AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Bulletin

ISSN: 0033-2909

https://doi.org/10.1037/bul0000368

A Systematic Review and Meta-Analysis of Growth Mindset Interventions: For Whom, How, and Why Might Such Interventions Work?

Jeni L. Burnette[1], Joseph Billingsley[2], George C. Banks[3], Laura E. Knouse[4], Crystal L. Hoyt[5], Jeffrey M. Pollack[6], and Stefanie Simon[7]

[1] Department of Psychology, College of Humanities and Social Sciences, North Carolina State University
[2] School of Science and Engineering, Tulane University
[3] Belk College of Business, University of North Carolina Charlotte
[4] School of Arts and Sciences, University of Richmond
[5] Jepson School of Leadership Studies, University of Richmond
[6] Poole College of Management, North Carolina State University
[7] School of Liberal Arts, Siena College

As growth mindset interventions increase in scope and popularity, scientists and policymakers are asking: Are these interventions effective? To answer this question properly, the field needs to understand the meaningful heterogeneity in effects. In the present systematic review and meta-analysis, we focused on two key moderators with adequate data to test: Subsamples expected to benefit most and implementation fidelity. We also specified a process model that can be generative for theory. We included articles published between 2002 (first mindset intervention) through the end of 2020 that reported an effect for a growth mindset intervention, used a randomized design, and featured at least one of the qualifying outcomes. Our search yielded 53 independent samples testing distinct interventions. We reported cumulative effect sizes for multiple outcomes (i.e., mindsets, motivation, behavior, end results), with a focus on three primary end results (i.e., improved academic achievement, mental health, or social functioning). Multilevel metaregression analyses with targeted subsamples and high fidelity for academic achievement yielded, $d = 0.14$, 95% CI [.06, .22]; for mental health, $d = 0.32$, 95% CI [.10, .54]. Results highlighted the extensive variation in effects to be expected from future interventions. Namely, 95% prediction intervals for focal effects ranged from −0.08 to 0.35 for academic achievement and from 0.07 to 0.57 for mental health. The literature is too nascent for moderators for social functioning, but average effects are $d = 0.36$, 95% CI [.03, .68], 95% PI [−.50, 1.22]. We conclude with a discussion of heterogeneity and the limitations of meta-analyses.

Public Significance Statement
Growth mindset interventions are increasing in popularity in education and are being applied to improving other areas of functioning as well; however, there is debate about how well they work. Despite the large variation in effectiveness, we found positive effects on academic outcomes, mental health, and social functioning, especially when interventions are delivered to people expected to benefit the most.

Keywords: mindset intervention, metaregression, meta-analysis, systematic review, heterogeneity

Growth mindset interventions, which seek to foster beliefs in the malleable nature of abilities, attributes, and traits, are rising in popularity and expanding in scope (Dweck & Yeager, 2019). This increased attention is illustrated not only by the creation of businesses, both nonprofit (e.g., Project for Education Research That Scales [PERTS]) and for profit (e.g., Mindsets Work) but also by application to diverse problems, such as mitigating the consequences of bias (Okonofua et al., 2016), understanding consumer behavior (Murphy & Dweck, 2016), and improving mental health (Schleider & Weisz, 2016). As growth mindset work proliferates, researchers are synthesizing the mounting empirical findings. Multiple meta-analytic reviews published in the last few years (Burnette et al., 2013; Burnette, Knouse, et al., 2020; Costa & Faria, 2018; Sarrasin et al., 2018; Schleider et al., 2015; Sisk et al., 2018) reflect the field's interest in reaching consensus on critical empirical questions: Do mindsets predict outcomes as expected? Are effects

Jeni L. Burnette https://orcid.org/0000-0001-7548-8782
Joseph Billingsley https://orcid.org/0000-0002-6577-175X
Laura E. Knouse https://orcid.org/0000-0001-7080-431X
Crystal L. Hoyt https://orcid.org/0000-0001-9108-3910
The authors would like to thank Scott Tonidandel for his analytic advice and help with the code. All data, code, and supplemental files are available at Open Science Framework (*OSF*; https://osf.io/2fgey).

Correspondence concerning this article should be addressed to Jeni L. Burnette, Department of Psychology, College of Humanities and Social Sciences, North Carolina State University, 640 Poe Hall, Campus Box 7650, Raleigh, NC 27607, United States. Email: jlburne5@ncsu.edu

robust? Can growth mindset interventions make a meaningful impact?

This work focuses on this last question of whether growth mindset interventions are effective. Yet, this question asks for a simple verdict—one that meta-analyses in fields with clear heterogeneity in outcomes should not try to deliver. Rather, heterogeneity in effects is expected for growth mindset interventions. For example, in an influential meta-analysis examining if growth mindset interventions improve academic performance, the authors report a weak average standardized mean difference between growth mindset treatment groups and control conditions, but much larger effects for subgroups (Sisk et al., 2018). Indeed, intervention effects for students from low-SES backgrounds were four times larger than the average effect (Sisk et al., 2018). Additionally, theoretical and empirical work details the conditions under which growth mindset interventions are expected to work (e.g., Yeager et al., 2022) or not (e.g., Ganimian, 2020). Thus, drawing conclusions based on average effects is misleading, as it suggests treatment effects are constant across populations, implementation strategies, and contexts (Raudenbush & Bryk, 1985). This meaningful heterogeneity in combination with different conclusions from large-scale replications (e.g., Foliano et al., 2019; Yeager et al., 2019) suggests that more work is needed to clarify boundary conditions. Accordingly, our aim in the present study is to conduct a heterogeneity-focused and theoretically informed meta-analysis of growth mindset interventions that address remaining issues (McShane & Böckenholt, 2018; Tipton et al., 2019b).[1]

Namely, we identified seven issues that are critical to consider before making pronouncements about the value, or lack thereof, of growth mindset theory and interventions (see Table 1). The sheer number of issues highlights the complexity of trying to answer the seemingly simple question of intervention effectiveness. First, we suggest that when examining intervention impact, researchers need to consider proximate outcomes (e.g., goal persistence) that may be of value, as well as potential end results that go beyond academic performance, to avoid premature foreclosure on interventions that could be efficacious for other purposes, such as improving mental health. Second, additional efforts in cumulative syntheses must be aimed at clarifying *for whom* these interventions work best. In this meta-analysis, we investigate intervention effectiveness based on whether the intervention is delivered or analyzed based on theoretically driven sample characteristics expected to impact effectiveness (i.e., with focal groups) versus including members of a more general population (e.g., Dodge, 2020). Third, important heterogeneity could also result from wide variation in *how* the intervention is implemented. Here, we examine if implementation fidelity, delivering the intervention as intended, impacts conclusions. Fourth, in parsing heterogeneity, researchers need to understand the context, which includes the culture and environments, in which growth mindset interventions are implemented—a concept recently referred to in the mindset literature as affordances (Hecht et al., 2021). Evidence suggests that encouraging growth mindsets will improve outcomes only in a context where there are affordances, or opportunities for students' beliefs "to take root and yield benefits" (Hecht et al., 2021, p. 2). Fifth, using metaregression techniques, we extract and interpret a focal effect—the intervention effect that emerges under a fruitful combination of moderator values. This focal effect provides the best estimate for drawing conclusions, especially if this effect accounts for meaningful heterogeneity.

Sixth, to address the *why* of mindset intervention effects, researchers must consider whether the relationships among proximate and ultimate outcomes reflect the processes of change specified in the theory of mindsets. In other words, is there evidence that interventions work for the reasons we think they do? Does the evidence from intervention studies support the mediating role of the proposed cognitive and behavioral processes? These questions are critical for mindset intervention development, as mediators offer additional potential points of intervention (Miller et al., 2017). Seventh, we suggest that the field needs to consider how to evaluate effect sizes from interventions targeting outcomes that are difficult to shift (e.g., grades; rates of depression). Here, we also encourage the inclusion of a return on investment (ROI) evaluation when determining benchmarks for worthwhile interventions.

In summary, we suggest that these seven issues, if left unexamined and unaddressed, could lead the field of growth mindset interventions toward conclusions and practical decisions both premature and regrettable. The purpose of the present work is to provide scientists and practitioners with data-driven recommendations regarding the impact (or lack thereof) of growth mindset interventions, based upon a nuanced understanding of the outcomes, people, and contexts for which they are potentially effective. For example, to the extent our proposed moderators help explain heterogeneity, we can outline how to customize training in a way that delivers the right intervention to the right population in the right way. Where sufficient data do not yet exist, our approach outlines well-defined directions for future research.

Our framework includes an overarching theoretical model that outlines a taxonomy of outcomes derived from mindset perspective using a structure inspired by the organizational training literature (Kirkpatrick, 1959; see Figure 1). We term our model *mindset intervention effectiveness* (MIE) and test each path outlined in the MIE model meta-analytically when possible, and narratively when there is not adequate data to report a quantitative effect. The model is broad in scope, examining intervention effectiveness across multiple outcomes (mindsets, motivation, and goal-directed behaviors) and assorted end results (academics, mental health, and social functioning). Yet, the model is also refined in its attempts to parse the substantial heterogeneity across studies in meaningful ways that elucidate the "who, how, and why" of growth mindset interventions (see Figure 2).

We proceed as follows. First, we review growth mindset theory and research, with a focus on theoretically driven sources of heterogeneity. Next, we take a closer look at our proposed MIE model and each of the questions derived from our integration of mindset theory with the organizational training literature. Then, we turn to the methodological and analytical approach taken in this systematic review.

Growth Mindset Theory and Interventions

Growth mindset interventions are built on theory detailing how beliefs about the nature of attributes, traits, and people predict the meaning assigned to experiences (Molden & Dweck, 2006). Mindsets fall on a continuum, with stronger growth, relative to fixed, mindsets reflecting a belief that attributes and people can and do

[1] We thank the reviewers and editor for their extensive feedback and suggestions.

Table 1
Overarching Framework

Question/s	Solution/s	Meta-analytic finding/s	Result/s table path/s
1. Can interventions improve important proximal outcomes, along with desired final results, and what is the between-study heterogeneity?	Investigate effects for proximal and distal outcomes using a taxonomy of growth mindset outcomes at multiple levels that provide estimates of effectiveness at each level, including three end results of academic performance, mental health, and social functioning (see Figures 1 and 2)	Intervention effects on levels: mindsets, expectations, behavior, and all three end results. Prediction intervals	Table 2 Paths A–F
2. Are results stronger among focal analyses and populations?	Investigate whether effects of targeting interventions to specific groups and subgroups versus more general intervention approach accounts for meaningful heterogeneity	Intervention effects broken down by focal group (yes vs. no). Prediction intervals & reduction in heterogeneity	Tables 3 and 4 Path G
3. Do different implementation practices impact effectiveness?	Investigate effects of implementation fidelity by identifying intervention components and methods of implementation that are associated with the strongest intervention effects	Fidelity moderator tests. Prediction intervals & reduction in heterogeneity	Tables 3 and 4 Path H
4. In what contexts are mindsets interventions most likely to yield results? What affordances are necessary?	Investigate the psychological and systemic characteristics of contexts that are associated with stronger intervention effects	Insufficient data to analyze	Not applicable
5. What are expected intervention effects at theoretically important moderator levels?	Investigate effects of growth mindset interventions when combining all available moderator effects using metaregression	Focal effects extracted from metaregression models. Prediction intervals and reduction in heterogeneity	Table 5 Paths: (NA)
6. Why do interventions work? What are the processes of change?	Investigate links among proximal and distal outcomes to verify whether data support the theory about the cognitive and behavioral processes of change (mediators)	Correlations among mindsets, motivation, behavior, and results	Table 6 Paths: I–N
7. What is a meaningful effect?	Considering that "effect sizes are the currency in psychological research" (Schäfer & Schwarz, 2019, p. 1), we need better benchmarks for understanding applied findings	Effect sizes across all findings but especially for focal effect	NA

