller studies in the model). To test whether some of the heterogeneity could be accounted for by moderator variables, we conducted mixed-effects meta-analytic modeling. These moderator analyses included theoretically motivated variables, methodological variables, and bias-related variables.

Table 5
Included Studies' Intervention Method and Outcome Measure Characteristics

Intervention/Outcome measure characteristic	Number of effect sizes	Number of students
Intervention type		
Interactive	88	93,648
Feedback	2	3,195
Passive	6	829
Mode of intervention		
Combination in-person and other mode	12	856
In-person	44	65,736
By teacher(s)	30	53,798
By researcher(s)	11	890
By teacher who is also the researcher	6	221
By both teachers and researchers	1	155
Other administrator(s)	8	11,528
Computer program	34	30,035
Reading materials	6	1,045
Mindset type		
Intelligence	74	91,377
Other (e.g., math ability, personality)	13	4,231
Intelligence + Other	9	2,064
Number of sessions		
One	35	63,301
More than one	61	34,371
Intervention context		
Outside regular classroom activities	59	81,527
Integrated into classroom activities	37	16,145
Time interval to outcome measure[a]		
Immediate (end of last session)	3	186
Short (within 4 months)	56 (55)	70,881 (70,316)
Long (longer than 4 months)	29 (30)	21,522 (21,994)
Unclear	8	6,403
Academic achievement measure[a]		
Standardized test	38	63,495
Laboratory measure	10	587
Course grade average (e.g., GPA)	28	26,392
Course grade	17	5,330
Course exam grade	13	4,050

Note. GPA = grade point average.
[a] The cumulative sample size is greater than 97,672 due to multiple measures from the same sample in different categories. Numbers in parentheses represent the numbers when replacing the effect associated with the longest time interval within the same academic context with the effect associated with the longest time interval regardless of context.

To evaluate the impact of publication bias on the overall effect, we conducted several publication bias analyses including examining a funnel plot, conducting Egger's regression analysis (Egger et al., 1997), assessing trim-and-fill results (Duval & Tweedie, 2000), and conducting a conditional PET-PEESE model. We used the Comprehensive Meta-Analysis Version 2 (Borenstein et al., 2005) software package for all analyses except the conditional PET-PEESE where we used R.

Results

Overall Results. Across all studies regardless of quality, the overall meta-analytic average standardized mean difference in academic achievement between students receiving a growth mindset

Table 6
Characteristics Associated With the Included Studies' Articles

Associated characteristic of article	Number of effect sizes	Number of students
Financial incentive to find positive effects		
One or more authors	28	23,276
No authors	68	74,396
Publication status[a]		
Published	37	8,949
Unpublished	59	88,723
Combinations of the above two factors		
Financial incentive + Published	19	6,581
No financial incentive + Published	18	2,368
Financial incentive + Unpublished	9	16,695
No financial incentive + Unpublished	50	72,028

[a] In cases where a published and unpublished report of the same study were accessible, we used the published version if both reports included the same number of participants. If study participants' data were excluded prior to publication, we used the unpublished version with more study participants.

intervention and students in a control group was $\bar{d} = 0.05$, 95% CI = [0.02, 0.09], $p = .004$. See Figure 5.

Moderator Analyses. The between-study variability in effect sizes due to heterogeneity rather than random error was moderate, $I^2 = 39.14$ ($\tau^2 = .005$). We investigated the source of this heterogeneity through moderator analyses. We conducted moderator analyses when there were at least five effect sizes contributing to a subgroup (Williams, 2012). See Tables 4 and 5 for the number of effect sizes and sample sizes contributing to each subgroup.

Mindset theory holds that growth mindset interventions will be especially important under certain circumstances. Counter to this assumption, no theoretically meaningful moderators were significant, indicating that the effect sizes did not differ significantly across levels of the moderator. See Table 7 for a summary of the results of the moderator analyses.

Some moderators were correlated with one another. For example, Intervention Context and Administrator were correlated: When the intervention was integrated with classroom activities, the classroom teacher, rather than the researcher, was more likely to be the intervention administrator. See Supplemental Materials for the moderator correlation matrix.

Bias Analyses. To test the extent to which our results were biased, we conducted several bias analyses. See Table 6 for the number of effect sizes and sample sizes contributing to each subgroup.

Financial Incentives. We found that the average effect size of studies where an author had a financial incentive to find positive results was not significantly different from the average effect size of studies where no authors had perceived financial conflicts of interest, see Table 8.

Financial Incentives in the Published Literature. As described earlier, following data collection we observed a distinction between authors with financial incentives versus authors without financial incentives in terms of which effects were reported in the published literature. Authors with financial incentives found both positive and null results, yet, often, only their larger, significant effects were published. Unpublished effects included studies yielding null results as well as subsamples yielding null results. For example, a subsample of over 6,000 participants included in a preprint (Yeager et al., 2018) yielding a null result was removed from the published version (Yeager et al., 2019). See Table 1.

Although not a preregistered analysis, we tested whether, within the published literature (studies coded as published), there was a significant difference between studies where an author had a financial incentive to report positive effects, and studies where no authors had known financial incentives to report positive effects. We found that the effect for published studies where an author had a financial incentive to report positive effects was significantly larger than the effect for published studies where no authors had a financial incentive to report positive effects, see Table 8 and Figure 3. Thus, financial incentives may influence which effects are published.

Publication Bias Analyses. We additionally conducted several publication bias analyses, each with their own strengths and weaknesses, to test whether multiple approaches supported the same conclusion.

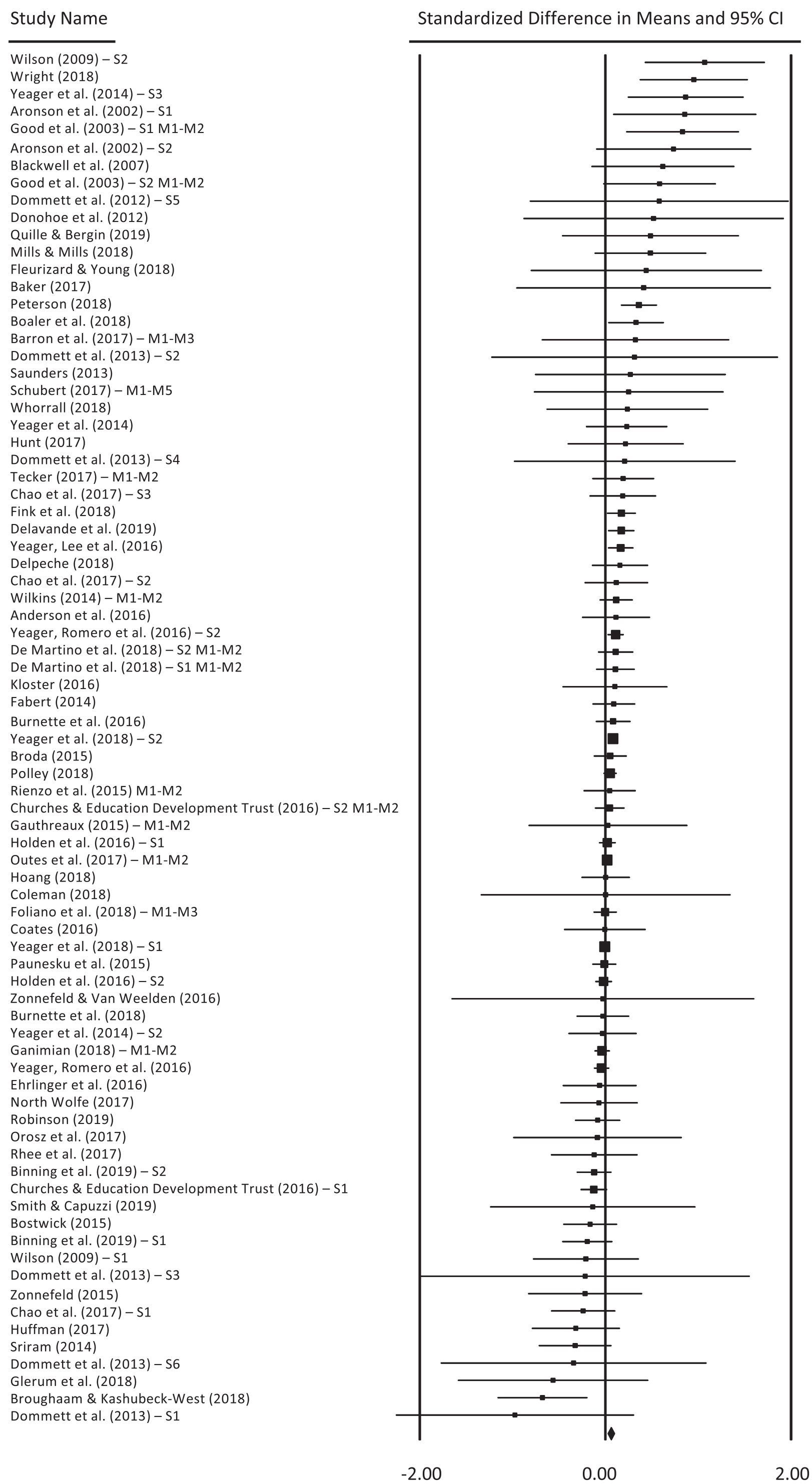
Funnel Plot and Egger's Regression. Funnel plots depict the relationship between a study's effect size and sample size. Larger studies, barring other forms of bias, should cluster around the meta-analytic mean effect. Smaller studies introduce more random sampling error and therefore should have a wider scatter around the mean. Importantly, these deviations should be random and thus equally likely to be higher or lower than the mean. An indicator of potential publication bias is if the funnel plot is asymmetrical, with smaller studies with smaller-than-average effects "missing" from the left side of the plot.

Visual inspection may lead to misleading impressions of asymmetry (Simmonds, 2015; Terrin et al., 2005). Therefore, funnel plot asymmetry was quantified by Egger's regression. For this meta-analysis, Egger's regression suggests multiple studies may have been conducted but are missing from the available literature and are not included in the present meta-analysis (recommended 1-tailed $p = .034$; Borenstein et al., 2005; 2-tailed $p = .067$), see Table 8. The positive intercept indicates that the funnel plot is biased toward the right of the mean (i.e., toward larger-than-average positive effects), which suggests studies with smaller-than-average effects are missing (Lin & Chu, 2018). See Figure 6. This approach does not estimate the number of studies missing or estimate the overall effect if these studies were not missing.

Trim and Fill. Duval and Tweedie's trim-and-fill method (Duval & Tweedie, 2000) estimated that 10 studies with smaller-than-average effect sizes were missing from the available literature. This is not surprising, as some unpublished effects are only accessible to the original study authors. For example, during the literature search for their meta-analysis, Sisk et al. (2018) describe contacting a researcher (who had a financial incentive to find positive effects [see Sisk et al.'s, 2018; Supplemental Materials]) who refused to provide information about an unpublished study because replications had failed, and also declined to provide those failed replication effects. If the estimated 10 studies missing from our meta-analysis were available and included, Duval and Tweedie's analysis estimated the overall effect would be nonsignificant, see Table 8. Thus, the meta-analytic mean findings reported above might be inflated due to publication bias.

Figure 5

Each Sample's Effect Size and 95% Confidence Interval in Meta-Analysis 1: No Quality Control



Note. Square size is proportionate to the effect's weight (larger samples contribute more weight) with a minimum size imposed for visibility. The diamond on the bottom row represents the meta-analytically weighted mean Cohen's *d*. For studies with multiple independent samples, the result for each sample (S1, S2, etc.) is reported separately. Multiple measures resulting from a single sample were combined and adjusted for dependency (e.g., M1–M2).

Table 7
Moderator Results, Meta-Analysis 1: No Quality Control

Moderator and levels	Result		
	Theoretical factors		
Developmental stage		$Q(2) = 0.47$, $p = .791$	
Adults	$\bar{d} = 0.07$	95% CI [−0.01, 0.14]	$p = .089$
Adolescents	$\bar{d} = 0.04$	95% CI [0.002, 0.09]	$p = .041$
Children	$\bar{d} = 0.09$	95% CI [−0.06, 0.23]	$p = .240$
Academic challenge status[a]		$Q(2) = 0.28$, $p = .870$	
High-challenge level (e.g., low grades)	$\bar{d} = 0.07$	95% CI [0.01, 0.13]	$p = .019$
Moderate-challenge level (e.g., new school)	$\bar{d} = 0.05$	95% CI [−0.02, 0.12]	$p = .200$
Low-challenge level	$\bar{d} = 0.05$	95% CI [−0.01, 0.11]	$p = .083$
Socioeconomic status[b]		$Q(1) = 2.10$, $p = .147$	
Middle–high	$\bar{d} = 0.04$	95% CI [−0.02, 0.09]	$p = .182$
Low	$\bar{d} = 0.15$	95% CI [0.002, 0.30]	$p = .047$
Time interval to outcome measure[c]		$Q(1) = 0.21$, $p = .647$	
Short (interval ≤4 months)	$\bar{d} = 0.07$	95% CI [0.02, 0.12]	$p = .011$
Long (interval >4 months)	$\bar{d} = 0.05$	95% CI [−0.01, 0.11]	$p = .131$
	Methodological factors		
Intervention type		$Q(1) = 0.39$, $p = .534$	
Interactive (e.g., "saying-is-believing" task)	$\bar{d} = 0.06$	95% CI [0.02, 0.10]	$p = .003$
Passive (e.g., only reading materials)	$\bar{d} = -0.001$	95% CI [−0.18, 0.18]	$p = .995$
Number of sessions		$Q(1) = 6.01$, $p = .014$	
Slope	$b = 0.01$	95% CI [0.001, 0.01]	$p = .014$
Intervention delivery mode		$Q(3) = 4.33$, $p = .228$	
Reading material	$\bar{d} = 0.07$	95% CI [−0.05, 0.20]	$p = .245$
Computer program	$\bar{d} = 0.04$	95% CI [−0.0002, 0.08]	$p = .051$
In person	$\bar{d} = 0.04$	95% CI [−0.02, 0.10]	$p = .235$
Combination of delivery modes	$\bar{d} = 0.35$	95% CI [0.05, 0.64]	$p = .022$
Administrator (of In-person delivery)		$Q(3) = 0.40$, $p = .939$	
Teacher	$\bar{d} = 0.07$	95% CI [−0.01, 0.15]	$p = .103$
Researcher	$\bar{d} = 0.16$	95% CI [−0.12, 0.43]	$p = .260$
Teacher who is also the researcher	$\bar{d} = 0.11$	95% CI [−0.42, 0.64]	$p = .682$
Other	$\bar{d} = 0.08$	95% CI [−0.06, 0.23]	$p = .266$
Context		$Q(1) = 2.99$, $p = .084$	
In the classroom	$\bar{d} = 0.11$	95% CI [0.03, 0.19]	$p = .006$
Outside the classroom	$\bar{d} = 0.03$	95% CI [−0.01, 0.07]	$p = .091$
Academic achievement measure		$Q(3) = 1.61$, $p = .657$	
Course exam grade	$\bar{d} = 0.06$	95% CI [−0.01, 0.13]	$p = .093$
Single course grade	$\bar{d} = 0.10$	95% CI [0.01, 0.19]	$p = .026$
Multicourse grade average (e.g., GPA)	$\bar{d} = 0.03$	95% CI [−0.02, 0.09]	$p = .214$
Standardized test score	$\bar{d} = 0.06$	95% CI [−0.01, 0.13]	$p = .105$
Laboratory vs. actual standardized test		$Q(1) = 0.18$, $p = .674$	
Laboratory-based standardized test	$\bar{d} = 0.12$	95% CI [−0.19, 0.42]	$p = .454$
Actual standardized test score	$\bar{d} = 0.05$	95% CI [−0.02, 0.12]	$p = .179$