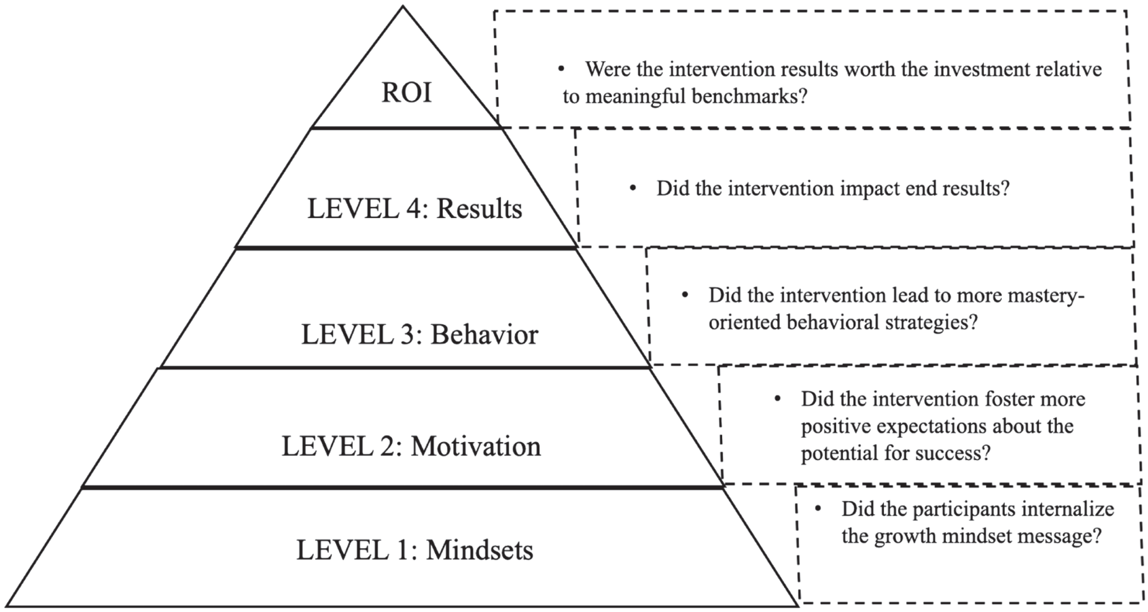
change. Across a variety of domains and contexts, growth mindsets predict motivation and self-regulatory strategies, especially when stressors arise. In turn, the psychological and behavioral processes used by those who have a stronger growth, relative to a fixed, mindset can predict goal achievement (Burnette et al., 2013). Considering these links, research investigating the potential to cultivate growth mindsets to enhance performance flourished, with the majority of early work focused on shrinking achievement gaps.

The first published mindset intervention sought to offset the deleterious effects of stereotype threat on both academic motivation and performance. African American students (and to some degree White students) in the growth mindset intervention, relative to the control condition, reported greater motivation and higher grades (Aronson et al., 2002). The second published growth mindset intervention supported the effectiveness of growth mindset interventions in improving standardized test scores, especially for female, minority, and low-income adolescents (Good et al., 2003). In the next published intervention, adolescents coping with the challenge of transitioning to middle school who were in the growth mindset intervention, relative to a control, reported greater motivation and a decreased downward trajectory in grades (Blackwell et al., 2007). The combined N for students assigned to the growth mindset condition across these three early studies is approximately 107 participants (Aronson et al., 2002, $n = 28$; Good et al., 2003, $n = 31$,[2] Blackwell et al., 2007, $n = 48$). These small sample sizes are subject to overestimating effects and inflating Type I errors.

With advances in the development of scalable online interventions, high-powered studies sought to replicate these early effects. Some of these studies found the anticipated effects, whereas others did not. For example, in a sample of close to 1,600 students, a growth mindset intervention enhanced grade point averages (GPAs) for students at risk of dropping out (Paunesku et al., 2015). Similarly, findings from another highly powered study ($N = 6{,}320$; National Study of Learning Mindsets)—which used a rigorous design including preregistration, a nationally representative sample of schools, and analyses conducted by statisticians blind to hypotheses—showed small but significant effects on academic performance for low-achieving students (Yeager et al., 2019). However, in contrast, another large-scale intervention ($N = 2{,}917$;

[2] Total, $N = 125$, and thus, with four experimental conditions having equal participants per condition, the result is approximately 31 in the growth mindset condition.

Figure 1
Training Effectiveness Model Applied to Growth Mindsets

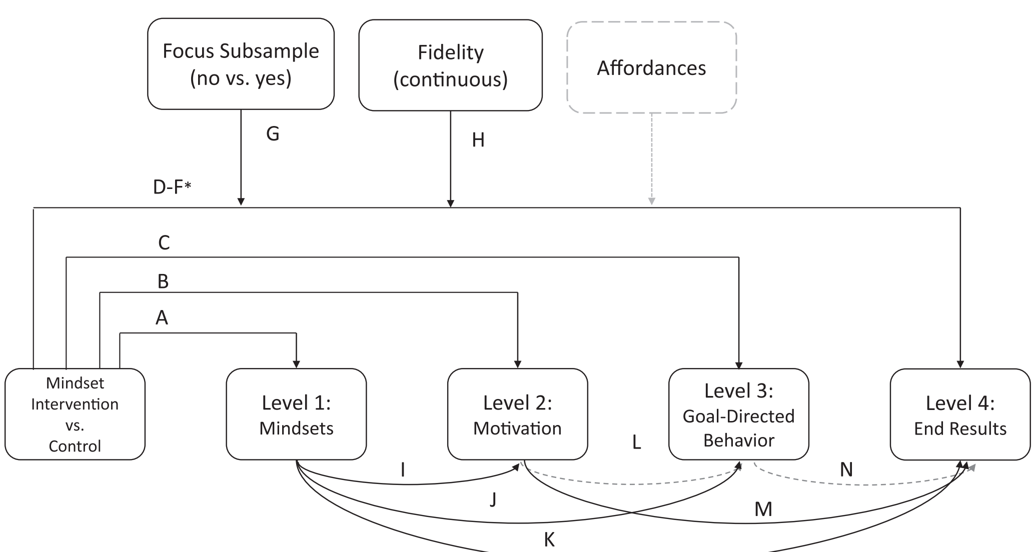


Note. ROI = return on investment.

Changing Mindsets Effectiveness Trial) failed to find significant effects, reporting an effect size of roughly zero for all academic outcomes, even among targeted subgroups (Foliano et al., 2019).

In addition to these large-scale replication efforts, conceptual tests of mindset intervention work sought to extend findings to pressing social issues that go beyond enhancing academic performance. Most of this research has tested the capacity to leverage growth mindset interventions to improve mental health and social functioning. As with academic performance, heterogeneity in findings for mental health is often the norm rather than the exception. For instance, although early growth mindset interventions reduced rates of depression for youth (Miu & Yeager, 2015; Schleider et al., 2020) and helped with anxiety (Schleider & Weisz, 2016), other work only replicated effects on reduced rates of depression for eighth graders, but not for adolescents in ninth grade (Calvete et al., 2019). When the goal is to improve social functioning, there is

Figure 2
Mindset Intervention Effectiveness (MIE) Model



* Indicates three end results of academic achievement, mental health, and social functioning.

limited work, but initial findings are fairly consistent and promising. For example, students in a growth mindset intervention, relative to a control condition, behaved less aggressively (Yeager, Miu, et al., 2013). This work extends beyond adolescents' relationships: A growth mindset intervention targeting adults promoted more positive attitudes and greater willingness to make concessions to outgroups (Goldenberg et al., 2017).

In summary, existing research on growth MIE focuses primarily on academic achievement, with mixed results. However, there are other outcomes (e.g., persistence) and end results (e.g., mental health) that are accruing enough data to examine in a cumulative synthesis. Furthermore, more work is needed to parse the heterogeneity of effects that characterize growth mindset interventions across the relevant end results—at least those with an adequate number of studies to test moderators.

Heterogeneity in Effects

A previous synthesis of growth mindset intervention effects on academic performance reported heterogeneity in outcomes, with wide prediction intervals as well as larger effects for some subgroups relative to others (Sisk et al., 2018). This is not unique to growth mindset interventions, as heterogeneity in psychological and behavioral interventions is common. For example, Rothman and Sheeran (2021) reviewed 46 meta-analyses of behavioral health interventions and found that 83% of these analyses identified significant heterogeneity across studies, which went largely unexplained despite the fact that 40 of the 46 studies tested an average of 13 moderators. Although intervention researchers may hope that moderator analyses will not be relevant to their work—that is, they hope for limited boundary conditions to the effectiveness of their intervention—investigating heterogeneity provides an opportunity to better understand for whom, how, and where interventions work, and to make more useful recommendations to practitioners and policymakers (Rothman & Sheeran, 2021). In this work, we suggest that the heterogeneity in mindset intervention outcomes can be partially explained by understanding three factors: (a) focal groups, or the desired sample that has the most potential for development; (b) implementation practices; and (c) contextual affordances.

Focal Groups

The nature of the intervention study sample is likely the most relevant contributor of heterogeneity in effects. Theory and evidence suggest that growth mindsets are most impactful when individuals feel threatened, are under stress or pressure, face uphill battles due to situational constraints (e.g., ego-threat; Burnette et al., 2013), are considered at risk (Yeager et al., 2019), or have weaker growth mindsets to start (Miu & Yeager, 2015; Yeager et al., 2022). Interventions applied to people with no particular indication of risk or vulnerability with respect to processes or desired outcomes of the intervention will probably show weaker effects. Just as people who are more at risk for a heart attack will likely benefit the most from blood pressure- and cholesterol-lowering drugs, interventions and/or analyses with subsamples of individuals with some indication that they may be vulnerable should demonstrate larger effects than interventions or analyses that highlight average effects across broad populations. The focal subgroups expected to benefit most are what the growth mindset literature customarily terms "at risk"—for example, students with indications of risks such as academic failure or students from underprivileged socioeconomic backgrounds. It is worth noting that growth mindset interventions in academic contexts are often delivered to all students to avoid stigmatizing students who may need the intervention. Thus, this moderator is often within study, rather than between studies. Overall, understanding *whom* to target in these interventions may help to explain meaningful heterogeneity.

Implementation Practices

In addition to heterogeneity due to differences in the focal subsamples, the interventions included in our analysis varied widely with respect to their implementation practices. Prior to drawing conclusions about the efficacy of growth mindset interventions, researchers should account for and optimize implementation fidelity in order to arrive at the appropriate inferences. Implementation fidelity is the degree to which an intervention is delivered as expected (Dumas et al., 2001; Dusenbury et al., 2003). In the absence of implementation fidelity, null results from intervention studies due to failure to implement can lead to the erroneous conclusion that the theoretical underpinnings of an intervention are not sound, also known as a Type III error (Dobson & Cook, 1980). Thus, a key question is as follows: How should practitioners best design and implement growth mindset interventions for maximum impact upon the designated focal groups? Unfortunately, mindset researchers have yet to provide a standardized approach for implementing these interventions across the diverse array of end results, and importantly there is not a gold-standard assessment of fidelity. Thus, we outline and test the main components related to *how* interventions are delivered. In our discussion, we note the need for the development of implementation-related tools that can help researchers to effectively design, apply, and report on growth mindset interventions across an array of end results.

Contextual Affordances

Finally, in addition to sampling and fidelity, the mindset literature and implementation science more generally (e.g., Rothman & Sheeran, 2021) highlight the crucial issue of the *contexts* in which interventions are most likely to translate into positive results. Using a "seed and soil" metaphor, Walton and Yeager (2020) describe how psychological interventions designed to help people adopt more adaptive beliefs may only be effective when the social context supports such mindsets, essentially providing a rich environment for growth. As a specific example, in a recent analysis of data from a randomized controlled trial (RCT) of a mindset intervention designed to boost academic performance, the mindsets of students' math teachers predicted the extent to which students themselves benefitted from the mindset intervention vis-à-vis improvement in their math grades (Yeager et al., 2022). Thus, variations in contextual factors influence the opportunity for changes in beliefs and behavior to "take root" and likely contribute to meaningful heterogeneity across mindset intervention studies. As is the case for focal groups and implementation practices, a better understanding of contextual "affordances" as moderators of intervention outcomes will be necessary for providing practitioners and policymakers with accurate information about for whom, how, and *in what context* mindset interventions work.

MIE Model

The above theoretical and narrative review highlights the complexity of our analysis that seeks to summarize growth MIE and emphasizes the need to provide an empirical synthesis that goes beyond average effects. To organize the expanding growth mindset intervention literature, we draw on a popular organization training model (Kirkpatrick, 1959; Kirkpatrick & Kirkpatrick, 2006) to provide a unifying framework for examining intervention effectiveness with an eye toward understanding multiple outcomes and meaningful heterogeneity. Our MIE model guides us in addressing the seven issues posed in the introduction (also see Table 1).

Issue 1: Effects Across Levels and End Results

In our MIE model, we outline the different levels as well as distinctive end results at which growth mindset interventions may exert effects. When examining the value of an intervention, the training effectiveness model recommends focusing on the desired end result or goal—Level 4—and then building back from that to understand how best to assess and operationally define the other levels. In adapting this model to examine growth mindset interventions—which are training sessions designed to teach individuals to believe in the malleable nature of human attributes—we offer an operational definition of each level (see Figure 1). At Level 4, we examine distal end results such as GPA, mental health, or social functioning—all end goals of growth mindset interventions with adequate data to examine meta-analytically. At Level 3, we focus on goal-related behaviors. For example, growth mindsets predict a willingness to exert effort despite obstacles, and such tenacity predicts greater goal achievement (e.g., Burnette et al., 2013). At Level 2, drawing on the interface between mindset theory and achievement motivation theory, we focus on a specific component of motivation: expectations for success. For example, we examine if growth mindset interventions foster stronger beliefs in the potential to learn and succeed in the future—a cornerstone of reaching goals. At Level 1, we examine if growth mindset interventions actually foster stronger growth mindsets.

In summary, as the field strives to understand the impact of growth mindset interventions, it is crucial to appreciate that the potential impact likely depends upon the *levels* and/or *goals*. Thus, our first tests examine the effect of growth mindset interventions, along with between-study heterogeneity, in effects across four levels—Level 1 (mindsets), Level 2 (motivation), Level 3 (goal-oriented behavior), and Level 4 (end results) as well as three end goals (academic achievement, mental health, social functioning; Figure 2, Paths A–F).

Issues 2–4: Moderators

In the MIE model, we propose three key theoretical moderators likely to explain variation in intervention effects upon end results with a focus on academic achievement and mental health—the social functioning literature is too nascent to test these moderators. First, we model and investigate the use of focal groups, which is typically reflected by within-study analyses that compare the magnitude of effects from targeted subsamples to the magnitude of effects from broader subsamples not expected to benefit as strongly (if at all) from the intervention (Figure 2, Path G). Second, we highlight the importance of implementation practices for understanding differences in intervention effects (Figure 2, Path H). Third, our model also includes a moderator for the context of the intervention and the extent to which it provides psychological affordances that support attitude and behavior changes prompted by the intervention (Walton & Yeager, 2020). Unfortunately, we found that there is not yet sufficient data to empirically test affordances. Finally, we note that our model limits its focus to three theoretically driven moderators outlined in the literature, but we recognize the potential impact of a great number of methodological and other study-related differences and discuss these in our sensitivity analyses as well as in the Limitations section.

Issue 5: Focal Effect

The MIE model provides a comprehensive framework for examining moderators collectively. Putting the empirical moderators together, at least those with enough data to test, brings us to what we refer to as the *focal effect.* This effect combines the "for whom" and the "how." When effect sizes represent findings from focal samples or targeted analyses conducted with those expected to benefit the most (*who*) and with powerful implementation strategies (*how*), what is the impact on end results? This focal effect describes the potential of growth mindset interventions to affect end results.

Issue 6: Mediators

The MIE model also provides a starting point for examining mechanisms of change. MIE includes an overall theoretical model of the processes that might lead to intervention effects—that is, the relationships between proximal outcomes and desired end results (Figure 2, Paths I–N). We test this model for academic outcomes, as this is where most of the work has been done and where the most data are available, although similar models are relevant to other end goals. For example, when the goal is to improve mental health, growth mindset interventions should not only enhance growth mindsets but also increase the value placed on seeking treatment (Level 2, motivation), improve help-seeking behaviors (Level 3), and ultimately reduce psychological distress (Level 4). Although for simplicity the proposed model is linear, we recognize that these paths are likely more complicated in nature.

Issue 7: Benchmarks and ROI

In the training effectiveness model that we used to integrate and summarize the growth mindset intervention literature, the top of the pyramid outlines the importance of ROI (see Figure 1). That is, when comparing intervention effects to benchmarks, do key stakeholders and policymakers find the benefits worth the costs? Intervention effect sizes can be interpreted relative to benchmarks in existing research, but comparisons can also be made with studies in related fields or with intuitively understood phenomena (Funder & Ozer, 2019; Hill et al., 2008). Importantly, in considering benchmarks, the best evidence for cumulative findings should include key moderators, rather than average effects (Cheung & Slavin, 2016). We pay particular attention to the focal effect—namely, the effect of growth mindset interventions for targeted subsamples at high fidelity for the primary end results of academic achievement and mental health.

Cumulative effects from mindset interventions can be compared to the effects from other types of interventions for similar outcomes. For example, in considering academic performance, a meta-analysis of interventions that eliminated upwardly biased effects by focusing on preregistered studies (Kraft, 2020) suggests effects are likely small, with a median effect close to $d = 0.05$. By way of further comparison, in one large study, an entire year of 55 min per school day of intensive math tutoring had an effect of $d = 0.19$–0.31 for math achievement test scores (Cook et al., 2015). Such a comparison also highlights the importance of ROI—the costs of a year of daily instruction are substantial. In considering mental health, gratitude interventions showed effects of $d = 0.17$–0.31 for well-being but much smaller effects for depression or anxiety (Davis et al., 2016; Dickens, 2017). A meta-analysis of antidepressant medication versus placebo reported effects of approximately $d = 0.20$ for mild-to-moderate depression (Fournier et al., 2010). Additionally, a meta-analysis of the effects of social media on depression in adolescents reported an estimated average effect of $r = 0.11$ ($d = 0.22$), but with high heterogeneity ($I^2 = 95.22\%$; Ivie et al., 2020). For social functioning comparisons, a meta-analysis of the effects of video games on aggressive behavior was $r = 0.21$ ($d = 0.45$; Burkhardt & Lenhard, 2022). A meta-analysis of intervention programs designed to reduce aggressive behavior reported an effect of $g = 0.60$, whereas the same meta-analysis reported an effect of $g = 0.23$ for interventions designed to promote prosocial behavior (Mesurado et al., 2019).

Overall, benchmarks for academic outcomes range from $d = 0.05$ on the conservative side to $d = 0.31$ for highly intensive types of teaching. For mental health, especially when trying to reduce or explain psychological distress, anxiety, and depression, effects of low-intensity interventions are, at best, around $d = 0.20$–0.30 or often smaller. For social functioning-related outcomes, effects are slightly larger when the goal is to reduce aggression but a bit smaller when the goal is to encourage prosocial behavior. Considering the typical low cost, both in terms of time and money (Yeager et al., 2019), of implementing growth mindset interventions as well as the practical value of improving grades, mental health, and social functioning, especially in populations that need it most, a Cohen's *d* effect that ranges from 0.10 to 0.20 likely falls within a reasonable benchmark for expected effects. A detailed ROI analysis should depend on scientists' and policymakers' own objectives and evaluations. The goal here is to encourage a discussion of these issues, not to make an ultimate pronouncement about meaningful effect sizes.

MIE Model Paths

The integration of mindset theory with the training effectiveness literature fills several gaps in existing syntheses of mindset research, allowing us to ask both broad and nuanced questions related to growth MIE. We build on this integration to develop the MIE model, which outlines each path we investigate in this meta-analytic review. First, across levels and end results (Paths A–F in Figure 2), we examine the average effect. However, critically, we also ask, what is the between-study heterogeneity in these effects? Is there an indication that important boundary conditions exist?

Path A: Mindsets

This path depicts the most proximal outcome and represents intervention effects on growth mindsets, which is often termed a manipulation check in psychological research. Are participants internalizing the key message of the intervention? Despite the importance of knowing if the intervention impacts mindsets, surprisingly, not all intervention studies[3] include this assessment (64% of studies examined intervention impact on mindsets). When growth mindsets are assessed, evidence for the potential to foster growth mindsets is promising. For example, in a large representative sample of lower achieving students in the United States (number of schools = 65, $N = 6{,}320$), an intervention strengthened growth mindset beliefs relative to the control condition, with a standardized mean difference of 0.33 (Yeager et al., 2019). Level 1, mindsets, are a critical component to assess as it indicates responsiveness to the materials as well as internalization of the message. Thus, if training fails at this level, the intervention is unlikely to impact more distal outcomes including the desired end goal. As such, we first ask: *Do interventions foster stronger growth mindsets, and what is the heterogeneity in those effects?*

Path B: Motivation

In the growth mindset literature, motivation is often assessed in terms of cognitive processes (such as a belief that exerting effort is valuable) and attributions regarding the meaning of failure (Dweck & Yeager, 2019). Mediators derived from mindset theory also point to the importance of self-regulatory processes such as goal setting (i.e., learning, rather than performance-focused goals) as well as goal monitoring which includes beliefs about one's potential to learn and succeed in the future (i.e., positive expectations; Burnette et al., 2013). Although a range of social-cognitive mechanisms is relevant, we focus on expectations, as this is the primary motivation-related variable assessed in the intervention studies available for inclusion in the meta-analysis. Of the included studies, 23% included an assessment (12 of 53 studies) related to positive expectations. For example, Burnette, Pollack, et al. (2020) assessed self-efficacy in the domain of entrepreneurship, using undergraduate students in an entrepreneurship course. Growth mindset messaging can enhance such expectations, as it implies that everyone has the capacity to learn and improve despite potential challenges. For example, following failure feedback, students with growth mindsets remain confident in their skills and capacity to reach future goals, whereas students with stronger fixed mindsets question their abilities and potential to succeed on future tasks (Dweck, 2000). Furthermore, expectations predict the relevant end results of growth mindset interventions such as academic performance and mental health (Burnette, Pollack, et al., 2020; Sriram, 2014). In summary, growth mindset interventions could also foster stronger beliefs in

[3] To minimize confusion, we try to be as precise as possible in using the related terms *article*, *study*, and *independent sample*. An *article* refers to an article, thesis, dissertation, or unpublished article that contains at least one study that meets the inclusion criteria. A *study* is a systematic analysis of a growth mindset intervention that meets the inclusion criteria. A given study may include one or more than one independent sample. An independent sample is defined as a unique set of participants from which one or more effect sizes are calculated. There are times when there is more than one study in a given article, or an article reports on multiple outcomes (e.g., both mindsets and end results) and thus contributes multiple effects.

one's future potential. Here, we ask: *Do growth mindset interventions impact participants' expectations and with how much heterogeneity?*

Path C: Goal-Directed Behavior

In the organizational training literature, this outcome refers to how well individuals are applying the training in terms of behavioral changes—sometimes called transfer. The behaviors that the training should modify directly align with the desired end result and are thus referred to as goal-directed behavior. For example, if the end goal is to improve academic performance, researchers should anticipate that students are putting more effort into studying. Yet, this behavioral component is infrequently tested in social psychology in general (Baumeister et al., 2007), and in growth mindset interventions specifically (42% of studies included this Level 3 outcome). Whether observed or assessed via self-report, goal-related behavior is a critical level at which to examine if growth mindset interventions are effective.

Correlational and experimental data point to various types of behaviors that help individuals reach their end goal. Individuals with growth, relative to fixed mindsets, are more likely to report mastery-oriented coping in the face of challenges, described as an overall "hardy response" (Dweck, 2000, p. 6). For example, in laboratory studies, students are more likely to engage in remediation following setbacks (Nussbaum & Dweck, 2008), and to increase their practice time (e.g., Cury et al., 2008). However, other work examining a general self-report measure of goal persistence fails to find a relation with growth mindsets of intelligence (Burgoyne et al., 2020). In contrast, within a health context, individuals with substance use problems exposed to a growth mindset, relative to a disease-fixed message, reported being more likely to engage in counseling and cognitive behavior therapy (Burnette et al., 2019). Examples of goal-directed behavior from included studies comprise outcomes such as seeking academic challenges in the form of preferring harder versus easier questions to answer (Bettinger et al., 2018; Rege et al., 2021), favoring complex versus easy tasks (Skipper, 2015), or taking a full course load the first year of college (Yeager, Walton, et al., 2016). Another mastery behavior frequently assessed was task persistence, whether as self-report in the domain of entrepreneurship skills (Burnette, Pollack, et al., 2020) or as observed number of questions attempted on a standardized test covering multiple domains of knowledge (Chao et al., 2017). And, as illustrated by the above examples, behavior is primarily assessed, thus far, in interventions designed to improve academic performance. Here, we ask: *Do growth mindset interventions impact participants' goal-directed behaviors, and what is the degree of heterogeneity?*

Paths D–F: End Results

In the growth mindset intervention literature, the end result or goal is typically focused on improved academic performance (68.1%), mental health promotion (17.0%), or social functioning (14.9%). In our model, we have three different end results, each with unique benchmarks and different policy implications. Thus, we report results separately to help the reader draw conclusions about effectiveness related to each end goal. Here, we ask: *What is the average effect of growth mindset interventions on academic performance, mental health, and social functioning, as well as the degree of variation in effects for each end result?*

Path G: Focal Groups

Importantly, these questions of average effects are expected to exhibit a large degree of heterogeneity. Thus, we also ask, for *whom* do growth mindset interventions work? Although there are theoretical reasons to expect that targeting individuals and groups expected to benefit most creates meaningful boundary conditions of effectiveness at all levels, there is limited inquiry regarding focal groups or analyses at Levels 1–3. There are also reasons to expect that the answer to the question of who will benefit the most from interventions might depend upon what level of outcome we are examining. For example, Rege et al. (2021) predicted that an academic mindset intervention would promote challenge-seeking motivation for all students but would only have an effect on GPA for lower achieving students. Due to limited data and different theorizing depending on the level of the outcome, we focus on moderation by the focal group for end results—namely academic performance and mental health, as the data are too nascent for social functioning. If there is moderation for these two end results, follow-up research can, and should, investigate where else in the chain moderation effects may also occur (Rothman & Sheeran, 2021).