Note. GPA = grade point average.
[a] Two studies provided information for high-challenge-level subsamples. When replacing the whole samples of these studies with these subsamples, the pattern of results is unchanged. [b] Studies not reporting student-level socioeconomic status were not included in this moderator analysis. Four studies provided information for low socioeconomic subsamples. When replacing the whole samples of these studies with these subsamples, the pattern of results is unchanged. [c] For seven samples, a longer time interval was available beyond the academic context in which the intervention was administered. When replacing effects with those from these longer intervals, the moderator remains nonsignificant and long intervals become marginally significant: $\bar{d} = 0.06$, 95% CI [0.001, 0.12], $p = .046$.

Like Egger's regression, the trim-and-fill method decreases in accuracy as heterogeneity increases. The method can then over- or underestimate the effect of bias (Carter et al., 2019). We observed a moderate amount of heterogeneity. We, therefore, conducted another publication bias analysis to test whether the results were also in line with the results from Egger's regression and the trim-and-fill analysis.

PET-PEESE. To further test for publication bias, we performed a conditional precision-effect test and precision-effect estimate with standard errors (PET-PEESE; Stanley & Doucouliagos, 2014). We performed the analyses in R using the PETPEESE functions available online (Hilgard, 2015, 2020; https://github.com/Joe-Hilgard/PETPEESE). We used restricted maximum-likelihood estimation for the analyses. As with Egger's regression and Duval and Tweedie's trim and fill, PET-PEESE has limitations. Simulation studies suggest that although PET-PEESE performs well when there is one true population effect (i.e., in a fixed-effect environment), it performs less well when there is meaningful heterogeneity in the population effect (i.e., in a random-effects environment; Alinaghi & Reed, 2018; Hong, 2019), as is the case for the present meta-analysis.

Table 8
Bias Analyses Results, Meta-Analysis 1: No Quality Control

	Financial incentives		
All effects		$Q(1) = 1.29$, $p = .256$	
Author(s) with a financial incentive	$\bar{d} = 0.07$	95% CI [0.02, 0.13]	$p = .006$
No authors with a financial incentive	$\bar{d} = 0.03$	95% CI [−0.02, 0.08]	$p = .187$
Published effects		$Q(1) = 9.66$, $p = .002$	
Authors(s) with a financial incentive	$\bar{d} = 0.18$	95% CI [0.07, 0.28]	$p = .001$
No authors with a financial incentive	$\bar{d} = -0.10$	95% CI [−0.24, 0.04]	$p = .164$
	Test of missing studies due to publication bias		
Egger's regression		$B_0 = 0.36$, $t(77) = 1.86$, $p = .034$	
	Publication bias-corrected estimates of the overall effect		
Trim and fill	$\bar{d} = 0.03$	95% CI [−0.01, 0.07]	$p = .097$
Precision effect test (PET)	$\bar{d} = 0.01$	95% CI [−0.03, 0.05]	$p = .667$

The conditional PET-PEESE analysis consists of two components: A PET analysis in which the effect size is regressed on its standard error, and a PEESE analysis in which the effect size is regressed on its variance. For both analyses, the intercept represents the estimated true effect after correction for publication bias and other small-study effects, and the statistical significance of the slope indicates the presence of publication bias (Carter et al., 2019). The analysis is considered "conditional" because the statistical significance of the PET estimate of the intercept is used to determine whether to use the PET or PEESE estimates.

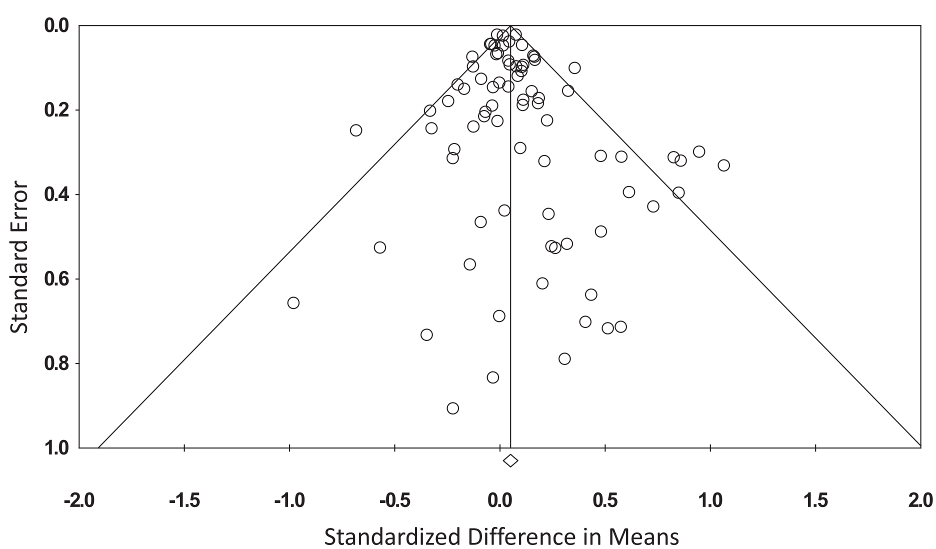
The PET estimate of the the intercept was not statistically significant. Therefore, the PET estimate is the appropriate effect size estimate from the conditional PET-PEESE analysis (Stanley & Doucouliagos, 2014). (The results of the PEESE analysis are provided in the Supplemental Materials.) The PET analysis revealed a statistically significant slope, $b = .42$, 95% CI [.07, .76], $SE = .18$, $p = .019$, indicating the presence of publication bias. Similar to the results of the trim and fill, the PET estimate of the true effect after correcting for publication bias and other small-study effects was nonsignificant, $\bar{d} = 0.01$, see Table 8.

Publication Bias Analysis Summary. As each publication bias analysis technique has its own strengths and limitations, it is important to use multiple techniques to ascertain whether there is converging evidence. All three tests indicated that the mindset intervention literature contains significant publication bias. Two of these tests estimate a bias-corrected meta-analytic mean effect size: These estimates were 0.03 and 0.01, respectively, both of which were not significantly different from zero. Taken together, the overall uncorrected effect size found in the present meta-analysis is likely overestimated.

Discussion

The overall effect of growth-mindset interventions in this model, if not accounting for publication bias, was small and statistically significant: $\bar{d} = 0.05$. Yeager and Dweck (2020) suggest that the largest effect one can expect from an educational intervention on

Figure 6
Funnel Plot of Standard Error by Standardized Difference in Means, Meta-Analysis 1: No Quality Control



real-world outcomes is $d = 0.20$. “If psychological interventions can get a meaningful chunk of a .20 effect size on real-world outcomes in targeted groups, reliably, cost-effectively, and at scale, that is impressive” (Yeager & Dweck, 2020, p. 13).

How impressive is an effect size of 0.05? One way to interpret an effect size is to examine the distribution of scores between the treatment and control groups. The closer an effect size is to zero, the more the two samples’ distributions of scores will overlap. Cohen’s $d = 0.05$ means that 98% of the treatment and control group’s academic achievement scores overlap. When considering the publication bias-corrected estimates, $\bar{d} = 0.03$ and $\bar{d} = 0.01$, there is a 99%–100% overlap between the treatment and control groups’ academic achievement distributions.

Another way to interpret effect sizes is to consider the probability that a randomly selected student from the treatment group will outperform a randomly selected student from the control group. The closer an effect size is to zero, the closer the probability is to mere chance (i.e., 50/50). Cohen’s $d = 0.05$ means there is a 51.4/48.6 chance that a randomly selected student from the treatment group will have a higher score on the measure of academic achievement than a randomly selected student from the control group (Magnusson, 2020). When considering the publication bias-corrected estimates, $\bar{d} = 0.03$ and $\bar{d} = 0.01$, there is a 50.8/49.2–50.3/49.7 chance that a randomly selected student from the treatment group will have a higher score on the measure of academic achievement than a randomly selected student from the control group (Magnusson, 2020).

We investigated whether factors fundamental to mindset theory moderated the effect of growth mindset interventions. No theoretically meaningful moderators were significant. Some theoretically predicted subgroups had significant effects, but because the moderator was nonsignificant, it is inappropriate to conclude that the intervention is more important for certain subgroups than others. For example, the effect was significant for low-SES students and nonsignificant for not low-SES students. However, the two groups’ effects did not differ significantly from each other. We, therefore, cannot conclude that growth mindset interventions are more important for low-SES students than not low-SES students. As another example, the effect was significant for academically struggling (i.e., high challenge) students, but was nonsignificant both for students facing situational challenges and for students facing minimal challenges. This finding runs counter to a major claim of mindset theory—that a growth mindset is especially important for students facing situational challenges (e.g., Rattan et al., 2015; Yeager & Dweck, 2012). Furthermore, the effects across levels of academic challenge did not differ. Therefore, we cannot conclude that students’ level of academic challenge is an important factor for determining the efficacy of a growth mindset intervention.

Aside from the absence of significant moderator analyses that would have supported theoretical claims and the presence of indicators of publication bias, there are other reasons to doubt the credibility of growth mindset interventions having an effect on academic achievement. In many of the studies included in this meta-analysis, there was poor adherence to best practices (see Table 2 and Figure 4). Thus, the results of Meta-Analysis 1 could be influenced by expectancy effects, confounds, or selective reporting of results.

Critically, 71 of the 96 effect sizes included in Meta-Analysis 1 were from interventions with no measurable impact on students’ mindsets. Growth mindset interventions are hypothesized to impact academic achievement by instilling a growth mindset in students. If we assume measures of mindset are valid and reliable, a better test of the importance of growth mindset is to only meta-analyze studies with evidence the intervention influenced students’ mindsets as intended.

Meta-Analysis 2: Minimal Standard of Evidence

Meta-Analysis 2 is identical to Meta-Analysis 1, except it only includes studies that demonstrated that the growth mindset intervention influenced students to have more of a growth mindset. Thus, assuming measures of mindset accurately reflect students’ beliefs, studies included in Meta-Analysis 2 meet the minimal standard of evidence to attribute effects to growth mindset.

Method

Ideally, researchers should test students’ mindsets before and after the intervention and provide evidence that students who received the growth mindset intervention shifted to greater growth mindset beliefs following the intervention, while students in the control group had no change in beliefs. Few studies met this criterion. Our criterion was more lenient: For a study to be included in this meta-analysis, the authors needed to provide evidence that students who received the growth mindset intervention shifted to greater growth mindset beliefs following the intervention relative to before the intervention (i.e., significant pre- to postintervention difference within the treatment group).

Posttest differences between treatment and control groups do not provide evidence that the growth mindset intervention influenced students’ mindsets because the groups could have differed at baseline or the control group could have become more fixed. Likewise, quizzes to determine if students read the materials do not establish that the growth mindset intervention influenced students’ beliefs. Only 25 of the 96 effect sizes were from studies that provided evidence that the growth mindset intervention influenced treatment students to have more of a growth mindset. See Table 9.