Focal groups are those for whom a growth mindset intervention is excepted to work best. Because this is often postulated a priori and is often a within-study moderator, we note that focal group also includes focal subsamples or analyses within a particular study. In academics, focal groups have included students at risk of failing, students from disadvantaged backgrounds, those with lower growth mindsets to start, and sometimes some combination of these. For mental health, focal groups have included participants who had experienced previous stress or trauma, who had low growth mindsets to start, or who had been hospitalized for psychiatric reasons. Here, we ask: *Are the effects of growth mindset interventions on academic achievement and mental health stronger for focal groups, relative to non-focal groups? Does this moderator significantly reduce between-study heterogeneity, and if so, by how much?*

Path H: Implementation Fidelity

Next, we ask, *how* might interventions work? What are the most powerful ways to deliver a growth mindset message? A central aim of our work is to highlight the importance of implementation fidelity for understanding differences in intervention effectiveness. Despite the significance of fidelity, there is not yet an agreed-upon standard of implementation outlined in the growth mindset literature, and practices vary widely. The primary implementation practice in the first three growth mindset interventions was to provide scientific evidence for the malleability of intelligence (e.g., learning helps the brain make new connections), and to invite students to reflect on and internalize the growth mindset message—a technique called "saying is believing" (e.g., Aronson et al., 2002). Nonetheless, even the saying is believing technique is not consistently implemented, and the use of best practices for message delivery varies, as does control over the delivery of the key message. All of these components can be aggregated to get a sense of overall implementation fidelity. Other implementation strategies vary, such as dose (both number of sessions and time spent delivering the message). However, because there is not an agreed-upon gold-standard implementation approach for these additional components, we did not include this as part of our overall fidelity score. Here, we outline the three key components that go into our continuous overall fidelity moderator: adherence, delivery competence, and message control.

Adherence. Adherence means including "core components" that constitute an intervention and is often defined as delivering the intervention as prescribed (Dane & Schneider, 1998). To be considered a growth mindset intervention, there must be a message that the attribute is malleable. Yet, to help make this attribute stick, interventions also often include an attitude change tactic such as a saying is believing activity. For example, this often involves writing a pen-pal letter to a struggling student—a note that stresses the opportunity for growth and development despite challenges. Having participants put the growth mindset message into their own words helps to instill the mindset belief. In the present work, 18.9% of interventions failed to include this type of attitude change tactic. Interventions that are implemented with this component and adhere to best practices are likely more impactful. Thus, those that included the saying is believing type of exercise received a score of 1, and those that did not, received a score of 0.

Delivery Competence. Delivery competence relates to the skillful delivery of the intervention content. Experts in mindset research delineate several components that should improve the delivery of growth mindset interventions, most of which are delineated in the context of mindsets of intelligence and academic performance. These components for mindset interventions are as follows: (a) include neuroscience information, (b) use credible sources, (c) provide scientific evidence of the potential for change, (d) respect autonomy (e.g., ask the participant to collaborate), (e) incorporate social norms and social modeling, (f) include content to avoid blame and encourage self-compassion, (g) make it personally relevant and (h) incorporate metaphors that make the message "sticky" (Yeager, Romero, et al., 2016). In developing the coding procedures for delivery competence, we arrived at five categories that could be coded in the existing data. First, was scientific evidence (including neuroscience) for malleability provided? Second, were social norms, peer norms, and/or social/role modeling used? Third, did the intervention include content designed to reduce blame and/or encourage compassion? Fourth, did the intervention contain content related to conveying benefits of growth mindsets and/or personal relevance of content? Fifth, did the intervention contain metaphors or activities designed to be sticky and memorable? Each of the five categories was coded "0" for absent, "1" for present, with the five responses then summed. No study received a summed score of 5, with values ranging from 0 to 4 ($Mdn = 2$).

Message Content Control. This variable assesses the level of control that the research team had over the curriculum and content of the intervention delivered to the desired end recipients (i.e., individual students). Values range from 3 (*full researcher control*) to 1 (*modest researcher control*). Interventions delivered directly to students, whether via computer (e.g., Paunesku et al., 2015) or in-person by members of the research team (e.g., Mills & Mills, 2018), represent full-delivery control, but other intervention formats involve intermediaries and are characterized by varying degrees of researcher control over the curriculum and message. In such cases, we distinguish levels of content control based upon the detail and structure of the curriculum that is specifically prepared for individual students. For example, Rienzo et al. (2015) student intervention exhibited substantial control: They provided their assistants with a detailed curriculum to be executed in the course of six 2-hr workshops, each with a specific mindset-relevant theme. In contrast, the intervention conducted by Good et al. (2003) exhibited relatively modest researcher control: mentors met with students in two unstructured sessions of 90-min each, and interacted via email, without specified pedagogical objectives. In short, greater researcher control over intervention content should improve fidelity, and lead to stronger outcomes. The range was 1–3 with a median score of 3.

Overall, these three components (adherence, delivery competence, and message content control) are scored to produce a continuous assessment of fidelity that ranged from 0 to 8 with a mean of 6.0 and a standard deviation of 1.45. Here, we ask: *Are the effects of growth mindset interventions on academic achievement and mental health stronger when interventions are implemented with better fidelity? Does implementation fidelity significantly reduce between-study heterogeneity in effects, and if so, to what extent?*

Focal Effects

Combining Paths G–H allows us to examine what we call our focal effect, which captures intervention effectiveness for end results when researchers incorporate focal groups (*who*) and use powerful implementation strategies (*how*)? Here, we ask: *What is the estimated effect, the prediction interval, and the reduction in heterogeneity relative to average effects for end results, when the intervention is implemented with focal groups and with high fidelity?*

Paths I–N: Potential Mediators

Why might growth mindset interventions work? Theoretically, mindsets should directly and indirectly predict academic achievement by influencing certain motivational processes (Dweck, 2000). The organizational literature notes that the direct association with Level 4, the end result, is usually not as strong as other levels that are more proximal to the intervention. And, importantly, each level should impact the final result, but effects should be stronger as the process gets closer in the chain to the end result. That is, Level 1 should have the strongest effects when considering intervention effectiveness, but when examining correlations among constructs, Level 3, goal-directed behavior, should have the strongest correlation with the end result, Level 4. If we find that this is indeed the case, conclusions about effectiveness, or lack thereof, can be informed. That is, if these interventions strongly shift growth mindsets and this, in turn, relates to motivation, which can impact behavior, which in turn correlates with the end results, then investments may be worth it. This example is purely serial, but there are a number of ways in which these processes may work. Here, we explore correlations among the levels.

Summary

We sought to provide a comprehensive empirical overview of effects of growth mindset interventions with a focus on meaningful heterogeneity. To examine the questions raised by the integration of the growth mindset approach with the training effectiveness literature, we provided an overarching theoretical framework. The MIE model and our proceeding review outlined each path that we propose and test (see Figure 2 and Table 1). We now turn our attention to the methodological and analytical approach and our findings related to the paths outlined in the model.

Method

Search Strategy and Inclusion Criteria

We conducted an initial search using the following electronic databases: ERIC, APA PsycINFO, ProQuest Dissertation and Theses, as well as Google Scholar. Search terms included various combinations of the keywords (e.g., "lay theories," "implicit theories," "mindset," "growth mindset," and "intervention"; see Figure 3). We also conducted a legacy search (i.e., "back-tracking" an article by its references to identify potentially useful articles) and a forward search on included articles. For the forward search, we used APA PsycINFO and Google Scholar to determine if any newer articles that had cited our included articles met the inclusion criteria of our meta-analysis. To obtain unpublished and in-press articles, we sent a request to the listserv for the Society for Personality and Social Psychology. Our search started with the original intervention (Aronson et al., 2002) and concluded with articles published before the end of 2020.

In assessing abstracts and full-text articles, five inclusion criteria had to be met. First, the study had to be a test of a growth mindset intervention designed to foster stronger beliefs in the malleability of an attribute. Thus, correlational or longitudinal studies that lacked an intervention (e.g., Canning et al., 2019) were excluded, as were theoretical overviews (e.g., Yeager & Dweck, 2012), articles reporting the development of a scale (e.g., Limpo & Alves, 2014), and lab-based studies that featured an experimental manipulation rather than an intervention in an applied setting (e.g., Da Fonseca et al., 2008). Second, eligible studies had to use random allocation and feature a comparable control group. Studies that did not—for example, Stern (2015) and Stevens (2018)—were excluded. Third, the intervention needed to be designed to foster a growth mindset and could not include adaptations that made it impossible to disentangle the effect of the malleability message from the effect of ancillary content unique to the treatment group. For example, Binning et al. (2019), which featured an intervention that taught a belonging mindset together with a growth mindset, was excluded, whereas Paunesku et al. (2015) were not excluded because the researchers used separate experimental conditions to introduce content beyond the growth mindset message (i.e., sense of belonging)—thus enabling us to isolate the growth mindset effect. Fourth, the intervention goal had to be to enhance academic outcomes, improve mental health, or seek to enhance social functioning. Here, we excluded studies with the goal of improving outcomes such as sports performance (Shaffer, 2014), outdoor personal development (O'Brien & Lomas, 2017), and physical health (Burnette & Finkel, 2012). Fifth, the authors had to report an effect size reflecting one of the primary outcomes (i.e., mindsets, expectations, goal-directed behavior, and end results), or information needed to compute this effect size had to be available (either in the article or from the authors). Overall, regarding these inclusion criteria, the vast majority of articles excluded from the meta-analysis were eliminated because they were not an intervention.[4] See Figure 3 for exclusion funnel.

Coding Strategy—Moderators

To code our moderators, the first and second authors began by coding the three chronologically oldest interventions—Aronson et al., 2002; Blackwell et al., 2007; Good et al., 2003—to refine the coding scheme. After doing so, both authors then independently coded the next three articles in alphabetical order. Cohen's κ was .89 across the moderator coding, indicating good interrater reliability. The remaining coding was done by the second author.

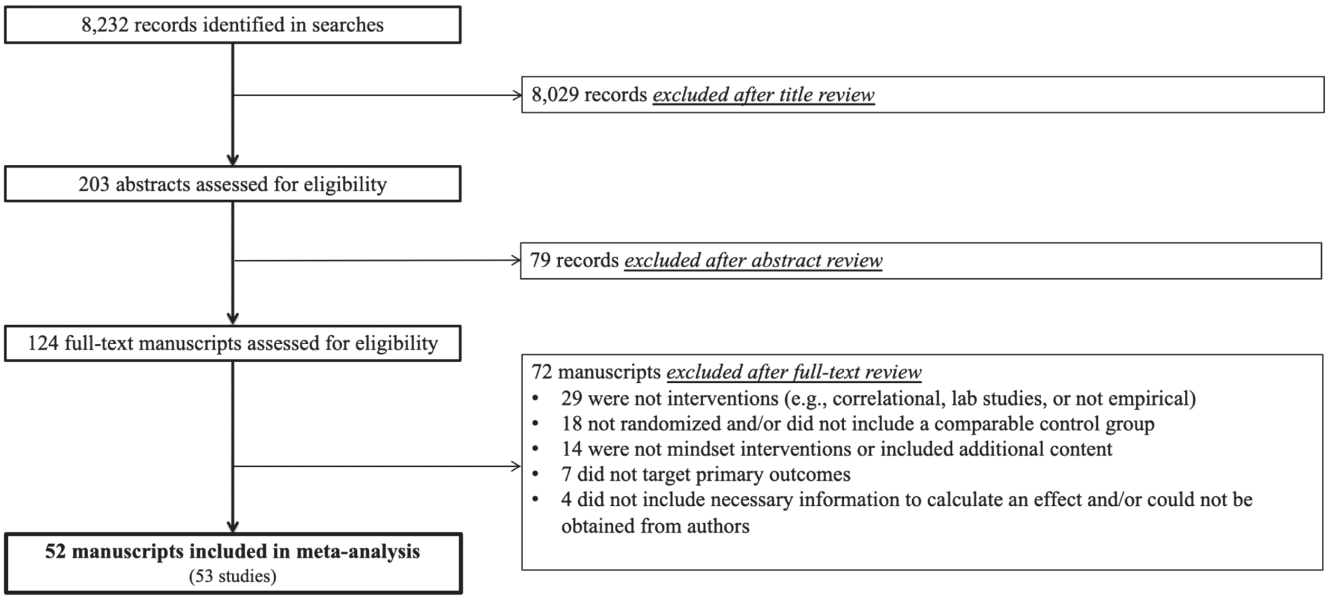
Focal Groups

The focal group moderator combines at risk and other indicators of a targeted sample for which effects are expected to be pronounced—that is, individuals for whom the intervention is expected to be most efficacious. The effect is coded as 0 (not focal) when the specified sample or subsample analysis is not expected to benefit most. The effect is coded as 1 (focal) when one of two conditions is met: (a) The entire study sample is considered to be targeted by the intervention or (b) the particular subsample featured in the analysis that produces the effect size is considered to be targeted by the intervention. We note that this latter condition occurs frequently in the data set because growth mindset interventions are often administered to most or all available students—even those not expected to benefit—in order to avoid stigmatizing treated students. In this latter case, the authors outlined a priori why they tested a particular individual difference as a moderator, and the effect in question represents an effect at the level of the moderator where the effect is expected to be present or especially strong. For example, Aronson et al. (2002) hypothesized that their growth mindset intervention would impact African American/Black students more strongly than White students, due to the challenges posed by stereotype threat. Thus, effects for the African American/Black students were coded "1" for focal group, while effects for White students were coded "0" for not the focal group. Similarly, Yeager, Walton, et al. (2016; Study 2) hypothesized that disadvantaged college students (specifically, those who were first-generation college students or from a disadvantaged ethnic minority) would benefit most from the growth mindset intervention. Thus, effects involving disadvantaged students were coded "1" for the focal group, while effects for other students were coded "0" for not focal. In some cases, the full sample for a study may be coded as the focal group, when study authors make such targeting clear. Alan et al. (2019, p. 1151) did so, for instance, when they reported that "[w]e deliberately target low-SES students for whom interventions of this type have been shown to be most effective." However, we do not code entire samples as focal when the author's rationale for doing so is based solely on ambiguous transitions or categories because such groups represent massive heterogeneity and lack theoretical clarity. For example, it seems unlikely that *all* students who live in rural areas of the South should be considered at risk (e.g., Burnette et al., 2018). We return to the complicated and nuanced issue of defining focal subgroups in the discussion.

Despite the theoretical assertion that growth mindsets are most influential in times of challenges, use of focal groups in analyses or populations was limited (57.4% of studies with a codable end

[4] See *OSF* for reference list of all studies included in the meta-analysis and for exclusion file detailing the reasons for exclusion.

Figure 3
Search Funnel for Meta-Analysis



result), and definitions of these challenges vary. Furthermore, reporting of the statistics necessary to pull apart simple effects was not always reported or available upon request from the authors (14.9% of studies with an end result have known missing simple effects involving moderation). In the discussion, we outline the need for greater theoretical precision as well as transparency in reporting of a priori versus exploratory analyses, specifically as it relates to "for whom" the intervention should be most effective. We also discuss the importance of reporting all simple effects.

Implementation Fidelity

For moderation by implementation fidelity, we created a continuous score that represents best implementation practices. This total score included three subcomponents: *adherence* (0 or 1); *delivery competence* (0–5); *message control* (1–3). Higher numbers indicate greater fidelity, with scores ranging from 2 to 8. Mean fidelity across all interventions was 6.0 (SD = 1.44). For the academic interventions, mean implementation fidelity was 5.78 out of 8 (SD = 1.58). Mental health interventions exhibited mean implementation fidelity of 6.13 (SD = .64), while interventions that targeted social functioning featured an average implementation fidelity of 6.14 (SD = 1.35). Further descriptive statistics pertinent to implementation fidelity are available in Online Material Table 1 (OM1).

Meta-Analytic Procedure

We conducted a random-effects meta-analysis to synthesize the primary samples and establish estimates of the population parameters (Borenstein et al., 2009; Hedges & Olkin, 1985). We report Cohen's d as the primary effect size estimate representing the difference in outcome between intervention and control groups in each study; hence, we converted other types of effect size information, such as correlations, odds ratios, as well as means and standard deviations, to Cohen's d. For Table 2, we report the number of effects, samples, and studies, the N, Cohen's d, the standard error, 95% confidence intervals (CIs), and τ^2 as well as prediction intervals to indicate the degree of heterogeneity (Borenstein et al., 2009). In Table 3 and Table 4, for moderators, we report the estimates, standard errors, 95% confidence intervals (CIs), τ^2, variance accounted for (VAF), and the prediction intervals. In Table 5, for the focal effects, we report the estimated effect at the specified moderator values, as well as the corresponding standard error, 95% confidence intervals for the estimate, and 95% prediction intervals (PIs). In Table 6, for the correlations among constructs, we report the summary correlation coefficient, along with its standard error, 95% confidence intervals, τ^2, and 95% prediction intervals.

Interdependency in the Data

There are a number of interdependencies in these data, including multiple effects derived from the same author across studies, multiple levels or outcomes within the same study (e.g., mindsets and results), as well as multiple waves of assessments and a host of operationalizations of the same outcome (e.g., mindsets assessed with different measures). The first two issues are inherent in our process. That is, we include multiple effects from the same authors and multiple effects across relevant outcomes (e.g., mindsets and end results). However, to address the remaining issues related to independencies in the data, we adopt best practices as recommended by Tipton et al. (2019a, 2019b). Specifically, we include *all* relevant effects from a given study, rather than picking single effects or computing an average effect for each study (which may bias results). By retaining all relevant effects, we take advantage of crucial within-study variation. For example, whenever possible, we include separate effects from subsamples pertinent to our key within-study moderator, targeting of focal groups.

Table 2
Multilevel Effects on Each Outcome

Outcome	Effects	Samples	Studies	*N*	*d*	*SE*	95% CI	τ^2 (Sum)	95% PI
Growth mindsets	49	37	34	30,403	0.46	0.06	[.34, .58]	0.105	[−0.19, 1.10]
Expectations	18	12	12	62,987	0.18	0.10	[−0.02, 0.38]	0.108	[−0.49, 0.85]
Mastery behaviors	38	22	22	58,305	0.18	0.05	[0.10, .27]	0.030	[−0.17, 0.54]
Academic performance	110	48	32	51,676	0.09	0.02	[0.05, 0.13]	0.012	[−0.13, 0.30]
Mental health	28	10	8	2,529	0.18	0.06	[0.06, 0.29]	0.019	[−0.12, 0.47]
Social functioning	36	8	7	659	0.36	0.17	[0.03, 0.68]	0.164	[−0.50, 1.22]

Note. Effects equals the number of total effects reported. This represents both multiple outcomes and timing. *Samples* equals the number of subsamples—this is larger than studies because it captures within-study heterogeneity of targeting specific focal subgroups. *Studies* equals the number of studies (a test of intervention effectiveness across independent samples). This does not match the total number of articles included in the search figure because some articles report more than one study and because some articles only report one or two of the outcomes or levels. The nesting structure is such that effects are nested within samples that are nested within studies. τ^2 (Sum) reflects true study variance. To calculate the sample size *N*, due to the multilevel context, we averaged within studies; for details, please see the R code on the Open Science Framework (OSF) Project Page. *SE* = standard error; 95% CI = 95% confidence interval; 95% PI = 95% prediction interval.

To account for the data dependency introduced by the inclusion of multiple relevant effects from each study, we employ multilevel metaregression using restricted maximum likelihood (REML) estimation. Namely, we nest effect sizes within independent samples within studies, using the metafor package in R (Viechtbauer, 2010). We report all models for growth mindsets, expectations, goal-directed behaviors, academic performance, mental health, and social functioning in Table 2 without moderators and using this nested structure. The same nesting structure was also used to calculate the Paths I–N reported in Table 6. In Tables 3 and 4, we used the same nesting structure, but added moderators in a stepwise fashion. We first included a basic model which reports the intercept and focal group moderator. In Model 2 we added the Fidelity moderator. Consequently, we compute and evaluate VAF, which reflects the percent change in estimated true study variance, or τ^2 (Sum), relative to the unconditional model in Table 2. Code is available at the Open Science Framework (OSF; Burnette et al. 2022).

Effect Size Extraction and Variance Calculations

Full details regarding the extraction, computation, and/or conversion of each effect size and its associated variance are available in the datafile posted at *OSF*. A brief overview of pertinent decisions that could have an impact is elaborated on below—note, we did not preregister the search protocol or decisions outlined below, and some decisions are a result of multiple rounds of reviewer feedback.

In terms of extracting effects, there are three key decisions worth noting. First, we excluded the effects from Rienzo et al. (2015) Study 2 because students were not the target of the intervention. Rather, the authors noted with respect to their teacher-targeted intervention (Study 2):

> How teachers put this learning into practice was at their discretion. The training was designed to give teachers an understanding of the general approach and specific techniques, which they could then apply to teaching and learning as they thought was appropriate. (p. 11)

This was excluded as a pure "train the teacher model" with no additional materials provided to help implement the growth mindset message.[5] Second, we chose not to include effects related to end results when they were not the primary goal of the intervention—what we termed *side effects*. If, for example, a article's stated goal was to improve mental health, but the authors also happened to report an effect for academic performance, we only included the primary end result—namely, mental health. The effect on academic performance was excluded as a side effect. This resulted in the exclusion of the three effects reported in Yeager et al. (2014) that pertained to academic achievement (i.e., Yeager et al., 2014), as well as a total of four effects on depression, one in Yeager, Miu, et al. (2013) and three in Schleider et al. (2020). This decision was made based on mindset theory and the organizational training framework we incorporated throughout. Interventions should be designed with the primary end goal in mind and other outcomes are considered secondary. As can be seen by the small number of studies, most work only focused on one stated end goal.

Third, there is the issue of intent-to-treat analyses (ITT) versus the local average treatment effect (LATE). If both ITT and LATE were clearly reported in the main text, we included only LATE, although such instances were rare (e.g., Outes-Leon et al., 2020; Porter et al., 2020). The reason is that LATE, or similar approaches, should capture the effect under more idealized circumstances, as it is the effect of the treatment for those participants who complied. We note that most studies reported ITT or similar procedures or failed to report their approach—unless LATE was clearly reported, we included the primary effect the authors do report. These three decisions, especially because they are not preregistered, could be seen as biased. To help avoid this concern, we also ran analyses with all these effect sizes included. Results from these additional sensitivity analyses are discussed below and more details regarding the excluded effects can be found on OSF.

Overall, in terms of computing effects, where possible, we culled reported effect sizes directly from the original report, converting them if necessary to the metric of Cohen's *d* using the methods of Lenhard and Lenhard (2016) as implemented in the effect size calculator at https://www.psychometrica.de/effect_size.html. When effect sizes were not directly reported, but the relevant statistical analysis was provided in the text, we computed Cohen's *d* from the available results (e.g., *t* or *F* statistics), again using the methods of Lenhard and Lenhard (2016). When the relevant statistical analysis

[5] We excluded these effects based on reviewer feedback. We retained the student intervention (Study 1) from Rienzo et al. (2015) but not the teacher intervention (Study 2).