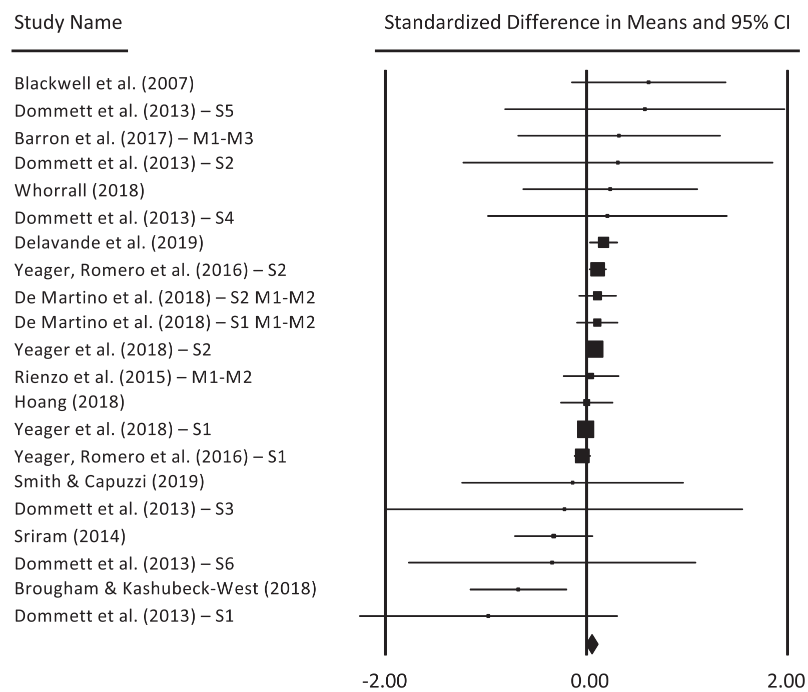
Table 9
Assessing Whether the Intervention Changed Students’ Mindsets

Methodological/Reporting characteristic	Number of effects	Number of students
Did the study test whether the growth mindset intervention changed treatment students’ mindsets?		
No report of this test	37	65,486
Tested, but results not reported[a]	10	9,053
Yes	49	23,133
Of the 49 effects with reported results, did the intervention influence treatment students’ mindsets as intended (significant pre–post change toward growth mindset)?		
No	24	4,778
Yes	25	18,355

[a] Six effects (5,483 students) were from studies that did not report mindset change results. Four effects (3,570 students) were from studies where whether the treatment students’ mindsets changed could not be determined from the results reported (no pre–post difference results).

Figure 7

Each Sample's Effect Size and 95% Confidence Interval in Meta-Analysis 2: Minimal Standard of Evidence



Note. Square size is proportionate to the effect's weight (larger samples contribute more weight) with a minimum size imposed for visibility. The diamond on the bottom row represents the meta-analytically weighted mean Cohen's *d*. For studies with multiple independent samples, the result for each sample (S1, S2, etc.) is reported separately. Multiple measures resulting from a single sample were combined and adjusted for dependency (i.e., M1–M2, M1–M3).

Results

Overall Results. For studies in which the intervention altered students' mindsets as intended, the overall meta-analytic average standardized mean difference between growth mindset treatment and control groups in academic achievement was nonsignificant, $\bar{d} = 0.04$, 95% CI = [−0.01, 0.10], $p = .146$. See Figure 7.

Moderator Analyses. The between-study variability in effect sizes due to heterogeneity rather than random error was moderate, $I^2 = 39.45$ ($\tau^2 = .004$). We investigated the source of this heterogeneity through moderator analyses. We conducted moderator analyses when there were at least five effect sizes contributing to a subgroup (Williams, 2012). The relatively small number of studies that met the minimal standard of evidence limited the moderator analyses we could conduct. See Tables 10 and 11 for the number of effect sizes and sample sizes contributing to each subgroup and Table 12 for the number of effect sizes and sample sizes associated with the articles' characteristics.

Mindset theory holds that growth mindset interventions will be especially important under certain circumstances. Yet, no theoretically meaningful moderators were significant, indicating that the effect sizes did not differ significantly across levels of the moderator. See Table 13 for the results of the moderator analyses.

Bias Analyses

Financial Incentives. It has been suggested that mindset proponents may be more thoughtful, careful, or skilled at creating the

Table 10

Student Characteristics of Included Studies' Samples, Meta-Analysis 2: Minimal Standard of Evidence

Sample characteristic	Number of effect sizes	Number of students
Developmental stage		
Adults	4	1,064
Adolescents	19	17,110
Children	2	181
Academic challenge		
High	8 (10)	8,563 (8,619)
Moderate	11	9,307
Low	6 (4)	485 (304)
Socioeconomic status		
Not low	7 (5)	16,852 (16,671)
Low	4 (6)	563 (635)
Not reported	14	940

Note. All samples in this model clearly fit one of the classifications for each sample characteristic variable. Numbers in parentheses represent the numbers when replacing whole samples' effects with relevant subsamples. For example, a study's reported high-challenge subsample replacing a low-challenge whole sample, increases the high-challenge effect size count and number of students while decreasing the low-challenge effect size count and number of students.

Table 11
Characteristics of Included Studies' Intervention Methods and Outcome Measures, Meta-Analysis 2: Minimal Standard of Evidence

Intervention/Outcome measure characteristic	Number of effect sizes	Number of students
Intervention type		
Interactive	25	18,355
Feedback	0	0
Passive	0	0
Mode of intervention		
Combination in-person and other mode	1	20
In-person	8	545
By teacher(s)	3	132
By researcher(s)	3	177
By teacher who is also the researcher	1	75
By both teachers and researchers	0	0
Other administrator(s)	2	181
Computer program	16	17,790
Reading materials	0	0
Mindset type		
Intelligence	24	18,146
Other (e.g., math ability, personality)	0	0
Intelligence + other	1	209
Number of sessions		
One	6	1,194
More than one	19	17,161
Intervention context		
Outside regular classroom activities	21	18,051
Integrated into classroom activities	4	304
Time interval to outcome measure[a]		
Immediate (end of last session)	0	0
Short (within 4 months)	9	4,805
Long (longer than 4 months)	15 (15)	13,816 (13,789)
Unclear	1	75
Academic achievement measure[a]		
Standardized test	11	990
Laboratory measure	6	241
Course grade average (e.g., GPA)	9	17,086
Course grade	5	847
Course exam grade	0	0

Note. GPA = grade point average.
[a] The cumulative sample size is greater than 18,355 due to multiple measures from the same sample in different categories. Numbers in parentheses represent the numbers when replacing the effect associated with the longest time interval within the same academic context (here, all long intervals) with the effect associated with the longest time interval regardless of context.

best circumstances for growth mindset interventions to impact academic achievement than researchers who are not active proponents of mindset theory. As Dweck (as cited in Chivers, 2017) stated when confronted with a failed replication attempt:

> Not anyone can do a replication. We put so much thought into creating an environment; we spend hours and days on each question, on creating a context in which the phenomenon could plausibly emerge. Replication is very important, but they have to be genuine replications and thoughtful replications done by skilled people. Very few studies will replicate done by an amateur in a willy-nilly way. (paras. 21–22)

Table 12
Characteristics of Included Studies' Articles, Meta-Analysis 2: Minimal Standard of Evidence

Associated characteristic of article	Number of effect sizes	Number of students
Financial incentive to find positive effects		
One or more authors	8	16,309
No authors	17	2,046
Publication status[a]		
Published	12	4,020
Unpublished	13	14,335
Combinations of the above two factors		
Financial incentive + Published	3	3,538
No financial incentive + Published	9	482
Financial incentive + Unpublished	5	12,771
No financial incentive + Unpublished	8	1,564

[a] In cases where a published and unpublished report of the same study were accessible, we used the published version if both reports included the same number of participants. If study participants' data were excluded, we used the version with more study participants.

If this is the case, authors with financial incentives to find positive effects, who we consider mindset proponents, should be better at creating a context that influences students' mindsets as intended. Yet, authors with financial incentives were no more likely to find evidence that their growth mindset interventions changed students' mindsets than authors without financial incentives, $\chi^2(1, N = 61) = 0.13$, $p = .911$. The average growth mindset treatment effect size of studies where one or more authors had a financial incentive to find positive results was not significantly different than the average growth mindset treatment effect size where no authors had a perceived financial conflict of interest, see Table 14.

Financial Incentives in the Published Literature. In line with the observations that authors with financial incentives appear less likely to publish weak results and that effects where the intervention measurably influenced mindsets were generally weak, there were not enough published studies by authors with financial incentives in this meta-analysis to conduct this moderator analysis. For published studies where no authors had a financial incentive to find positive results, the effect was significant, but negative, suggesting growth mindset interventions that changed students' mindsets harmed students' academic achievement, see Table 14.

Publication Bias Analyses. No publication bias analyses suggested that this model's effect was impacted by publication bias, see Table 14.

Discussion

When limiting the analysis to studies in which the growth mindset intervention measurably shifted students' mindsets, there was no significant growth mindset intervention effect on academic achievement. That said, the quality control in this meta-analysis is minimal and inadequate study design (e.g., multiple differences between treatment and control groups) and flawed reporting (e.g., removing subgroups) could still influence outcomes. To better understand the potential importance of growth mindset we should focus on the highest quality studies providing the best available evidence. We did this in the final meta-analytic study.

Table 13
Moderator Results, Meta-Analysis 2: Minimal Standard of Evidence

Moderator and levels	Result		
	Theoretical factors		
Developmental stage		NA	
Adults	—	—	—
Adolescents	$\bar{d} = 0.04$	95% CI [−0.02, 0.10]	$p = .225$
Children	—	—	—
Academic challenge status[a]		$Q(2) = 2.98$, $p = .226$	
High-challenge level (e.g., low grades)	$\bar{d} = 0.09$	95% CI [0.04, 0.14]	$p < .001$
Moderate-challenge level (e.g., new school)	$\bar{d} = -0.01$	95% CI [−0.10, 0.09]	$p = .909$
Low-challenge level	$\bar{d} = 0.05$	95% CI [−0.21, 0.32]	$p = .690$
Socioeconomic status[b]		$Q(1) = 1.01$, $p = .314$	
Middle–high	$\bar{d} = 0.05$	95% CI [−0.01, 0.11]	$p = .119$
Low	$\bar{d} = 0.13$	95% CI [−0.01, 0.27]	$p = .066$
Time interval to outcome measure[c]		$Q(1) = 0.02$, $p = .896$	
Short (interval ≤4 months)	$\bar{d} = 0.03$	95% CI [−0.10, 0.16]	$p = .630$
Long (interval >4 months)	$\bar{d} = 0.04$	95% CI [−0.02, 0.10]	$p = .170$
	Methodological factors		
Intervention type		NA	
Interactive (e.g., "saying-is-believing" task)	$\bar{d} = 0.04$	95% CI [−0.01, 0.10]	$p = .146$
Passive (e.g., only reading materials)	—	—	—
Number of sessions		$Q(1) = 0.01$, $p = .916$	
Slope	$b = -0.002$	95% CI [−0.03, 0.03]	$p = .916$
Intervention delivery mode		$Q(1) = 0.16$, $p = .686$	
Reading material	—	—	—
Computer program	$\bar{d} = 0.05$	95% CI [−0.01, 0.10]	$p = .088$
In person	$\bar{d} = -0.03$	95% CI [−0.42, 0.35]	$p = .867$
Combination of delivery modes	—	—	—
Administrator (of in-person delivery)		NA	
Teacher	—	—	—
Researcher	—	—	—
Teacher who is also the researcher	—	—	—
Other	—	—	—
Context		NA	
In the classroom	—	—	—
Outside the classroom	$\bar{d} = 0.04$	95% CI [−0.02, 0.10]	$p = .169$
Academic achievement measure		$Q(2) = 3.98$, $p = .137$	
Course exam grade	—	—	—
Single course grade	$\bar{d} = 0.18$	95% CI [0.03, 0.32]	$p = .018$
Multicourse grade average (e.g., GPA)	$\bar{d} = 0.03$	95% CI [−0.05, 0.10]	$p = .453$
Standardized test score	$\bar{d} = -0.02$	95% CI [−0.19, 0.14]	$p = .765$
Laboratory vs. actual standardized test		$Q(1) = 0.04$, $p = .838$	
Laboratory-based standardized test	$\bar{d} = -0.08$	95% CI [−0.66, 0.49]	$p = .780$
Actual standardized test score	$\bar{d} = -0.02$	95% CI [−0.19, 0.15]	$p = .819$

Note. — = not enough effects available to include in analysis. GPA = grade point average.
[a] One study was available that provided information for a high-challenge subsample. The results do not change when replacing the whole sample with this subsample, although with this replacement there were no longer enough low-challenge samples to include in the moderator analysis. [b] Studies not reporting student-level socioeconomic status were not included in this moderator analysis. Not enough low-SES samples were available for moderation analysis, unless we replaced whole samples with available low-SES subsamples. The results in the table reflect results with low-SES subsamples replacements. [c] For six samples, a longer time interval was available beyond the academic context in which the intervention was administered. When replacing effects with those from longer time intervals, the pattern of results is unchanged.

Meta-Analysis 3: Best Available Evidence

Meta-Analysis 3 provides the best available evidence to evaluate the impact of growth mindset interventions on academic achievement. Our goal was to include studies that provided evidence the intervention influenced students' mindsets and met all the best practices criteria:

1. The treatment group was compared with an active control group.
2. Only the critical ingredient differed between the treatment and control groups.
3. An a priori power analysis was conducted.
4. Individual students were randomly assigned to condition.
5. Students, administrators, and teachers were blind to condition and hypotheses.
6. A test of whether the treatment influenced students' mindsets was conducted.
7. Hypotheses, methods, and planned analyses were preregistered.
8. Results of the participants who participated were reported.

Table 14
Bias Analyses Results, Meta-Analysis 2: Minimal Standard of Evidence

Financial incentives			
All effects		$Q(1) = 0.08$, $p = .781$	
Author(s) with a financial incentive	$\bar{d} = 0.04$	95% CI [−0.02, 0.11]	$p = .222$
No authors with a financial incentive	$\bar{d} = 0.02$	95% CI [−0.10, 0.14]	$p = .731$
Published effects		NA	
Authors(s) with a financial incentive	—	—	—
No authors with a financial incentive	$\bar{d} = -0.37$	95% CI [−0.63, −0.10]	$p = .007$
Test of missing studies due to publication bias			
Egger's regression	$B_0 = -0.09$, $t(19) = 0.25$, $p = .402$		
Publication bias-corrected estimates of the overall effect			
Trim and fill	$\bar{d} = 0.04$	95% CI [−0.01, 0.10]	$p = .146$
Precision effect test (PET)	$\bar{d} = 0.05$	95% CI [−0.02, 0.13]	$p = .137$

9. Results for the whole sample (or all subsamples) were reported.
10. No authors had a financial incentive to find positive effects.

There were no studies that met all the best practices criteria. Our plan in this case was to lower the proportion of best practices criteria required to be met until at least five studies could be included in the meta-analysis. Zero studies that measurably changed students' mindsets met 100% or 90% best practices criteria, one study met 80% best practices criteria, two studies met 70% best practices criteria, and three met 60% best practices criteria. All other studies that measurably influenced students' mindsets met ≤50% best practices. This process, therefore, resulted in lowering the threshold proportion of best practices met to 60%, yielding six studies with eight samples and 10 effects ($N = 13{,}571$). Thus, this model is not free from flaws or biases, but represents the best evidence currently available to evaluate the effect and mechanism of growth mindset interventions on academic achievement.

These eight samples and their interventions are generally representative of circumstances in which mindset interventions are hypothesized to be especially helpful. Four samples consisted primarily of students experiencing a high degree of academic challenge, three consisted of students facing situational challenges (transitioning to a new school), and only one sample had a low-challenge level. Two of the five samples with reported socioeconomic status information were low SES. Five of the eight samples consisted of adolescents. Six of the 10 effects measured academic achievement after a long period following the intervention. All samples' treatment groups were exposed to an interactive (e.g., saying-is-believing) intervention. All samples demonstrated that the growth mindset intervention shifted treatment students' mindsets toward more of a growth mindset.

Method

Coding and Inclusion. Cases were coded as having met or not having met each of the 10 best practices criteria. The proportion of best practices met was then calculated for each sample. If a sample contributed multiple effects with inconsistent application of a best practice, we coded the combined effect as not having met the criterion. We reasoned that inconsistent application of best practices is not following best practices.

Results

When examining the best available evidence, the effect of growth mindset interventions on academic achievement was nonsignificant: $\bar{d} = 0.02$, 95% CI = [−0.06, 0.10], $p = .666$. See Figure 8.

This meta-analysis had substantial heterogeneity, $I^2 = 62.71\%$ ($\tau^2 = .006$). There were not enough high-quality studies available within any subgroups to conduct moderator analyses. No publication bias analyses suggested that this model's effect was impacted by publication bias.

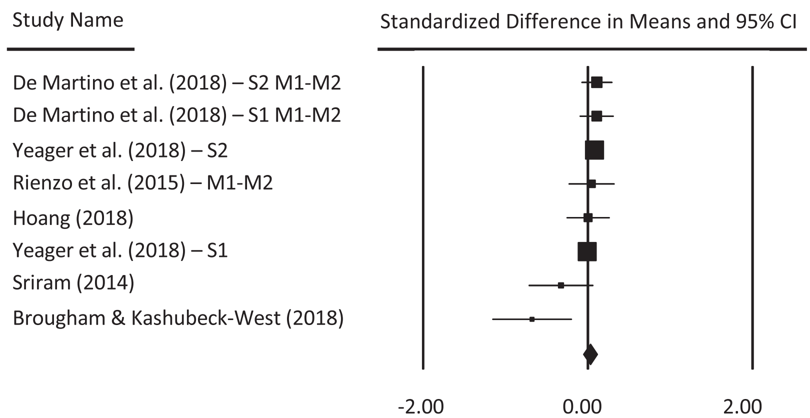
Discussion

The studies included in Meta-Analysis 3 provided evidence the intervention influenced students' mindsets and met at least 60% of the best practices criteria. When examining these interventions, there was no significant growth mindset intervention effect on academic achievement. Although these studies represent the highest-quality evidence currently available, as a whole they still failed to meet multiple best practices criteria, which likely erroneously increased the size of the effect. Therefore, the overall effect size found in this model ($\bar{d} = 0.02$) might still be inflated.

This meta-analysis included studies that met the most best practices criteria regardless of which best practices they met. It could be argued that not all best practices are equally important, and researchers may disagree on which combination of the best practices criteria should be considered in a model examining the highest quality studies. We, therefore, also tested the effect when studies adhered to every possible combination of best practices.

In total, we conducted an additional 230 meta-analytic models of interventions that adhered to every number and combination of best practices criteria, with and without the requirement of a successful manipulation check, when at least five studies were available for the model. As the number of best practices met increased, the number of significant models quickly decreased, such that no models where studies adhered to any combination of three or more best practices were significant, see Table 15. After correcting for publication bias, there were no significant models that adhered to any number or combination of best practices criteria, see Table 15. See Supplemental Materials for a summary of model results. Taken together along with Meta-Analysis 3, our results suggest that apparent effects of growth

Figure 8
Each Sample's Effect Size and 95% Confidence Interval in Meta-Analysis 3: Best Available Evidence



Note. Square size is proportionate to the effect's weight (larger samples contribute more weight) with a minimum size imposed for visibility. The diamond on the bottom row represents the meta-analytically weighted mean Cohen's *d*. For studies with multiple independent samples, the result for each sample (S1, S2) is reported separately. Multiple measures resulting from a single sample were combined and adjusted for dependency (i.e., M1–M2).

mindset interventions may be due to inadequate study designs, reporting flaws, and/or bias.

General Discussion

We conducted a systematic review and multiple meta-analyses of the growth mindset intervention literature. Our goal was to answer two questions: (a) Do growth mindset interventions generally improve students' academic achievement? and (b) Are growth mindset intervention effects due to instilling growth mindsets in students or are apparent effects due to shortcomings in study designs, analyses, and reporting? To answer these questions, we systematically reviewed the literature and conducted multiple meta-analyses imposing varying degrees of quality control. Our results indicated that apparent effects of growth mindset interventions are possibly due to inadequate study designs, reporting flaws, and bias. In particular, the systematic review yielded several concerning patterns of threats to internal validity.

First, almost all growth mindset interventions allow multiple factors to covary with experimental condition. That is, the treatment group is taught that intelligence or another characteristic is malleable (growth mindset) and is additionally given, for example, encouragement to work harder, strategies for taking on challenges, inspirational stories, and/or individualized study plans that the control group does not receive. When multiple characteristics of the protocol differ between groups, the mechanism of any effect is undeterminable. Group differences in achievement could be due to growth mindset; alternatively, it could be that encouragement to work harder caused students to work harder thereby improving achievement while having no impact on their mindsets. Of the few studies that only allowed the key factor of teaching attribute malleability to vary between the groups, none produced evidence the treatment influenced students to have more of a growth mindset.

A second common threat to internal validity found in the literature is potential expectancy bias. Fewer than one in four interventions attempted to reduce expectancy bias by keeping students, study

Table 15
Summary of Number of Meta-Analytic Models for Every Number and Combination of Best Practices Criteria Met, With and Without the Requirement of a Successful Manipulation Check

Best practices criteria met	Possible models	Models with ≥5 studies available	Significant models ($p < .05$)	Significant models after correcting for publication bias
1	20	15	5	0
2	90	44	4	0
3+	1,936	171	0	0

Note. Best practices criteria met: 1 = adhered to a single given best practice, 2 = adhered to a given pair of best practices; 3+ = adhered to any given combination of three or more best practices. Possible models include the total number of every unique combination of best practices for that row's number of best practices (see left-most column), with and without the requirement of a successful manipulation check (e.g., row 1 = 10 possible models where studies adhered to each of the 10 best practices plus another 10 models where studies adhered to each of the 10 best practices and reported a successful manipulation check).

administrators, and teachers blind to condition. Given promises of profound effects of growth mindset interventions, teachers may unconsciously view students in the growth mindset treatment as having made more academic progress than if teachers were blind to condition. Study administrators who expect the growth mindset intervention to be beneficial may deliver the treatment differently than the control protocol. In several studies, the classroom teacher designed the study, delivered the intervention, graded students' academic achievement, and authored the study. Expectations of effects, especially when grading students, may be the source of group differences in academic achievement.

We also observed several issues related to reporting of significant effects. One such issue occurs when studies assign classes or schools to condition rather than individual students, which inflates the Type I error rate (Hox, 1998; McCoach & Adelson, 2010). Adjustments for cluster designs can be applied but often were not, calling into question the significance of some effects reported in the literature.

Another issue associated with reporting significant effects in the growth mindset intervention literature is that some study authors found null effects but described them as though they were significant. This was the case for about 10% of the null effects in our meta-analysis. When study authors discuss null effects as though they were significant, readers may come away with a mistaken understanding of growth mindset intervention effectiveness.

A final issue with reporting significant effects is one of potential bias. We found that, relative to authors with no perceived financial conflict of interest, studies with one or more authors with a financial incentive to report positive effects were more than two and half times as likely to report significant effects. We also found different publication patterns between authors with and without a financial incentive to report positive effects: Authors with financial incentives to report positive effects appeared less likely to publish weaker effects than authors without such a financial incentive (significant Publication status × Financial incentive status interaction).

Additional evidence of publication bias came from uncovering several unpublished reports of studies that were later published. The majority of them (5 of 7) excluded participants prior to publication. The number of participants removed was as high as 6,222 students, half the total sample. Excluding data and removing subsamples prior to publication leaves a skewed view of interventions' effects in the published literature.

Publication bias appeared to be a factor in our meta-analysis as well. The first meta-analytic model we conducted considered all evidence regardless of quality or whether treatments measurably influenced students' mindsets (63 studies, N = 97,672). Growth mindset interventions appeared to lead to a small overall difference between treatment and control groups in academic achievement: $\bar{d}$ = 0.05, 95% CI = [0.02, 0.09], p = .004. Critically, when correcting for publication bias, the effect was nonsignificant, $\bar{d}$ = 0.03, 95% CI [−0.01, 0.07], p = .097, or $\bar{d}$ = 0.01, 95% CI [−0.03, 0.05], p = .667, depending on the bias correction estimate. This result suggests that findings unsupportive of growth mindset hypotheses are suppressed from publication. This publication suppression biases the available evidence to appear more supportive of growth mindset intervention effects on students' academic achievement than may be warranted.

Growth mindset interventions are hypothesized to work because they influence students to have more of a growth mindset, which leads to changes in motivations and behaviors resulting in improved academic achievement. Therefore, evidence the intervention influenced students to have more of a growth mindset is needed to attribute the underlying intervention mechanism to instilling a growth mindset. We conducted a second meta-analytic model of studies providing evidence the intervention influenced students to have more of a growth mindset. Only 26% of the effect sizes came from studies that met this minimal standard of evidence. When examining these cases (13 studies, N = 18,355), the effect of growth mindset interventions on academic achievement was nonsignificant: $\bar{d}$ = 0.04, 95% CI = [−0.01, 0.10], p = .146.

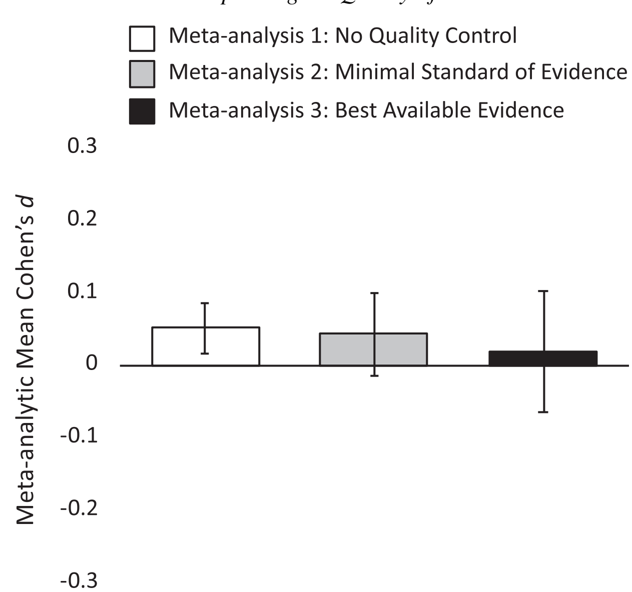
We next evaluated the highest-quality evidence—studies that demonstrated the intervention influenced students to have more of a growth mindset and adhered to the most best practices criteria. No studies adhered to all of the best practices criteria. Only 10% of the effects came from studies that both provided evidence the intervention changed students' mindsets and met more than half the best practices criteria. When examining these cases (10 effects from 6 studies, N = 13,571), the effect of growth mindset interventions on academic achievement was nonsignificant: $\bar{d}$ = 0.02, 95% CI = [−0.06, 0.10], p = .666.

See Table 16 for descriptive statistics and Figure 9 for overall growth mindset intervention results of the three main meta-analytic models.

Table 16
Descriptive Statistics for Each Meta-Analytic Model

Model	Number of effect sizes (number of samples)	Number of students
Model 1: No quality control	96 (79)	97,672
Model 2: Minimal standard of evidence	25 (21)	18,355
Model 3: Best available evidence	10 (8)	13,571

Figure 9
Overall Differences Between Treatment and Control Groups' Academic Achievement Depending on Quality of Studies



Note. Error bars represent 95% confidence intervals.

We also conducted additional meta-analytic models examining effects of studies that adhered to every number and combination of best practices criteria when enough studies were available. As the number of best practices met increased, the number of significant models decreased. After correcting for publication bias, there were no significant models that adhered to any number of any combination of best practices criteria.

Taken together, our findings indicate that studies adhering to best practices are unlikely to demonstrate that growth mindset interventions benefit students' academic achievement. Instead, significant meta-analytic results only occurred when quality control was lacking, and these results were no longer significant after adjusting for publication bias. This pattern suggests that apparent effects of growth mindset interventions on academic achievement are likely spurious and due to inadequate study design, flawed reporting, and bias.

Implications

Should parents and schools allocate time and money to growth mindset interventions when they do not appear to influence academic achievement? One argument for continuing to promote growth mindset interventions in schools is that, even if the effects do not flow all the way downstream to academic achievement, growth mindsets may promote other benefits such as motivation to try harder, a greater focus on learning, and embracing mistakes and challenges. If this is the case, mindset theorists should first ensure these associations are not artifacts of measurement issues (see, e.g., Burgoyne et al., 2018; Limeri et al., 2020), and then restrict the scope of their claims to these outcomes. Additionally, researchers should attempt to uncover why these motivational and behavioral changes resulting from a growth mindset do not impact academic achievement as hypothesized.