Table 3
Moderators of Multilevel Effects for Academic Performance

Model	Moderator	Estimates	*SE*	95% CI	τ^2 (Sum)	VAF	95% PI[a]
Model 1	Intercept	0.04	0.03	[−0.01, 0.09]	0.009	25%	[−0.15, 0.24]
	Focal group: yes	0.10	0.04	[0.02, 0.18]			
Model 2	Intercept	0.04	0.03	[−0.01, 0.09]	0.010	17%	[−0.16, 0.25]
	Focal group: yes	0.10	0.04	[0.02, 0.18]			
	Fidelity	0.00	0.01	[−0.03, 0.02]			

Note. VAF = variance accounted for; *SE* = standard error; 95% CI = 95% confidence interval; 95% PI = 95% prediction interval. The nesting structure is such that effects are nested within samples that are nested within studies. τ^2 (Sum) reflects true study variance. VAF indicates variance accounted for, and quantifies the percent reduction in heterogeneity achieved relative to the unconditional model in (Table 2). Fidelity is median-centered based on descriptives reported in Online Material Table OM1. Focal group is coded such that not a focal subsample = 0 and subsamples or analyses examining focal groups = 1. For example, the intercept represents the effect size for *not* focal group and the regression coefficient represents the difference between not a focal and yes a focal group. Thus, the sum of the two (e.g., Model 1: 0.04 + 0.10 = 0.14) equals the effect for those anticipated to benefit most.
[a] This PI corresponds to the prediction interval for the intercept only, and thus represents the expected range of effects when the intervention does not have a focal group and is conducted with median fidelity.

was not reported, we computed Cohen's *d* from means, standard deviations, and group *n*'s. In cases where none of the above information was available, we requested information directly from the authors. In all cases, we adjusted the sign of Cohen's *d* such that a positive value reflects a result in accord with the expected direction of effects. Thus, a positive *d* in the context of academics indicates that the intervention improved performance. Likewise, a positive *d* in the context of mental health indicates that the intervention improved mental health, even if the actual reported outcome was a *reduction* in anxiety or depression, as predicted by theory. Obviously, if a negative value represented an actual effect counter to expectations, it remained negative.

For associated variances, we took steps to ensure that our procedures properly accounted for the clustering exhibited in the design of many primary studies in the data set. For studies that employed multilevel analyses, where possible we converted cluster-robust standard errors or *p* values to variances using the approximation recommended by Borenstein and Hedges (2019), or by Altman and Bland (2011). For studies that did not employ multilevel analyses, we computed the variance using the standard formula provided by Borenstein and Hedges (2019). We note that for some primary studies, the data exhibited multilevel structure, but the study authors did not employ multilevel techniques in their analysis. In such cases, we abjured the use of corrective techniques (e.g., Borenstein & Hedges, 2019), given that the information necessary to do so was almost always lacking.

Heterogeneity Indices

To assess heterogeneity, we focus our reporting on two statistics: τ^2 and the 95% prediction interval. The statistic τ^2 captures the variance in true effects—that is, the variance not attributable to sampling error (Viechtbauer, 2010). The nested structure of our data (effects within samples within studies) entails that there are two sources of true-effect variance: the sample level and the study level. To simplify our presentation, we sum these two sources of variance to provide a single τ^2 value (Konstantopoulos, 2011). To quantify the effectiveness of our moderators, we compute the percent reduction in τ^2 achieved by the conditional model, relative to the unconditional model—what we term VAF.

Based upon the dispersion of true effects, 95% prediction intervals can be calculated. These prediction intervals provide a range within which we can expect future true-effect sizes from randomly selected populations to fall (Borenstein & Hedges, 2019). Prediction intervals are distinct from the 95% confidence intervals for the mean effect size: Whereas the latter only provides information about the precision with which the mean effect size is estimated, the former addresses heterogeneity in the overall distribution of true effects.

Table 4
Moderators of Multilevel Effects for Mental Health

Model	Moderator	Estimates	*SE*	95% CI	τ^2 (Sum)	VAF	95% PI[a]
Model 1	Intercept	0.05	0.02	[0.02, 0.09]	0.000	100%	[0.02, 0.09]
	Focal group: yes	0.23	0.05	[0.14, 0.33]			
Model 2	Intercept	0.09	0.05	[−0.01, 0.18]	0.004	79%	[−0.06, 0.24]
	Focal group: yes	0.18	0.09	[0.01, 0.35]			
	Fidelity	0.03	0.06	[−0.10, 0.15]			

Note. VAF = variance accounted for; *SE* = standard error; 95% CI = 95% confidence interval; 95% PI = 95% prediction interval. The nesting structure is such that effects are nested within samples that are nested within studies. τ^2 (Sum) reflects true study variance. VAF indicates variance accounted for, and quantifies the percent reduction in heterogeneity achieved relative to the unconditional model in (Table 2). Fidelity is median centered, based on the descriptives reported in Online Materials Table OM1. Focal group is coded such that not a focal subsample = 0 and subsamples or analyses examining focal groups = 1. For example, the intercept represents the effect size for *not* focal group and the regression coefficient represents the difference between not a focal and yes a focal group. Thus, the sum of the two (e.g., Model 1: 0.05 + 0.23 = 0.28) equals the effect in subsamples expected to benefit most.
[a] This PI corresponds to the prediction interval for the intercept only, and thus represents the expected range of effects when the intervention does not have a focal group and is conducted with median fidelity.

Table 5
Isolating Focal Intervention Effects in the Context of Moderation

Focal effects[a]	Estimate	*SE*	95% CI	95% PI
Academic results	0.14	.04	[.06, .22]	[−.08, .35]
Mental health	0.32	.11	[.10, .54]	[.07, .57]

Note. *SE* = standard error; 95% CI = 95% confidence interval; 95% PI = 95% prediction interval.
[a] The focal effect refers to the effect for an intervention that is directed to a *focal subgroup* and is of *high*-implementation fidelity (scoring 8 out of an observed range of 8). Note that the median fidelity score is 6. For more details on calculations, see R code on Open Science Framework (OSF).

We extract prediction intervals from our statistical models using the predict function in the metafor package for R (Viechtbauer, 2010). We note that in the context of metaregression, prediction intervals are available for each combination of moderator values. We, therefore, report prediction intervals for both the intercept of each model (all moderator values = 0) and a "focal effect" (described subsequently) that reflects moderator values of theoretical import.

Sensitivity Analyses: Additional Moderators

In line with past meta-analyses, best practices, and reviewer suggestions, we also examined if conclusions are robust when including additional moderators. The primary question here is whether effects (i.e., the focal effect) are robust to methodological rigor in the underlying studies, which we labeled "research quality." Research quality refers to features of the study related to best practices. Because quality determinations are so variable among scientists and given the limited evidence regarding how each component of research quality alone, or in combination, may impact results, it is not recommended practice to use research quality as part of the selection criterion (e.g., Johnson, 2021). Rather, we use it as part of our sensitivity moderation analyses. Namely, we focus on two broad areas: best practices for transparency and replication (i.e., open science) and best practices for RCTs. For open science, we focused on the use of intervention reporting standards, open access materials, and incorporating preregistration. For RCT practices, we included the intervention control group type and blinding. We used these subcomponents to create a continuous variable of research quality (range = 0–9). More detailed coding information pertaining to research quality is available in the coding guide located on OSF, as is a descriptive summary table (Online Material Table OM1).

Within all the noise associated with research quality, our moderator analysis may yet detect signal, but it is not entirely clear what the expected direction of the interaction effect should be. The most plausible case can probably be made that stronger effects are likely to emerge in conjunction with lower research quality, where for example, lack of preregistration permits researchers "opportunistic degrees of freedom," which might inflate the rate of false positive results as well as effect sizes (Simmons et al., 2011; Wicherts et al., 2016). Furthermore, reducing demand characteristics likely reduces treatment variance and can make effects seem smaller. Alternatively, researchers who are using best practices are also likely those engaging in best implementation practices more generally (e.g., stricter adherence, better delivery strategies) and thus effects for studies with higher research quality may be stronger. Indeed, research quality is positively correlated with fidelity in the present work ($r = 0.35$, $p < .001$). Then, there is also of course the possibility that research quality, with all the noise and potential confounds, has no meaningful impact on effect sizes. Thus, this moderator is not included in our theoretical model but is nonetheless included as a moderator in sensitivity analyses.

Furthermore, reviewers asked us to also include publication status (yes vs. no) and the simple binary breakdown of control group type (active vs. no treatment) as additional moderators. In our first sensitivity analysis, we included our main moderators (focal group and fidelity) as well as publication status and research quality. In our second sensitivity analysis, to avoid overlap, yet examine the binary categorical variable of active versus no-treatment control, we ran another analysis with our two primary moderators (focal group and fidelity) and publication status, while substituting the control group type moderator for research quality.

Sensitivity Analyses: Using Selection Models to Investigate Publication Bias

In a further set of sensitivity analyses, we focus specifically on the threat of publication bias, in which findings from published studies differ systematically from findings from the entire

Table 6
Mediators: Links Between All Levels

Relationship	Effects	Samples	Studies	*N*	*r*	*SE*	95% CI	τ^2 (Sum)	95% PI
Mindsets and									
Expectations	4	3	3	958	.25	0.10	[0.06, 0.44]	0.024	[−0.11, 0.61]
Behavior	18	5	5	26,406	.20	0.08	[0.06, 0.35]	0.026	[−0.15, 0.56]
Results	24	12	11	3,044	.17	0.04	[0.08, 0.25]	0.016	[−0.10, 0.43]
Expectations and									
Results	4	3	3	958	.05	0.18	[−0.29, 0.39]	0.088	[−0.63, 0.73]

Note. Effects equals the number of total effects reported. This represents both multiple outcomes and timing. *Samples* equals the number of subsamples—this is larger than studies because it captures within-study heterogeneity of targeting. *Studies* equals the number of studies (a test of intervention effectiveness across independent samples). This does not match the total number of articles included in the search figure because some articles report more than one study and because some articles only report one or two of the outcomes or levels. The nesting structure is such that effects are nested within samples that are nested within studies. τ^2 (Sum) reflects true study variance. r = observed correlation coefficient; *SE* = standard error; 95% CI = 95% confidence interval; 95% PI = 95% prediction interval.

population of the completed studies, particularly the unpublished literature (Vevea et al., 2019). A number of popular techniques exist to identify and correct for publication bias—including, for instance, the oft-employed trim and fill method (Duval & Tweedie, 2000). Popular techniques such as the trim and fill method, however, suffer from two key shortcomings: (a) they have not been extended to a multilevel context and (b) they do not address publication bias in the context of moderator analyses. To address these shortcomings, and given our focus on heterogeneity, we employed a mixture of alternative approaches that allowed us to retain a focus on moderation, while simultaneously accounting for our multilevel data structure as much as possible.

The core of our sensitivity analysis is the use of selection models, as developed and implemented by Vevea and Hedges (1995) and Vevea and Woods (2005). We utilize these models because, unlike other techniques, they offer the crucial advantages of analyzing publication bias in the context of moderation (Vevea et al., 2019), and of remaining robust to heterogeneity (McShane et al., 2016), which is expected in our data set. Fundamentally, these selection models "adjust meta-analytic data sets by specifying a model that describes the mechanism by which effect sizes may be suppressed" (Vevea et al., 2019, p. 396). In the models examined here, the suppression mechanism is assumed to be meaningful cutoffs for p values. Our sensitivity analyses involve two such models.

Our two selection models derive from the methods of Vevea and Woods (2005). With these models, the researcher provides the p value cut points of interest together with a hypothetical set of weights that reflects a possible suppression mechanism. Using this approach, we construct one selection model with p value cut points and weights that correspond to "moderate" publication bias, and a second model with cut points and weights that correspond to "severe" publication bias, as demonstrated in Vevea et al. (2019). Results from these two models provide the means and variance components that would emerge under those conditions, allowing us to determine the extent to which effects (including moderated effects) remain, or do not remain, robust.

Despite the manifold advantages offered by the selection models described above, the methodology has not yet been elaborated in a multilevel context. We address this shortcoming by following the precedent of Dworkin et al. (2017) and utilize a random sampling approach to assemble a data set of independent observations. While such an approach has the clear merit of producing independent observations in a way that avoids systematic bias, results may be sensitive to the specific sample set that is actually selected. To partially mitigate this limitation, we randomly select one effect from each independent sample, construct the two aforementioned selection models using the assembled data set, and extract the desired mean and variance component estimates. We repeat this process multiple times to obtain three sets of parameter estimates (one from each randomly selected data set), then use the mean of each set as our final point estimate for each parameter. Publication bias analyses are conducted in R, with selection models generated using the weightr package (Coburn & Vevea, 2019).[6]

Sensitivity Analyses: Model Specification

Finally, no one model completely accounts for all dependency in the data, and assumptions underlying the models are rarely met perfectly. In light of this and recent developments in best practices related to multilevel and multivariate meta-analytic techniques, we also ask if conclusions are robust to different techniques for modeling data dependency. To address this question, we draw upon recent refinements to research synthesis methodology, based upon the use of *robust variance estimation* (RVE; Hedges et al., 2010; Pustejovsky & Tipton, 2022; Tipton & Pustejovsky, 2015). A key advantage of the RVE framework is that it allows researchers to make reliable, statistically valid inferences from metaregression even when the effect sizes in the model are characterized by interdependency (Pustejovsky & Tipton, 2022). RVE achieves this by providing alternative 95% confidence intervals and standard errors (see Hedges et al., 2010, for details). Furthermore, recent extensions of this framework provide a specific working model—the "correlated and hierarchical effects" model, or CHE—that is explicitly designed to accommodate data sets such as ours, where interdependency among effect sizes is complex and largely unknown but is marked by both cluster effects and correlations within clusters due to multiple outcome measures and repeated observations over time (Pustejovsky & Tipton, 2022). Accordingly, and following the procedures outlined by Pustejovsky and Tipton, we conduct a sensitivity analysis in which we use the CHE model as a working model that is then analyzed using the alternative standard errors and confidence intervals provided by robust variance estimation. As argued by Pustejovsky and Tipton, this approach serves as a valuable technique to guard against misspecification of multivariate meta-analytic models.