That said, the argument that growth mindsets lead to meaningful changes in motivation and behavior is not well supported. In a meta-analysis of measured mindsets and motivations, Burnette et al. (2013) generally found weak associations. Likewise, in a preregistered, large-sample, direct test of mindset's associations with several motivations and behaviors assumed to originate from mindset (e.g., goal orientation, challenge seeking, resilience), we (Burgoyne et al., 2020) found weak and null relationships. The strongest association was counter to mindset theory. We concluded, "Our results suggest that the foundations of mind-set theory are not firm and, in turn, call into question many assumptions made about the importance of mind-set" (p. 266).

A goal of future research should be to meta-analyze relationships between mindsets, motivations, and behaviors considering studies' adherences to best practices. Our research indicates that treating all studies as having the same quality and ability to interpret mechanisms, and not accounting for bias, may lead to misleading conclusions. By examining both the quantity and quality of evidence, researchers should be able to assess whether (a) true relationships exist between mindsets and motivations and behaviors, or whether (b) apparent relationships are due to measurement problems, inadequate study design, reporting flaws, and/or bias, as appears to be the case in the growth mindset intervention literature.

Another argument for continuing to promote growth mindset interventions is that these interventions interact with recursive processes (Yeager & Walton, 2011), producing enduring changes that compound benefits over time (Dweck et al., 2014). That is, even if studies do not demonstrate effects in the short term, there may be long-term benefits of growth mindset interventions. We know of no evidence to support this claim. Our moderator analysis testing whether the treatment effect varied by the length of time between the intervention and the measure of academic achievement was nonsignificant. Furthermore, most studies collect postintervention measures of mindset within a year of administering a growth mindset intervention. To our knowledge, no studies have tested whether growth mindset intervention benefits are compounded over longer timescales (i.e., over years).

A final argument for continuing to promote growth mindset interventions is that mindset interacts with the level of challenge such that interventions are most effective for struggling students, even if there is no effect on achievement on average (see, e.g., Yeager et al., 2019). The evidence for this claim is limited. Our moderator analysis testing whether the treatment effect varied by level of challenge was nonsignificant. Additionally, as described in the State of the Literature section, claims about growth mindset interventions being particularly beneficial for particular student subgroups are often not properly tested. If researchers hypothesize that an intervention will benefit one group more than another, they should test whether intervention effectiveness significantly differs by group. A goal of future research may be to substantiate these claims using adequate study design and analyses.

It is perhaps surprising that after nearly 40 years of mindset research (Dweck first introduced the theory in 1983; see Dweck & Bempechat, 1983) with millions of dollars in funding, the theoretical underpinnings of mindset remain largely unsubstantiated. Researchers should endeavor to determine whether mindset's theoretical premises are accurate or identify the circumstances under which they are accurate before claiming that growth mindsets benefit students.

Limitations

This study is the most comprehensive systematic review of the empirical evidence of growth mindset interventions on academic achievement to date, assessing both the quantity and quality of the evidence. However, there are several limitations worth mentioning.

All meta-analyses are limited by their timeframes and the studies that were available during the literature search. If the number of growth mindset interventions continues to increase, there may be many more studies available at the time of publication than were available when we stopped our search. There is no reason to believe that these new studies would substantially change the overall effect sizes, but it is a possibility. Likewise, we were limited in the studies we could access. We were aware of studies we could not access—when researchers had unpublished studies they declined to provide us the details of—and our trim-and-fill analysis estimated we were missing 10 studies with smaller-than-average effect sizes.

We also were limited in our ability to assess documents in languages other than English. Mindset theory developed in the U.S. and the majority of the mindset research has been conducted in the U.S. with English-speaking students. Nonetheless, mindset interventions have been conducted around the world. Of the 4,832 independent records revealed in our search, 138 were in a language other than English. Approximately 1% of the total searched records were relevant to the present meta-analysis. Based on this percentage, we would expect one additional study to have been included if we were able to evaluate articles written in all languages. This estimate, however, is influenced by factors that bias this estimate higher and other factors that bias this estimate lower. Many studies conducted in countries where the dominant language is not English nevertheless write the study report in English in an effort to publish in English-language journals (11% of the studies in the current meta-analysis). Thus, all relevant studies may be included. Alternatively, we might be missing more than one otherwise relevant study because non-English articles were likely underrepresented in our search given the databases we used (e.g., APA PsycInfo), though Google Scholar searches across languages. There is no known evidence suggesting there are many non-English articles relevant to the current meta-analysis or that such studies would produce very different results from the studies currently included in the present meta-analysis. Regardless, it is possible that if we had been able to include more languages in our search that this could have affected the results of our meta-analysis.

We made (preregistered) decisions about the scope of this meta-analysis. For example, we only included growth mindset interventions that taught growth mindset of a human attribute. Broadening this scope to include growth mindsets of stress, belonging, and/or willpower might have produced smaller or larger effect sizes. Narrowing this scope to only include a particular human attribute (e.g., math ability) likewise might have produced smaller or larger effects. Similarly, we evaluated intervention and control content for meaningful and clearly codable categories in the moderator analysis. More nuanced evaluation of features (e.g., "saying-is-believing" activity vs. teacher-led discussion vs. gamified interactive activities) may have provided more insight into which intervention features are more or less impactful. We note that such features are not mutually exclusive, which would make coding these characteristics for moderator analyses challenging. A more fruitful approach for comparing nuanced features of an intervention may be to conduct experiments manipulating these features.

We based Meta-Analysis 2: Minimal Standard of Evidence and Meta-Analysis 3: Best Available Evidence on the assumption that measures of mindset collected before and after an intervention are valid and reliable. Recent research (Burgoyne et al., 2018; Limeri et al., 2020) has questioned this assumption. If mindset measures do not accurately reflect students' mindsets, then restricting models to only include studies with successful manipulation checks might skew the results. To account for this possibility, we conducted meta-analytic models for every number and combination of best practices criteria met when at least five studies were available, with and without the requirement that the included studies demonstrated a successful manipulation check. No model, regardless of this requirement, revealed a significant growth mindset intervention effect after correcting for likely publication bias.

We used Cohen's d because it is more commonly used in the literature outside of meta-analyses. Meta-analysts often use Hedges' g, which corrects for the slight overestimation of effect sizes Cohen's d produces for small samples, especially for very small samples (e.g., $df < 10$; Borenstein et al., 2009; Hedges, 1981). Our meta-analysis included one very small sample (Fleurizard & Young, 2018, $N = 10$). Thus, overall effect sizes we observed are slightly upwardly biased.

A further limitation is that we used Comprehensive Meta-Analysis software and therefore cannot publish our code. We provide open data (https://osf.io/ajhxv/) that includes study characteristics, effect sizes, and variances, along with information and explanations for which studies were included in each model. Thus, researchers should be able to reproduce all our results.

We observed heterogeneity—differences in effects across studies—in each of our meta-analyses. As heterogeneity increases, the variance of the true effect sizes (τ^2) increases, meaning the overall (pooled) true effect size might be higher or lower than what our models revealed. Heterogeneity is captured in the 95% confidence intervals around our reported effect sizes. See Figure 9.

Likely the biggest limitation of our review comes from the level of quality of the included studies. Insufficient quality of studies poses considerable barriers to scientific understanding. Likewise, conducting a meta-analysis without considering study quality can limit the conclusions that can be drawn from synthesizing a body of research. To better estimate the true effect, we constructed multiple meta-analytic models that differed by the quality of the included studies. We found that even when considering the best evidence available (Meta-Analysis 3), we had to considerably lower quality standards to have enough studies to meta-analyze.

Recommendations

To improve our scientific understanding and ability to make well-informed decisions, we present three recommendations for researchers and educators considering implementing a growth mindset intervention.

Improving Reports of Growth Mindset Interventions

In this article, we identified multiple best practices and other interpretation considerations in intervention design, analysis, and reporting, most of which are applicable to any intervention. In addition to adhering to these best practices and study designs to provide evidence for the treatment mechanism, we recommend that researchers produce transparent study reports where readers can (a) assess the size of the effect, (b) see the variance in the groups, (c) understand the analyses conducted (including the weighting scheme of covariates and whether the effect differs without covariates), and (d) rely on appropriate interpretations of the results from the authors.

Researchers should also strive to reduce potential bias in growth mindset intervention reports. Researchers should acknowledge perceived financial conflicts of interest including employment in or consulting for entities that promote growth mindset products or services, royalties from self-help books recommending growth mindset, and when one serves as a speaker-for-hire about the benefits of growth mindset. To reduce publication bias, researchers should publish the results from all growth mindset intervention studies regardless of their outcome. At a minimum, study results remaining unpublished should be made publicly available on a server such as PsyArXiv.

Avoiding Hype

We recommend that researchers avoid overhyping their results in future studies. Many claims about the benefits of growth mindset interventions outweigh the evidence (Ritchie, 2020). Big claims appear in popular press self-help books, commercial growth mindset intervention marketing, and op-eds where they reach a large number of laypeople and educators.

Growth mindset hype also appears in the scientific literature, making the theoretical underpinnings of growth mindset interventions appear more robust than the evidence suggests. For example, an article published in *American Psychologist* titled, "Mindsets and human nature: Promoting change in the Middle East, the schoolyard, the racial divide, and willpower" claimed "an emphasis on growth not only increases intellectual achievement but can also advance conflict resolution between long-standing adversaries, decrease even chronic aggression, foster cross-race relations, and enhance willpower" (Dweck, 2012, p. 614). As another example, in a highly cited publication, readers are told how interventions can seem "magical" (Yeager & Walton, 2011, p. 267). While the authors assure readers interventions are not magic, they emphasize that interventions "can lead to large gains in student achievement and sharply reduce achievement gaps even months and years later" (p. 267). Such claims, whether in the scientific literature or the popular press, should be tempered until well-conducted studies consistently show these effects.

Weighing Potential Benefits and Harms Before Implementing Interventions

Educators and policymakers planning to implement growth mindset interventions should weigh the potential benefits of doing so against the potential harms. First, interventions require time and resources that could be allocated elsewhere. Proponents of growth mindset state that growth mindset interventions are inexpensive relative to other interventions and can be brief (e.g., Dweck, 2018). Yet, if these interventions are ineffective, the money and time spent on them, however small, might be better spent on other programs or on enhanced learning opportunities (e.g., new textbooks, access to databases, tutoring services) that may better serve students.

Further, when attempting to foster growth mindsets in students, misunderstandings and misapplications that potentially harm students are common. These misunderstandings and misapplications occur so frequently that Dweck updated her 2006 popular press book in 2016 to describe these issues as what she calls "the false growth mindset"—for example, overly focusing on praising effort, absolving oneself from teaching struggling students, and blaming students all in the name of growth mindset. Relatedly, at least some types of growth mindset interventions have demonstrated they increase participants' propensity to blame themselves and others, which decreases participants' mental health and increases their feelings of anger, prejudice, and intolerance of others (Hoyt et al., 2017; Hoyt & Burnette, 2020). Finally, some equity-focused educators have raised concerns that growth mindset interventions implemented to reduce outcome inequalities are potentially harming the students they seek to benefit: Efforts to change students who are targeted for systemic oppression redirects efforts away from changing the inequitable systems in schools that disenfranchise these students in the first place (Gorski, 2011, 2017, 2019; Russell, 2019; Thomas, 2016, 2018).

Policymakers should consider what types of additional teacher training may be necessary to avoid unintentionally harming students, keeping in mind that (a) additional training requires more resources than the intervention might originally appear to cost, (b) it is unclear how to reduce harms resulting from growth mindset interventions because this topic is underresearched, and (c) time, effort, and money spent on growth mindset interventions and related training are resources not being spent on other, potentially more effective and equitable, endeavors.

Future Directions in Evaluating Evidence

A skewed perception of effects due to hype, bias, and problematic study designs makes it difficult for researchers, educators, and policymakers to accurately evaluate the evidence and—in the case of educators and policymakers—to make well-informed decisions about implementation. To aid educators and policymakers in making informed decisions about "what works" in education, the What Works Clearinghouse was established by the Department of Education in 2002 to review educational programs, products, practices, and policies.[5]

The What Works Clearinghouse uses a number of best practices (e.g., whether the study used individual-level random assignment to condition) to evaluate the scientific evidence for educational interventions.[6] Nevertheless, the What Works Clearinghouse has been criticized for ignoring other important best practices, thereby undermining confidence in their evaluations (see Simons et al., 2016). For example, the What Works Clearinghouse does not score double-blind studies higher than studies without blinding. Passive controls (e.g., teaching-as-usual) are not scored lower than active controls. Likewise, differences in encouragement, expectations, additional skills taught, inspirational stories, or other messages between treatment and control groups besides the critical ingredient are not recognized as confounds. Finally, though the What Works Clearinghouse (2020) has a conflict of interest policy, developers of intervention products (e.g., companies selling interventions) are exempt from this policy.

For the reasons just described, educators and policymakers should not have high confidence in the What Works Clearinghouse study evaluations. Thus, unfortunately, there is currently no central, trusted source to know what works. The What Works Clearinghouse should address these shortcomings. Until then, we recommend educators and policymakers seeking to evaluate interventions assess the study reports for adherence to best practices.

Conclusions

Mindset theory and growth mindset interventions are appealing because they claim to provide the answer to how to succeed and fulfill one's potential (Dweck, 2006, 2016). Growth mindset interventions have received substantial attention from researchers, governments, funding agencies, the media, educators, and entrepreneurs. Multiple companies and organizations sell growth mindset intervention products or provide growth mindset coaching to individuals and businesses. Calls have been made to make growth mindset interventions a national funding priority (Rattan et al., 2015). The White House even convened a special meeting in 2013 on mindsets in education (Shankar & Kalil, 2013). These drives for growth mindset practices in schools have been summarized as "the mindset revolution that is reshaping education" (Boaler, 2013, p. 143).

[5] https://ies.ed.gov/ncee/wwc/WhoWeAre (retrieved September 17th, 2020).

[6] https://ies.ed.gov/ncee/wwc/ (retrieved September 17th, 2020).

Despite mindset's popularity, no systematic review evaluating the quality and quantity of growth mindset interventions has previously been conducted. The first meta-analysis of growth mindset interventions in 2018 (which did not account for study quality) elicited conflicting conclusions: The meta-analysis authors (Sisk et al., 2018) concluded that effects of growth mindset interventions on academic achievement were small and cautioned they may be due to extraneous factors; Dweck (2018), the developer of mindset theory, concluded that the same effects were substantive and supported mindset theory. To help resolve this conflict, we reexamined the literature, this time providing a systematic review of best practices and varying the quality control between models to better evaluate the evidence.

We observed that methodological, analytical, reporting, and conflict of interest standards are relatively lax in the mindset literature; this flexibility in standards might be the source of most significant growth mindset intervention effects that are reported. In our extensive review, we found: (a) significant effects were most likely to be reported by authors who had financial incentives to report significant effects; (b) some highly cited studies interpreted nonsignificant effects as though they were significant (e.g., Aronson et al., 2002; Good et al., 2003); (c) many studies failed to adjust for the design effect, calling into question the veracity of some reported significant effects (e.g., Blackwell et al., 2007); and (d) most studies included major threats to internal validity, preventing any significant effects from being clearly attributable to growth mindset as the mechanism. Thus, apparent effects of growth mindset interventions may be spurious and due to inadequate study design, inappropriate interpretation of results, and bias.

In addition to our review, we conducted several meta-analyses. Our meta-analytic results further suggested that apparent growth mindset intervention effects may be due to problematic study designs, flawed reporting, and bias. When no quality control was imposed, the meta-analytic model was significant. However, when imposing quality control by examining interventions that changed students' mindsets as intended or examining the highest-quality evidence, significant effects were conspicuously absent. Further, after correcting for publication bias, no meta-analytic models were significant.

These results undermine the theoretical claims of mindset theory. Specifically, the claim that changing students' mindsets will impact their academic achievement is undermined by the null result for interventions that measurably changed students' mindsets. Likewise, claims that the effect of holding a growth mindset varies based on students' level of challenge or claims that the benefits of a growth mindset compound over time are undermined by the null results for all theoretical moderators, including student level of challenge and amount of time from the intervention to the measure of achievement.

In sum, despite the popularity of growth mindset interventions, the current evidence does not support claims that growth mindsets are beneficial for students' academic achievement. Research with higher-quality standards than the current lax standards in the field, along with a better understanding of potential harms from these interventions, is needed before recommending growth mindset interventions for use in educational contexts. Alternatively, rather than continuing to funnel resources into mindset research, efforts might be better focused on investigating more promising educational practices: Practices built on strong theoretical foundations with high potential for genuinely benefitting students' academic achievement.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Ainsworth, H., Hewitt, C. E., Higgins, S., Wiggins, A., Torgerson, D. J., & Torgerson, C. J. (2015). Sources of bias in outcome assessment in randomised controlled trials: A case study. *Educational Research and Evaluation*, *21*(1), 3–14. https://doi.org/10.1080/13803611.2014.985316

Alimohamadi, Y., & Sepandi, M. (2019). Considering the design effect in cluster sampling. *Journal of Cardiovascular and Thoracic Research*, *11*(1), Article 78. https://doi.org/10.15171/jcvtr.2019.14

Alinaghi, N., & Reed, W. R. (2018). Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work? *Research Synthesis Methods*, *9*(2), 285–311. https://doi.org/10.1002/jrsm.1298

*Anderson, D., Lammers, B., Nunnley, L., & Davis, S. (2016). *The role of goal commitment and rapport in student learning and performance* [Unpublished manuscript].

APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*(9), 839–851. https://doi.org/10.1037/0003-066X.63.9.839

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

*Aronson, J., 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. https://doi.org/10.1006/jesp.2001.1491

*Baker, J. L. W. (2017). *Growth mindset and its effect on math achievement*. [Unpublished doctoral dissertation]. California State University.

*Barron, K. E., Hulleman, C. S., Hartka, T. A., & Inouye, R. B. (2017). *Using improvement science to design and scale up social psychological interventions in schools: The case of the growth mindset app* [Unpublished manuscript].

Bekelman, J. E., Li, Y., & Gross, C. P. (2003). Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *Journal of the American Medical Association*, *289*(4), 454–465. https://doi.org/10.1001/jama.289.4.454

*Binning, K. R., Wang, M.-T., & Amemiya, J. (2019). Persistence mindset among adolescents: Who benefits from the message that academic struggles are normal and temporary? *Journal of Youth and Adolescence*, *48*(2), 269–286. https://doi.org/10.1007/s10964-018-0933-3

*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. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Boaler, J. (2013). Ability and mathematics: The mindset revolution that is reshaping education. *Forum*, *55*(1), 143–152. https://doi.org/10.2304/forum.2013.55.1.143

*Boaler, J., Dieckmann, J. A., Pérez-Núñez, G., Sun, K. L., & Williams, C. (2018). Changing students minds and achievement in mathematics: The impact of a free online student course. *Frontiers in Education*, *3*, Article 26. https://doi.org/10.3389/feduc.2018.00026

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, *8*(4), 445–454. https://doi.org/10.1177/1745691613491271

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis (version 2)*. Biostat.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

*Bostwick, K. C. (2015). *The effectiveness of a malleable mindset intervention in an introductory psychology course* [Unpublished master's thesis]. Oregon State University.

Bostwick, K. C., & Becker-Blease, K. A. (2018). Quick, easy mindset intervention can boost academic achievement in large introductory psychology classes. *Psychology Learning & Teaching*, *17*(2), 177–193. https://doi.org/10.1177/1475725718766426

Broda, M., Yun, J., Schneider, B., Yeager, D. S., Walton, G. M., & Diemer, M. (2018). Reducing inequality in academic success for incoming college students: A randomized trial of growth mindset and belonging interventions. *Journal of Research on Educational Effectiveness*, *11*(3), 317–338. https://doi.org/10.1080/19345747.2018.1429037

*Broda, M. D. (2015). *Three perspectives on postsecondary attainment: Students, teachers, and institutions* [Unpublished doctoral dissertation]. Michigan State University.

*Brougham, L., & Kashubeck-West, S. (2018). Impact of a growth mindset intervention on academic performance of students at two urban high schools. *Professional School Counseling*, *21*(1), 1–9. https://doi.org/10.1177/2156759X18764934

Burgoyne, A. P., Hambrick, D. Z., & Macnamara, B. N. (2020). How firm are the foundations of mind-set theory? The claims appear stronger than the evidence. *Psychological Science*, *31*(3), 258–267. https://doi.org/10.1177/0956797619897588

Burgoyne, A. P., Hambrick, D. Z., Moser, J. S., & Burt, S. A. (2018). Analysis of a mindset intervention. *Journal of Research in Personality*, *77*, 21–30. https://doi.org/10.1016/j.jrp.2018.09.004

Burgoyne, A. P., & Macnamara, B. N. (2021). Reconsidering the use of the Mindset Assessment Profile in educational contexts. *Journal of Intelligence*, *9*(3), Article 39. https://doi.org/10.3390/jintelligence9030039

Burnette, J. L., & Finkel, E. J. (2012). Buffering against weight gain following dieting setbacks: An implicit theory intervention. *Journal of Experimental Social Psychology*, *48*(3), 721–725. https://doi.org/10.1016/j.jesp.2011.12.020

*Burnette, J. L., Hoyt, C. L., Russell, V. M., Lawson, B., Dweck, C. S., & Finkel, E. (2016). *A growth mind-set intervention improves interest but not academic performance in the field of computer science* [Unpublished preprint of methods and results].

Burnette, J. L., Hoyt, C. L., Russell, V. M., Lawson, B., Dweck, C. S., & Finkel, E. (2019). A growth mind-set intervention improves interest but not academic performance in the field of computer science. *Social Psychological & Personality Science*, *11*(1), 107–116. https://doi.org/10.1177/1948550619841631

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin*, *139*(3), 655–701. https://doi.org/10.1037/a0029531

*Burnette, J. L., Russell, M. V., Hoyt, C. L., Orvidas, K., & Widman, L. (2018). An online growth mindset intervention in a sample of rural adolescent girls. *The British Journal of Educational Psychology*, *88*(3), 428–445. https://doi.org/10.1111/bjep.12192

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

*Chao, M. M., Visaria, S., Mukhopadhyay, A., & Dehejia, R. (2017). Do rewards reinforce the growth mindset?: Joint effects of the growth mindset and incentive schemes in a field intervention. *Journal of Experimental Psychology: General*, *146*(10), 1402–1419. https://doi.org/10.1037/xge0000355

Cheung, S. F., & Chan, D. K.-S. (2004). Dependent effect sizes in meta-analysis: Incorporating the degree of interdependence. *Journal of Applied Psychology*, *89*(5), 780–791. https://doi.org/10.1037/0021-9010.89.5.780

Cheung, S. F., & Chan, D. K.-S. (2008). Dependent correlations in meta-analysis: The case of heterogeneous dependence. *Educational and Psychological Measurement*, *68*(5), 760–777. https://doi.org/10.1177/0013
164408315263

Chivers, T. (2017). A mindset "revolution" sweeping Britain's classrooms may be based on shaky science. *BuzzFeed*. https://www.buzzfeed.com/tomchivers/what-is-your-mindset

Chivers, T. (2019). Does psychology have a conflict-of-interest problem? *Nature*, *571*(7763), 20–23. https://doi.org/10.1038/d41586-019-02041-5

Chow, S.-C., & Liu, J.-P. (2004). *Design and analysis of clinical trials: Concepts and methodologies*. Wiley.

Churches & Education Development Trust. (2016). *Closing the gap: Test and learn*. National College for Teaching and Leadership.

Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(31), 8664–8668. https://doi.org/10.1073/pnas.1608207113

*Coates, K. (2016). *An evaluation of growing early mindsets (GEM™)* [Unpublished doctoral dissertation]. University of Oregon.

Coates, K. (2020). *The girls' guide to growth mindset: A can-do approach to building confidence, courage, and grit*. Rockridge Press.

Coe, R. (2002). *It's the effect size, stupid: What effect size is and why it is important* [Paper presented]. The Annual Conference of the British Educational Research Association, University of Exeter, England. http://www.leeds.ac.uk/educol/documents/00002182.htm

*Coleman, A. (2019). *The effect of a growth mindset program on mathematics achievement of high school upperclassmen* [Unpublished master's thesis]. Goucher College.

Cooper, H., DeNeve, K., & Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods*, *2*(4), 447–452. https://doi.org/10.1037/1082-989X.2.4.447

Dai, T., & Cromley, J. G. (2014). Changes in implicit theories of ability in biology and dropout from STEM majors: A latent growth curve approach. *Contemporary Educational Psychology*, *39*(3), 233–247. https://doi.org/10.1016/j.cedpsych.2014.06.003

*De Martino, S., Porter, T., Martinus, A., Ross, R., Cyster, C. F., Pons, G., & Trzesniewski, K. H. (2019). *Changing learner beliefs in South African townships—an evaluation of a growth mindset intervention* [Unpublished manuscript].

*Delavande, A., Del Bono, E., Holford, A., & Sen, S. (2019). *Skills accumulation with malleable ability: Evidence from a growth mindset intervention* [Unpublished manuscript].

*Delpéche, H. E. (2018). *The implicit and the explicit: The impact of teaching academic mindsets and reading strategies on beginning college learners' reading comprehension* [Unpublished doctoral dissertation]. University of Delaware.

Dickerson, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis prevention, assessment and adjustments* (pp. 11–34). Wiley. https://doi.org/10.1002/0470870168.ch2

*Dommett, E. J., Devonshire, I. M., Sewter, E., & Greenfield, S. A. (2013). The impact of participation in a neuroscience course on motivational measures and academic performance. *Trends in Neuroscience and Education*, *2*(3–4), 122–138. https://doi.org/10.1016/j.tine.2013.05.002

*Donohoe, C., Topping, K., & Hannah, E. (2012). The impact of an online intervention (Brainology) on the mindset and resiliency of secondary school pupils: A preliminary mixed methods study. *Educational Psychology*, *32*(5), 641–655. https://doi.org/10.1080/01443410.2012.675646

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040–1048. https://doi.org/10.1037/0003-066X.41.10.1040

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success* (1st ed.). Random House.

Dweck, C. S. (2007a). Boosting achievement with messages that motivate. *Education Canada*, *47*(2), 6–10. https://www.edcan.ca/wp-content/uploads/EdCan-2007-v47-n2-Dweck.pdf

Dweck, C. S. (2007b). The perils and promises of praise. *Educational Leadership*, *65*(2), 34–39. http://mail.rpforschools.net/articles/Mindsets/Dweck%202007%20The%20Perils%20and%20Promises%20of%20Praise.pdf

Dweck, C. S. (2008a, Winter). Brainology: Transforming students' motivation to learn. *Independent School Magazine*. https://www.nais.org/magazine/independent-school/winter-2008/brainology/

Dweck, C. S. (2008b). Mindsets: How praise is harming youth and what can be done about it. *School Library Media Activities Monthly*, *24*(5). https://www.highpoint.edu/qep/files/2015/01/Dweck-Mindset-Article.pdf

Dweck, C. S. (2009). Mindsets: Developing talent through a growth mindset. *Olympic Coach*, *21*(1), 4–7. https://www.teamusa.org/-/media/TeamUSA/SportPerformance/coaching/Olympic-Coach-E-Mag/Winter-2009.pdf?la=en&hash=45B6302AE529D0E130B358D82323F2E5A32043FD

Dweck, C. S. (2010). Mind-sets and equitable education. *Principal Leadership*, *10*(5), 26–29. http://www.my-ecoach.com/online/resources/3865/Equitable_Mindsets.pdf

Dweck, C. S. (2012). Mindsets and human nature: Promoting change in the Middle East, the schoolyard, the racial divide, and willpower. *American Psychologist*, *67*(8), 614–622. https://doi.org/10.1037/a0029783

Dweck, C. S. (2016). *Mindset: The new psychology of success* (2nd ed.). Random House.