Transparency and Openness

We have posted all data and code on OSF in addition to other files to help with transparency. There are three folders, the first of which is the Data, Coding, and Script Folder. This folder contains 10 csv files that can be imported into R to reproduce the major analyses reported below, an R script with labels for each of the corresponding tables in the article, and a coding document that provides detailed criteria for coding decisions. The second folder is titled "Sensitivity Analyses" and includes the R scripts and additional datafiles needed to run all sensitivity analyses, as well as a table that summarizes the results derived from our selection models. The third folder is titled "Online Materials Folder"; it includes three Excel files that were used to elaborate on the nature of the studies in terms of populations (e.g., age, country of origin, sex), as well as publication status, type of control group, and additional details

[6] Selection models for some randomly selected data sets failed to converge. In addition, we attempted to generate a third selection model based on the methods of Vevea and Hedges (1995) in which the researcher specifies a meaningful interval (or set of intervals) of p values, from which the software calculates weights that represent the relative likelihood of survival for an effect size that falls within the given interval. The software then generates a random-effects model that is adjusted based upon the weighted likelihoods. Using this approach, researchers thus obtain not only bias-corrected mean and variance components but also weighting parameters that quantify the relative likelihood of effect sizes appearing in each interval. This approach requires that researchers specify the relevant p value intervals based on theoretical understanding of the suppression mechanism and that each interval contains at least some observed p values. Because our data set is relatively small, we specified only one interval for this model, opting for a cut point of $p = .025$, which reflects the positive tail in a two-tailed significance test with α at .05. Even with this limited set of intervals, however, the data were inadequate for our specified model (with moderators) to converge.

provided related to the studies. We separated these files in order to obtain study-level characteristics as the data analysis files report multiple effects across a given study. Given the interdependent nature of the data and effects, these distinct files provided individual study information. Additionally, in this third folder, we include an Excel file with descriptive information concerning the implemental fidelity and research quality of included studies, an Excel file documenting all excluded studies, and a document containing the references to all studies included in the meta-analysis.[7]

Results

Descriptive Analysis

Before moving into the findings (broken down by path in Figure 2), we contextualize the results by outlining descriptive features of the primary studies included in analyses. We first outline the number of effects for each level. Next, we include information related to the population, intervention, and context.[8] We then include information related to methodological features of the studies. The details are broken down by end result. For additional details on the populations, intervention, comparison groups, and outcomes (PICO), see Table 7.

Altogether, 52 articles supplied 55 studies, which resulted in 53 distinct interventions[9] testing intervention effectiveness, with 279 effects in total across the outcomes. In terms of descriptive information for each of our outcomes and analyses (Table 2), for Levels 1–3, growth mindsets were reported in 34 studies, expectations in 12 studies, and goal-directed behaviors were tested in 22 studies. In 47 of the 53 distinct interventions, an effect size relevant to an end result was provided—academic performance (32 studies); mental health (8 studies); or social functioning (7 studies).

As for descriptive information, for academic achievement, in terms of samples, 97% are drawn from Western, educated, industrialized, rich, and democratic (WEIRD) nations—those that are Western, educated, industrialized, rich, and democratic (Henrich et al., 2010). Here, 70% are from the United States, with 55.3% women on average, and 62% of the sample were primary or high school aged. In terms of research design, 75.7% used an active control group and 73% randomized at the individual level. The majority of the studies (78.4%) were conducted in the last 5 years and 78.4% were published. As for features related to the intervention delivery, information was not uniformly available regarding key characteristics of the interventions used in the analytic sample. Yet, information related to dose, in terms of sessions and duration, was more routinely accessible. Most academic interventions consisted of relatively few sessions (the median number was two), and were relatively brief (the median duration was 75 min), attesting to the comparatively low cost of most growth mindset interventions. In these interventions, all but three interventions focused on fostering stronger growth mindsets of intelligence (91.2%).

For mental health, in terms of samples, 100% were WEIRD and 87.5% were from the United States, with 54.4% women on average, and 87.5% of the sample were youth—namely, middle or high school aged. In terms of research design, 100% used an active control group, and 100% were randomized at the individual level. 62.5% of the studies were conducted in the past 5 years, and 100% within the last 10 years, showing the novelty of leveraging growth mindsets to improve mental health. Additionally, of those studies examining mental health as the primary outcome, 75% were published. As for features related to the dose, the median number of sessions was one, and the median intervention duration was 26 min. The majority of these interventions (87.5%) focused on mindsets of personality with one intervention seeking to foster stronger growth mindsets of emotion (Smith et al., 2018).

For social functioning, in terms of samples, 100% were WEIRD, and 71% were from the United States, with 55.6% women on average. 28.6% of the samples were youth—namely, middle or high school age, with adults representing 57% of the sample. In terms of research design, 100% of studies used an active control group, and 85.7% were randomized at the individual level. Forty-three percent of the studies were conducted within the last 5 years, with 71.4% published articles. As for features related to the intervention dose, the median number of sessions was one, and the median intervention duration was 90 min. All of these interventions focused on fostering stronger growth mindsets of personality.

In terms of the research quality of included studies, we created a continuous variable (range = 0–9), that included components related to transparency and best practices for testing intervention effectiveness. For the academic interventions, mean research quality was 3.72 (out of 9; $SD = 1.73$). Mental health interventions exhibited a mean research quality of 5.38 ($SD = 1.19$). Interventions that targeted social functioning featured an average research quality of 4.29 ($SD = .76$). In terms of reliability, when reported, reliability was generally good. Although, for academic performance outcomes (e.g., GPA) there is no reliability information, and for mental health outcomes, there were limited data available (25% of studies). For mental health, the average reliability (Cronbach's α) was .89, with an emphasis on assessments of depression (e.g., Beck et al., 1996). For studies designed to improve social functioning, psychometric information was more readily available: Of these, 77% of effects involved a measure that was accompanied by Cronbach's α. The average reliability for these outcomes was .83 (Helmreich & Stapp, 1974).

Question 1: Paths A–F (Table 2)

We begin by presenting findings at each of the levels of our taxonomy and for each end result. For the first three outcomes of mindsets, expectations, and goal-directed behavior, we include all relevant studies, regardless of whether the ultimate goal of the intervention was to improve academic performance, mental health, or social functioning. We made this decision as the vast majority of studies including these outcomes focused on improving academic performance and because such processes are similar. However, for end results, we breakdown effects by end goal. For Paths A–F, the effects are averages, but we are especially interested in indicators

[7] For questions or additional details related to coding or analyses, please contact the second and third authors.

[8] Details regarding these demographic and study-level descriptives may be found on OSF in the descriptives subfolder.

[9] We found two instances of the same interventions tested in the same sample with the same end result reported in different articles (Heslin et al., 2005, 2006; Schleider & Weisz, 2016, 2018). For these duplicates, to extract descriptive statistics related to each end result, we went with the first published article. Additionally, there was a duplicate intervention with partial overlap between the samples (Rege et al.'s, 2021, Study 1 and Yeager et al., 2019). For the partial duplicate, we went with the larger, more inclusive sample for the purpose of descriptive analyses.

Table 7
PICO Table

P	I	C	O
Population	Intervention	Comparison	Outcomes
What are the characteristics of the populations included in analyses?	What is the intervention?	What are the primary comparison groups?	What are the possible outcomes?
We provide descriptive information related to the samples broken down by end results: Academic achievement: 70% United States 55.3% female 45.3%White 17.5 mean age Mental health: 87.5% United States 54.4% female 43.2% White 13.9 mean age Social functioning: 71% United States 55.6% female 49.3% White 28.5 mean age	• Growth mindsets are defined as beliefs about the malleable versus fixed nature of human attributes, abilities, and traits. • We only included growth mindset interventions. For an intervention to be considered a growth mindset intervention, the study authors had to report seeking to foster stronger growth mindsets. • We did not include other types of interventions that did not target growth mindsets or that included blended approaches or confounds such as those that fostered stronger growth mindsets along with a sense of belonging.	We included randomized designs with active, attention matched, and no-treatment control groups. Here, we provide descriptive information related to type of control condition broken down by the end result: Academic achievement: No-treatment control: 9/37 = 24.3% Active control: 28/37 = 75.7% Mental health: No-treatment control: 0/8 = 0% Active control: 8/8 = 100% Social functioning: No-treatment control: 0/7 = 0% Active control: 7/7 = 100%	We included six possible outcomes: 1. *Mindsets,* defined as beliefs about the malleable versus fixed nature of human attributes, abilities, and traits. 2. *Expectations,* defined as beliefs about the potential for future success and confidence in one's abilities. 3. *Goal-directed behavior,* defined as persistence and mastery-oriented behaviors such as help-seeking in the pursuit of goals. 4. *Academic achievement,* defined as grades, GPA, test scores, or other school-related performance outcomes. 5. *Mental health,* defined as well-being and the absence of psychological distress such as anxiety and depression. 6. *Social functioning,* defined as an individual's ability to engage in interpersonal relationships in constructive rather than destructive ways.

Note. PICO = populations, intervention, comparison groups, and outcomes; GPA = grade point average.

of their heterogeneity—namely τ^2 and the 95% prediction intervals. We do not interpret effect sizes in the Results section—rather, we defer this to the discussion when we outline the focal effect, which includes the key moderators. Similarly, we do not visually illustrate the average effect in the form of a forest plot[10]; instead, we defer graphical representation of results until the section on moderation, where we illustrate key findings using the moving constant technique (Johnson, 2021).

Mindset (Path A)

For the intervention effect on growth mindsets, findings indicate a standardized mean difference estimated to be $d = 0.46$, 95% CI [.34, .58]. On average, the standardized mean difference is statistically greater than zero, but τ^2 and the 95% prediction intervals indicate substantial heterogeneity in this effect (see Table 2).

Motivation (Path B)

For the intervention effect on positive expectations, findings indicate a standardized mean difference of $d = 0.18$, 95% CI [−.02, .38]. On average, the standardized mean difference is not statistically distinguishable from zero. Also, the τ^2 and the 95% prediction intervals indicate substantial heterogeneity in this effect (see Table 2).

Goal-Directed Behavior (Path C)

For the intervention effect on goal-directed behaviors, the summary effect size is $d = 0.18$, 95% CI [.10, .27]. On average, the standardized mean difference is greater than zero, but τ^2 and the 95% prediction intervals indicate heterogeneity in this effect (see Table 2).

Academic Performance End Results (Path D)

For the intervention effect on academic achievement, the summary effect size is $d = 0.09$. Given the moderately large number of effects and the large *N*, this summary effect is estimated with a fair degree of precision: 95% confidence intervals range from 0.05 to 0.13. However, the τ^2 and prediction intervals indicate heterogeneity in this effect (see Table 2).

Mental Health End Results (Path E)

For the intervention effect on mental health, the summary effect is $d = 0.18$, 95% CI [.06, .29]. Given the small number of effects and *N*, this summary effect should be interpreted with caution. Additionally, both τ^2 and prediction intervals indicate substantial heterogeneity in this effect (see Table 2).

[10] Except in the case of social functioning, where we did not conduct moderation analyses due to the low number of studies.

Social Functioning End Results (Path F)

For the intervention effect on social functioning, the summary effect is $d = 0.36$, 95% CI [.03, .68]. Given the small number of effects and N, this summary effect should be interpreted with caution. Additionally, both τ^2 and prediction intervals indicate massive heterogeneity in this effect (see Table 2). Due to the lack of sufficient studies to test for moderation for this outcome, we present effect sizes in a forest plot (see Figure 4).