Dweck, C. S. (2018, June 26). Growth mindset interventions yield impressive results. *The Conversation*. https://theconversation.com/growth-mindset-interventions-yield-impressive-results-97423

Dweck, C. S., & Bempechat, J. (1983). Theories of intelligence and achievement motivation. *Learning and motivation in the classroom*. Erlbaum.

Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological Inquiry*, *6*(4), 267–285. https://doi.org/10.1207/s15327965pli0604_1

Dweck, C. S., Walton, G. M., & Cohen, G. L. (2014). *Academic tenacity mindsets and skills that promote long-term learning*. Bill & Melinda Gates Foundation. https://ed.stanford.edu/sites/default/files/manual/dweck-walton-cohen-2014.pdf

Dweck, C. S., & Yeager, D. S. (2019a). *A simple re-analysis overturns a "failure to replicate" and highlights an opportunity to improve scientific practice: Commentary on Li and Bates (in press)* [Unpublished manuscript]. https://www.researchgate.net/profile/David-Yeager/publication/337856605_A_Simple_Re-Analysis_Overturns_a_Failure_to_Replicate_and_Highlights_an_Opportunity_to_Improve_Scientific_Practice_Commentary_on_Li_and_Bates_2019/links/5def33d292851c83647068f8/A-Simple-Re-Analysis-Overturns-a-Failure-to-Replicate-and-Highlights-an-Opportunity-to-Improve-Scientific-Practice-Commentary-on-Li-and-Bates-2019.pdf

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

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

*Ehrlinger, J., Mitchum, A. L., & Dweck, C. S. (2016). Understanding overconfidence: Theories of intelligence, preferential attention, and distorted self-assessment. *Journal of Experimental Social Psychology*, *63*(4), 94–100. https://doi.org/10.1016/j.jesp.2015.11.001

Eich, E. (2014). Business not as usual. *Psychological Science*, *25*(1), 3–6. https://doi.org/10.1177/0956797613512465

*Fabert, N. S. (2014). *Growth mindset training to increase women's self-efficacy in science and engineering: A randomized-controlled trial* [Unpublished doctoral dissertation]. Arizona State University.

*Fink, A., Cahill, M. J., McDaniel, M. A., Hoffman, A., & Frey, R. F. (2018). Improving general chemistry performance through a growth mindset intervention: Selective effects on underrepresented minorities. *Chemistry Education Research and Practice*, *19*(3), 783–806. https://doi.org/10.1039/C7RP00244K

Flanigan, A. E., Peteranetz, M. S., Shell, D. F., & Soh, L.-K. (2017). Implicit intelligence beliefs of computer science students: Exploring change across the semester. *Contemporary Educational Psychology*, *48*(7), 179–196. https://doi.org/10.1016/j.cedpsych.2016.10.003

*Fleurizard, T. A., & Young, P. R. (2018). Finding the right equation for success: An exploratory study on the effects of a growth mindset intervention on college students in remedial math. *Journal of Counseling Psychology*, *2*(1), Article 3. https://digitalcommons.gardner-webb.edu/jcp/vol2/iss1/3

*Foliano, F., Rolfe, H., Buzzeo, J., Runge, J., & Wilkinson, D. (2019). Changing mindsets: Effectiveness trial. *Evaluation report*. Education Endowment Foundation.

Food and Drug Administration. (2007). *Section 801 of the Food and Drug Administration amendments act of 2007*. https://www.govinfo.gov/content/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf#page=82

Franco, A., Malhotra, N., & Simonovits, G. (2014). Social science. Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Friese, M., & Frankenbach, J. (2020). p-Hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods*, *25*(4), 456–471. https://doi.org/10.1037/met0000246

Ganimian, A. J. (2018). *Growth mindset interventions at scale: Experimental evidence from Argentina* [Unpublished manuscript]. New York University.

Garg, A. X., Adhikari, N. K. J., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., Sam, J., & Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *Journal of the American Medical Association*, *293*(10), 1223–1238. https://doi.org/10.1001/jama.293.10.1223

*Gauthreaux, E. F. (2015). *Effect of specific feedback on growth mindset and achievement*. [Unpublished doctoral dissertation]. Louisiana State University.

*Glerum, J., Loyens, S. M. M., & Rikers, R. M. J. P. (2020). Is an online mindset intervention effective in vocational education? *Interactive Learning Environments*, *20*(7), 821–830. https://doi.org/10.1080/10494820.2018.1552877

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights*, *1*(3), 291–308. https://doi.org/10.1257/aeri.20180633

Goldenberg, A., Cohen-Chen, S., Goyer, J. P., Dweck, C. S., Gross, J. J., & Halperin, E. (2018). Testing the impact and durability of a group malleability intervention in the context of the Israeli-Palestinian conflict. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(4), 696–701. https://doi.org/10.1073/pnas.1706800115

*Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, *24*(6), 645–662. https://doi.org/10.1016/j.appdev.2003.09.002

Gorski, P. C. (2011). Unlearning deficit ideology and the scornful gaze: Thoughts on authenticating the class discourse in education. *Counterpoints*, *402*, 152–173. https://www.jstor.org/stable/42981081

Gorski, P. C. (2017). *Reaching and teaching students in poverty: Strategies for erasing the opportunity gap* (2nd ed.). Teachers College Press.

Gorski, P. C. (2019). Avoiding racial equity detours. *Educational Leadership*, *76*(7), 56–61. https://www.fairforall.org/content/pdfs/haakmat-consulting/avoiding-racial-equity-detours.pdf

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. https://doi.org/10.1111/cdev.12955

Halperin, E., Russell, A. G., Trzesniewski, K. H., Gross, J. J., & Dweck, C. S. (2011). Promoting the Middle East peace process by changing beliefs about group malleability. *Science*, *333*(6050), 1767–1769. https://doi.org/10.1126/science.1202925

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hilgard, J. (2015). *PETPEESE. Source code*. https://github.com/Joe-Hilgard/PETPEESE

Hilgard, J. (2020). *PETPEESE. Source code*. https://github.com/Joe-Hilgard/PETPEESE

*Hoang, T. V. (2018). *Growth mindset and task value interventions in college algebra* [Unpublished doctoral dissertation]. Texas State University.

Hoewe, J. (2017). Manipulation check. In J. Matthes (Ed.), *The international encyclopedia of communication research methods* (pp. 1–5). Wiley. https://doi.org/10.1002/9781118901731.iecrm0135

*Holden, L. R., Moreau, D., Greene, D., & Conway, A. R. A. (2016). *The role of mindset: Investigating performance feedback and learning strategies in an online statistics course* [Poster presented]. The 57th annual meeting of the Psychonomic Society, Boston, MA.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal*, *319*(7211), 670–674. https://doi.org/10.1136/bmj.319.7211.670

Hong, S. (2019). Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work? A replication study of Alinaghi & Reed (Research Synthesis Methods, 2018). *International Journal for Re-Views in Empirical Economics (IREE)*, *3*(2019-4), 1–22. https://doi.org/10.18718/81781.13

Hox, J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways. Studies in classification, data analysis, and knowledge organization*. Springer. https://doi.org/10.1007/978-3-642-72087-1_17

Hoyt, C. L., & Burnette, J. L. (2020). Growth mindset messaging in stigma-relevant contexts: Harnessing benefits without costs. *Policy Insights from the Behavioral and Brain Sciences*, *7*(2), 157–164. https://doi.org/10.1177/2372732220941216

Hoyt, C. L., Burnette, J. L., Auster-Gussman, L., Blodorn, A., & Major, B. (2017). The obesity stigma asymmetry model: The indirect and divergent effects of blame and changeability beliefs on antifat prejudice. *Stigma and Health*, *2*(1), 53–65. https://doi.org/10.1037/sah0000026

*Huffman, G. C. (2017). *An intelligence mindset intervention for first semester college students* [Unpublished doctoral dissertation]. Indiana University of Pennsylvania.

*Hunt, J. F. (2017). *Academic math mindset interventions in first-year college calculus* [Unpublished doctoral dissertation]. University of Texas-Austin.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLOS ONE*, *10*(8), Article e0132382. https://doi.org/10.1371/journal.pone.0132382

Kish, L. (1965). *Survey sampling*. Wiley.

*Kloster, K. R. (2016). *Facet-level personality development: An intervention for developing student self-discipline and orderliness* [Unpublished master's thesis]. St. Cloud State University.

Lakens, D. (2014). Observed power, and what to do if your editor asks for post-hoc power analyses. *The 20% Statistician*. http://daniellakens.blogspot.com/2014/12/observed-power-and-what-to-do-if-your.html

Levy, S. R., & Dweck, C. S. (1999). The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development*, *70*(5), 1163–1180. https://doi.org/10.1111/1467-8624.00085

Li, Y., & Bates, T. C. (2019). You can't change your basic ability, but you work at things, and that's how we get hard things done: Testing the role of growth mindset on response to setbacks, educational attainment, and cognitive ability. *Journal of Experimental Psychology: General*, *148*(9), 1640–1655. https://doi.org/10.1037/xge0000669

Li, Y., & Bates, T. C. (2020). Testing the association of growth mindset and grades across a challenging transition: Is growth mindset associated with grades? *Intelligence*, *81*(2), Article 101471. https://doi.org/10.1016/j.intell.2020.101471

Limeri, L. B., Choe, J., Harper, H. G., Martin, H. R., Benton, A., & Dolan, E. L. (2020). Knowledge or abilities? How undergraduates define intelligence. *CBE Life Sciences Education*, *19*(1), Article ar5. https://doi.org/10.1187/cbe.19-09-0169

Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics*, *74*(3), 785–794. https://doi.org/10.1111/biom.12817

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis*. SAGE Publications.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*(5), 304–305. https://doi.org/10.1037/h0025105

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40. https://doi.org/10.1037/1082-989X.7.1.19

Macnamara, B. N., & Rupani, N. S. (2017). The relationship between intelligence and mindset. *Intelligence*, *64*(3), 52–59. https://doi.org/10.1016/j.intell.2017.07.003

Magnusson, K. (2020). Interpreting Cohen's d effect size: An interactive visualization (Version 2.2.0) [Web App]. *R Psychologist*. https://rpsychologist.com/d3/cohend/

McCoach, D. B., & Adelson, J. L. (2010). Dealing with dependence (Part I): Understanding the effects of clustered data. *Gifted Child Quarterly*, *54*(2), 152–155. https://doi.org/10.1177/0016986210363076

*Mills, I. M., & Mills, B. S. (2018). Insufficient evidence: Mindset intervention in developmental college math. *Social Psychology of Education*, *21*(5), 1045–1059. https://doi.org/10.1007/s11218-018-9453-y

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moreau, D., Macnamara, B. N., & Hambrick, D. Z. (2019). Overstating the role of environmental factors in success: A cautionary note. *Current Directions in Psychological Science*, *28*(1), 28–33. https://doi.org/10.1177/0963721418797300

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest control group designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology*, *75*(1), 33–52. https://doi.org/10.1037/0022-3514.75.1.33

*North Wolfe, S. (2017). *Effects of a growth mindset intervention on first-year college student academic performance* [Unpublished master's thesis]. Central Washington University.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783. https://doi.org/10.1037/h0043424

*Orosz, G., Péter-Szarka, S., Bőthe, B., Tóth-Király, I., & Berger, R. (2017). How not to do a mindset intervention: Learning from a mindset intervention among students with good grades. *Frontiers in Psychology*, *8*, Article 311. https://doi.org/10.3389/fpsyg.2017.00311

Outes, I., Sanchez, A., & Vakis, R. (2017). Growth mindset at scale: Impact of a psychosocial intervention on secondary school attainment in Peru. *Project: Growth Mindset at Scale—Increasing school attainment by affecting the mindset of pupils and teachers*. [Unpublished project report]. https://riseprogramme.org/sites/default/files/inline-files/Outes-Leon,%20Ingo,%20Sanchez,%20Alan,%20Vakis,%20Renos.%20%20Project-%20Growth%20Mindset%20at%20Scale.pdf

*Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science*, *26*(6), 784–793. https://doi.org/10.1177/0956797615571017

Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology*, *92*(1), 128–150. https://doi.org/10.1037/0021-9010.92.1.128

Perlis, R. H., Perlis, C. S., Wu, Y., Hwang, C., Joseph, M., & Nierenberg, A. A. (2005). Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *The American Journal of Psychiatry*, *162*(10), 1957–1960. https://doi.org/10.1176/appi.ajp.162.10.1957

*Peterson, L. (2018). *Can explicit teaching of a growth mindset in early elementary increase student ability?* [Unpublished doctoral dissertation]. Concordia University.

Peterson, V. (2019). How book advances and royalties work. *The Balance Careers*. https://www.thebalancecareers.com/book-advances-and-royalties-2799832#:~:text=What%20Is%20a%20Book%20Royalty,as%20a%20percentage%20of%20sales

*Polley, T. (2018). *Believing is achieving: Targeting beliefs about intelligence to increase learning in Bangladeshi secondary schools* [Unpublished job market paper]. Duke University.

Porta, N., Bonet, C., & Cobo, E. (2007). Discordance between reported intention-to-treat and per protocol analyses. *Journal of Clinical Epidemiology*, *60*(7), 663–669. https://doi.org/10.1016/j.jclinepi.2006.09.013

*Quille, K., & Bergin, S. (2019). CS1: How will they do? How can we help? A decade of research and practice. *Computer Science Education*, *29*(2–3), 254–282. https://doi.org/10.1080/08993408.2019.1612679

Rattan, A., Savani, K., Chugh, D., & Dweck, C. S. (2015). Leveraging mindsets to promote academic achievement: Policy recommendations. *Perspectives on Psychological Science*, *10*(6), 721–726. https://doi.org/10.1177/1745691615599383

Redick, T. S., & Webster, S. B. (2014). Videogame interventions and spatial ability interactions. *Frontiers in Human Neuroscience*, *8*, Article 183. https://doi.org/10.3389/fnhum.2014.00183

*Rhee, J., Johnson, C., & Oyamot, C. M. (2017). *Preliminary findings using growth mindset and belonging interventions in a freshman engineering class* [Paper presented]. The 2017 Conference of American Society for Engineering Education, Columbus, Ohio, United States.

*Rienzo, C., Wolfe, H., & Wilkinson, D. (2015). *Changing mindsets: Evaluation report and executive summary*. Education Endowment Foundation.

Ritchie, S. (2020). *Science fictions: How fraud, bias, negligence, and hype undermine the search for truth*. Henry Holt and Company.

*Robinson, K. A. (2019). *Supporting multiple paths to success: A field experiment examining a multifaceted, multilevel motivation intervention* [Unpublished doctoral dissertation]. Michigan State University.

Roseman, M., Milette, K., Bero, L. A., Coyne, J. C., Lexchin, J., Turner, E. H., & Thombs, B. D. (2011). Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments. *Journal of the American Medical Association*, *305*(10), 1008–1017. https://doi.org/10.1001/jama.2011.257

Rosenthal, R. (1976). *Experirmenter effects in behavioral research*. Irvington Publishers.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16–20. https://doi.org/10.1007/BF02322211

Rosenthal, R., & Rubin, D. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, *1*(3), 377–386. https://doi.org/10.1017/S0140525X00075506

Russell, T. V. (2019, April 20). The trouble with grit and growth mindset. *Mrs. Russell's Room*. http://tamaravrussell.com/2019/04/the-trouble-with-grit-growth-mindset/

*Saunders, S. A. (2013). *The impact of a growth mindset intervention on the reading achievement of at-risk adolescent students* [Unpublished doctoral dissertation]. University of Virginia.

*Schubert, L. K. (2017). *Exploring the connections between students' mindsets and their writing: An intervention study with a course-embedded writing tutor* [Unpublished doctoral dissertation]. Indiana University of Pennsylvania.

Shankar, M., & Kalil, T. (2013). Leveraging mental muscle for academic excellence. *Obama White House Blog*. https://obamawhitehouse.archives.gov/blog/2013/06/28/leveraging-mental-muscle-academic-excellence

Shaya, F. T., & Gu, A. (2007). Critical assessment of intent-to-treat analyses. *Journal of Medical Economics*, *10*(2), 171–177. https://doi.org/10.3111/200710171177

Sheiner, L. B. (2002). Is intent-to-treat analysis always (ever) enough? *British Journal of Clinical Pharmacology*, *54*(2), 203–211. https://doi.org/10.1046/j.1365-2125.2002.01628.x

Simmonds, M. (2015). Quantifying the risk of error when interpreting funnel plots. *Systematic Reviews*, *4*(24), Article 24. https://doi.org/10.1186/s13643-015-0004-8

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To what extent and under which circumstances are growth mindsets important to academic achievement? Two meta-analyses. *Psychological Science*, *29*(4), 549–571. https://doi.org/10.1177/0956797617739704

*Smith, T. F., & Capuzzi, G. (2019). Using a mindset intervention to reduce anxiety in the statistics classroom. *Psychology Learning & Teaching*, *18*(3), 326–336. https://doi.org/10.1177/1475725719836641

*Sriram, R. (2014). Rethinking intelligence: The role of mindset in promoting success for academically high-risk students. *Journal of College Student Retention*, *15*(4), 515–536. https://doi.org/10.2190/CS.15.4.c

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

*Tecker, S. (2017). *Bridging the gap—Growth mindset research and educators' practice* [Unpublished doctoral dissertation]. Concordia University Irvine.

Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology*, *58*(9), 894–901. https://doi.org/10.1016/j.jclinepi.2005.01.006

Thomas, P. L. (2016, August 10). Failing still to address poverty directly: Growth mindset as deficit ideology. *Medium*. https://plthomasedd.medium.com/failing-still-to-address-poverty-directly-growth-mindset-as-deficit-ideology-33e552f92223

Thomas, P. L. (2018, May 30). More on rejecting growth mindset, grit. *Medium*. https://medium.com/@plthomasedd/more-on-rejecting-growth-mindset-grit-52e5eb47374e

Tillis, G. E. (2019). Toward a transformative transition: A critical pedagogical approach to social-psychological interventions in first-year seminar. In J. Hoffman, P. Blessinger, & M. Makhanya (Eds.), *Strategies for fostering inclusive classrooms in higher education: International perspectives on equity and inclusion (innovations in higher education teaching and learning* (Vol. 16, pp. 183–195). Emerald Publishing Limited. https://doi.org/10.1108/S2055-364120190000016015

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

What Works Clearinghouse. (2020). *Procedures handbook, Version 4.1*.

*Whorrall, A. L. (2018). *Academic motivation and achievement of middle school Hispanic students: Exploring the use of mindset and Brainology in education* [Unpublished doctoral dissertation]. Northcentral University.

*Wilkins, P. (2014). *Efficacy of a growth mindset intervention to increase student achievement* [Unpublished doctoral dissertation]. Gardner-Webb University.

Williams, R. (2012). *Moderator analyses: Categorical models and meta-regression* [Paper presented]. The annual Campbell Collaboration Colloquium, Copenhagen, Denmark.

*Wilson, A. R. (2009). *Malleable view of intelligence as intervention for stereotype threat: Overcoming math underperformance in women* [Unpublished master's thesis]. East Carolina University, Greenville, North Carolina.

World Bank Blogs. (2017). The power of believing you can get smarter: The impact of a growth-mindset intervention on academic achievement in Peru (English). *News Break*. https://www.newsbreak.com/news/1504098819617/the-power-of-believing-you-can-get-smarter-the-impact-of-a-growth-mindset-intervention-on-academic-achievement-in-peru-english

Wright, D. B. (2006). Comparing groups in a before-after design: When *t* test and ANCOVA produce different results. *The British Journal of Educational Psychology*, *76*(3), 663–675. https://doi.org/10.1348/000709905X52210

*Wright, J. C. (2018). *The challenge of community college student academic motivation: The Go for Growth! Intervention* [Unpublished doctoral dissertation]. University of North Carolina at Chapel Hill.

Yan, V. X., Thai, K. P., & Bjork, R. A. (2014). Habits and beliefs that guide self-regulated learning: Do they vary with mindset? *Journal of Applied Research in Memory and Cognition*, *3*(3), 140–152. https://doi.org/10.1037/h0101799

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

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies? *American Psychologist*, *75*(9), 1284–1284. https://doi.org/10.1037/amp0000794

*Yeager, D. S., Hanselman, P., Walton, G. M., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C., Hinojosa, C., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang Hooper, S., Murray, J., Carvalho, C., . . . Dweck, C. S. (2018). Where and for whom can a brief, scalable mindset intervention improve adolescents' educational trajectories? *PsyArXiv*. https://doi.org/10.31234/osf.io/md2qa

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., . . . Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y

*Yeager, D. S., Johnson, R., Spitzer, B. J., Trzesniewski, K. H., Powers, J., & Dweck, C. S. (2014). The far-reaching effects of believing people can change: Implicit theories of personality shape stress, health, and achievement during adolescence. *Journal of Personality and Social Psychology*, *106*(6), 867–884. https://doi.org/10.1037/a0036335

*Yeager, D. S., Lee, H. Y., & Jamieson, J. P. (2016). How to improve adolescent stress responses: Insights from integrating implicit theories of personality and biopsychosocial models. *Psychological Science*, *27*(8), 1078–1091. https://doi.org/10.1177/0956797616649604

*Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., Lee, H. Y., O'Brien, J., Flint, K., Roberts, A., Trott, J., Greene, D., Walton, G. M., & 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. https://doi.org/10.1037/edu0000098

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*(2), 267–301. https://doi.org/10.3102/0034654311405999

Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., Kamentz, D., Ritter, G., Duckworth, A. L., Urstein, R., Gomez, E. M., Markus, H. R., Cohen, G. L., Dweck, C. S., & Dweck, C. S. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(24), E3341–E3348. https://doi.org/10.1073/pnas.1524360113

Zonnefeld, V. (2019). Implications of training in incremental theories of intelligence for undergraduate statistics students. In M. S. Hannula, G. C. Leder, F. Moreselli, M. Vollstedt, & Q. Zhang (Eds.), *Affect and mathematics education*. Springer. https://doi.org/10.1007/978-3-030-13761-8_9

*Zonnefeld, V. L. (2015). *Mindsets, attitudes, and achievement in undergraduate statistics courses* [Unpublished doctoral dissertation]. Dordt College.

*Zonnefeld, V. L., & Van Weelden, K. (2016). *Mindset training for undergraduates in developmental mathematics* [Poster presented]. Faculty Work, Dordt College.

(*Appendix follows*)

Appendix

Deviations From and Decisions Not Explicit in the Preregistration

Praise-Only/Struggle Story Manipulations

We did not include treatments where the manipulation only praised effort or attributed success to effort or struggle (e.g., teaching that Einstein struggled) because those manipulations did not directly teach students that a human characteristic was malleable.

Intent-to-Treat Analyses

High-profile, large-scale growth mindset intervention studies (e.g., the National Learning Mindset Study, Yeager et al., 2019) were published or otherwise became available in the last few years (that we became aware of after preregistering) that used intent-to-treat analysis, that is, including in the analyses students who did not receive or comply to the treatment. Given these studies' relevance it would have been inappropriate to exclude them from the present meta-analyses. We, therefore, allowed studies that only provided effects from intent-to-treat analyses into the meta-analyses.

Default Intraclass Correlation

Hox (1998) states that the typical intraclass correlation for school effects is $\rho = .10$. However, Hox (1998) does not provide a citation for this value. Upon further investigation, we found no rationale for a correlation of $\rho = .10$, and that the What Works Clearinghouse. (2020) under the Department of Education uses a default intraclass correlation of $\rho = .20$ for achievement outcomes if no intraclass correlation is provided by the study. We use this same standard in the present study.

Coding of Unpublished Studies

We preregistered that if an unpublished and a published version of the same study were both available, we would use whichever had the larger sample size in the meta-analyses, but code it as published because the study (in some form) had been published. However, we had not anticipated the sometimes vast differences between unpublished and published study versions. For example, the unpublished version of Yeager et al. (2019, 2018) included the results for 12,542 students, whereas the published version included the results for 6,320 students; the unpublished version met the best practice criterion of reporting the whole sample whereas the published version did not. Upon reflection, we decided to code the publication status to match the reference and coding we were using in the meta-analyses. We reasoned that coding features as published that were not published could be confusing.

Adjusting for Prior Academic Achievement

When enough information was provided by the study, we adjusted for groups' prior academic achievement (e.g., difference of change scores).

Financial Stake

We included book sales as having a financial stake if the book promotes growth mindset.

Preregistrations

We accepted clearly stated study aims in place of hypotheses in study preregistrations.

Power Analyses

In our preregistration, we noted that adequate power cannot be determined without a preregistration that included an a priori power analysis and stopping rule. As a proxy, we planned to conduct post hoc power analyses and to code samples as adequate in size if there was >.80 power. However, post hoc power does not reflect the true power of the test and is more closely aligned with the p value than the sample size or effect size (Lakens, 2014). For example, in our data set only 1 of the 13 independent samples with $N > 1{,}000$ had post hoc power >.80. In contrast, samples as small as $n = 14$ per group had post hoc power >.80. We changed the criterion for a study to meet this best practice as having reported conducting an a priori power analysis.

Criteria Removed Based on Reviewer Recommendations

One or more reviewer(s) suggested we remove several criteria we originally put forth as best practices (described next with rationale). These criteria increase the interpretability of the results but do not necessarily impact the quality of the study design or the size of the effect. Results with these criteria included can be found in the Supplemental Materials.

Reporting Standards

We preregistered several best practice criteria in reporting that do not impact the effect size: controlling the familywise Type I error rate; testing for differences among subsamples when differences were hypothesized/assumed by the study authors; reporting variance around means (e.g., standard deviations, confidence intervals); reporting effect sizes for key results; and appropriately interpreting results. We removed the reporting criteria that did not impact the effect size entered in the meta-analysis.

We also removed *conducting and reporting results interpretable to most readers*. We describe this further in the Supplemental Materials.

Conducting Theory-Driven Analyses to Test Hypotheses

We preregistered that conducting theory-driven, rather than data-driven, analyses to test hypotheses is a best practice. While an explicit data-driven method can be discerned from a description, a

(*Appendix continues*)

reviewer pointed out that analyses chosen from data cannot. (We originally coded all studies as meeting this criterion.)

Equivalence at Baseline

We preregistered as a best practice criterion that the control group should be comparable to the treatment group on the outcome measure (i.e., academic achievement) at baseline. We found that 58% of the samples—37% of the total students—included in the meta-analysis reported equivalent academic achievement between the treatment and control groups at baseline. However, a reviewer pointed out that equivalent baseline performance is not a best practice: baseline differences will occur by chance with randomization and should be unbiased in the long run. We, therefore, removed this as a best practice criterion.

Meta-Analytic Models With Studies Meeting at Least Half the Best Practices Criteria

The results of these models are presented in the Supplemental Materials.

Alternate Best Practices Models

The meta-analyses of studies adhering to every combination of best practices criteria were not preregistered but were suggested by a reviewer.

Received November 15, 2020
Revision received January 21, 2022
Accepted February 4, 2022 ■