Questions 2–4 Moderators: Paths G–H (Tables 3 and 4)

Here, we tested the moderators for which we had sufficient data: focal groups and implementation fidelity. We present the effects as well as the reduction in heterogeneity when each moderator is added to the model. As described in the Method section, we quantify the reduction in heterogeneity as percent decrease in tau-squared achieved by the conditional model (with moderators), relative to the unconditional model (without moderators). Thus, if the τ^2 value for the unconditional model were .50, for example, and the τ^2 value for the conditional model was .25, the reduction in heterogeneity would be 50%.

First, for academic performance, in Model 1, Table 3, running analyses with focal groups improves the estimate by $d = 0.10$. Thus, the effect is $d = 0.14$ for academic performance, and this moderator accounts for a 25% reduction in heterogeneity relative to the unconditional model (see the VAF column in Table 3). In Model 2, Table 3, fidelity adds little in terms of the estimate of effect or to a reduction in heterogeneity—this is counter to expectations.

Second for mental health, in Model 1, Table 4, focal groups improve the estimate by 0.23. Thus, the effect is $d = 0.28$ on mental health in Model 1, and the focal groups moderator accounts for virtually all heterogeneity relative to the unconditional model. In Model 2, Table 4, effects are slightly larger with fidelity but it does not result in a reduction in heterogeneity. Overall, targeting focal groups is a critical moderator for understanding effects and explaining heterogeneity for both academic performance and mental health. Yet, for mental health outcomes, we suggest caution in drawing broad conclusions because of the small number of studies included.

Question 5: Focal Effect (Table 5)

For the focal effect, we outline the effect with moderators at particular values. Namely, we ask what is the mean effect of the intervention on end results when focal groups are analyzed, and implementation fidelity is high? For academic performance, the mean expected effect under these conditions is $d = 0.14$, 95% CI [.06, .22]. The 95% prediction intervals range from −.08 to .35. For mental health, the mean expected effect under these conditions is $d = 0.32$, 95% CI [.10, .54]. The 95% prediction intervals range from .07 to .57.

Given our emphasis on this focal effect model, in keeping with current best practices (Johnson, 2021), we eschew presentation of results using a forest plot. Instead, we visually represent key results for academic achievement and mental health using the moving constant technique (Johnson & Huedo-Medina, 2011), which is recommended for depicting results from models with multiple moderators. This model visually depicts the focal effect sizes at varying levels of fidelity and for targeted versus not-targeted subgroups for both academic end results (see Figure 5) and mental health (see Figure 6).

Question 6 Mediators: Paths I–N (Table 6)

As shown in Table 6, we estimated four pathways: (a) mindsets to motivation operationalized as expectations; (b) mindsets to goal-oriented behavior; (c) mindsets to academic end results; and (d) expectations to academic end results. We could not test expectations to goal-directed behavior, nor goal-directed behavior to end results as there were insufficient data. Findings indicate that, on average, there is an association of growth mindset with expectations, $r = .25$, 95% CI [.06, .44], goal-directed behavior, $r = .20$, 95% CI [.06, .35], and academic end results, $r = .17$, 95% CI [.08, .25]. The τ^2 values and 95% prediction intervals indicate a great deal of heterogeneity in these effects (see Table 6). There is no evidence for a link between expectations and end results on average, $r = .05$, 95% CI [−0.29, 0.39]. Both τ^2 and the 95% prediction intervals indicate heterogeneity in these effects (see Table 6). In the discussion, we outline the importance of exploring boundary conditions of each of these links as well as the need for more work investigating different operationalizations of motivation and additional assessments of goal-directed behaviors.

Sensitivity Analyses

Given the relative lack of data on mental health and social functioning, our sensitivity analyses are focused on academic performance.

Moderation by Publication Status, Research Quality, and Type of Control Group

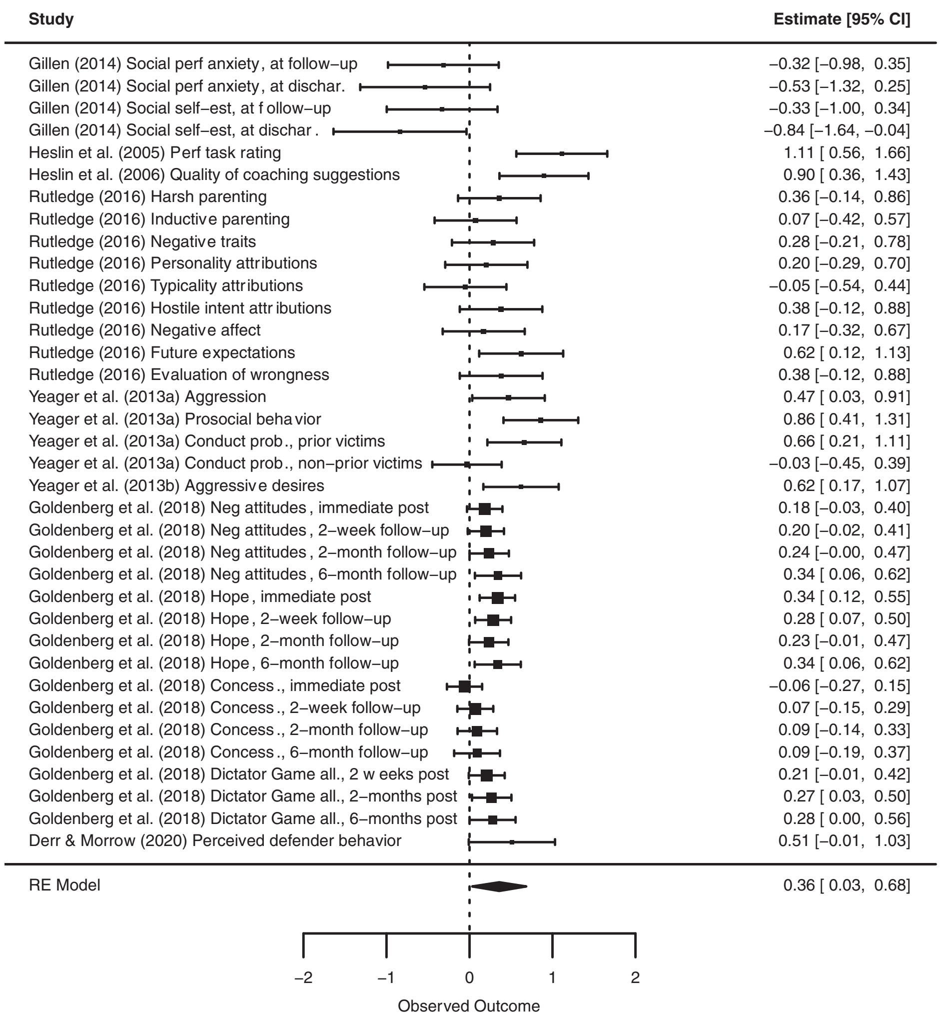
To assess the sensitivity of our conclusions to variation in publication status, research quality, and type of control group used, we constructed two metaregression models. In the first (Model S1), we introduced publication status and research quality as moderator variables, along with the primary moderator variables—focal group and implementation fidelity—reported earlier. In the second model, we simply substituted type of control group for research quality, given that information about the type of control group was a component of the research quality variable. The predictors in the second model were thus focal group, implementation fidelity publication status, and control group type. As noted earlier, these analyses are limited to effects involving academic results.

Model S1: Examining Publication Status and Research Quality as Potential Moderators. Results from this model indicated that neither publication status (published vs. unpublished) nor research quality (ranging from a low of 1 to a high of 9) significantly moderated intervention impacts on academic performance, controlling for focal group, and implementation fidelity.[11,12] Controlling for research quality and publication status (as well as for implementation fidelity), the use of a targeted focal group remained a

[11] Parameter estimates for publication status: $b = -0.034$, 95% CI [−0.14, 0.07], $SE = .053$, $p = .523$. To further explore publication status as a possible moderator, we ran the same metaregression model as above, but omitting the research quality variable. The conclusion was unchanged: $b = -0.051$, 95% CI [−0.15, 0.05], $SE = 0.050$, $p = .310$.

[12] Parameter estimates for research quality (which was median-centered): $b = -0.021$, 95% CI [−0.05, 0.01], $SE = .015$, $p = .153$. To further explore research quality as a possible moderator, we ran the same metaregression model as above, but omitting the publication status variable. The conclusion was unchanged: $b = -0.023$, 95% CI [−0.05, 0.01], $SE = 0.014$, $p = .104$.

Figure 4
Forest Plot for Social Functioning



Note. RE indicates that the summary effect size estimate is produced from a random effects model.

significant moderator of the intervention effect on academic performance, $b = 0.13$, 95% CI [0.04, 0.22], $SE = .046$, $p = .005$.

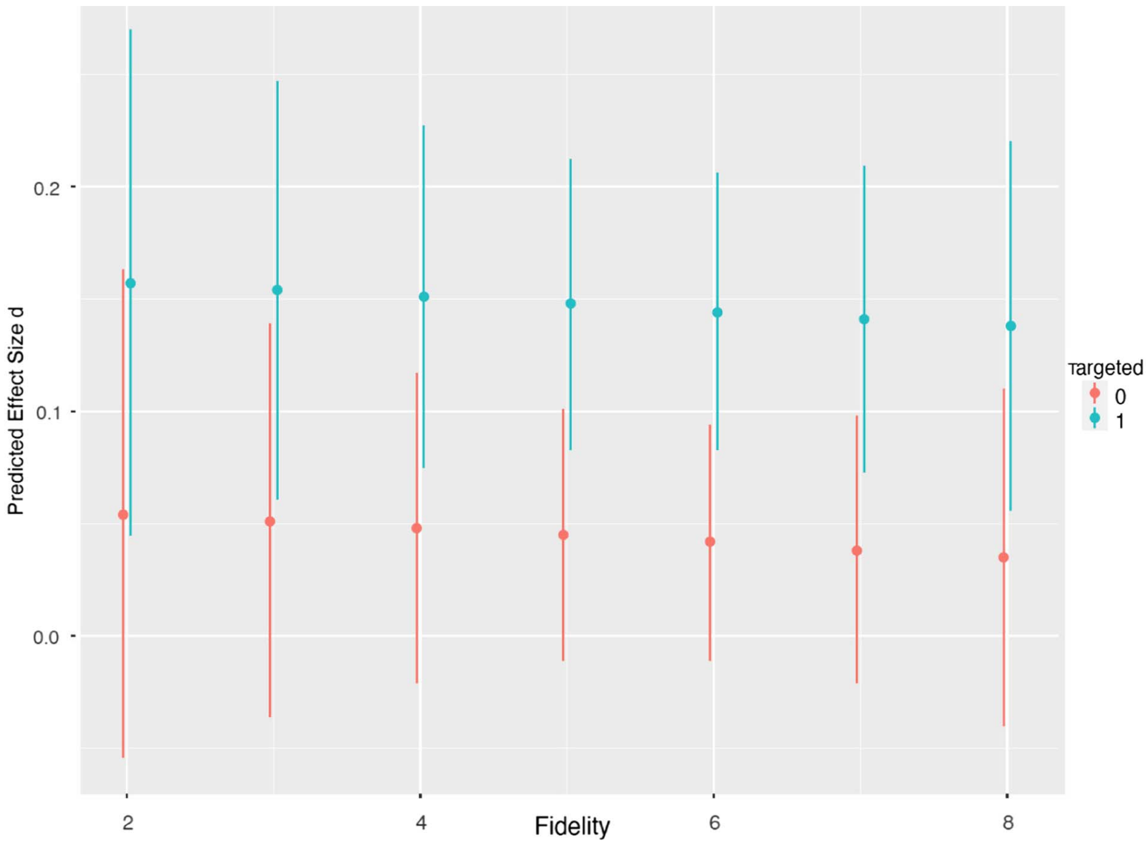
Model S2: Examining Control Group Type as a Potential Moderator. Results from this model indicated that type of control group (passive vs. active) did not significantly moderate intervention impacts on academic performance, controlling for focal group, implementation fidelity, and publication status.[13] In this model, publication status was nonsignificant as well ($p = .683$). Controlling for type of control group and publication status (as well as for implementation fidelity), the use of a targeted focal group remained a significant moderator of the intervention effect on academic performance, $b = 0.11$, 95% CI [0.03, 0.20], $SE = .044$, $p = .010$.

Using Selection Models to Investigate Publication Bias

Results obtained from our two selection models (Vevea & Woods, 2005) can be found on our OSF page. Using conventional criteria for evaluating the impact of publication bias (Kepes et al., 2014), the moderating effect of using focal subsamples is robust to both moderate and severe publication bias. Indeed, the data suggest that, if anything, the effect of targeting particular analyses with focal groups might actually be a bit obscured rather than inflated, if publication bias infects the literature. Because this moderating effect remains robust to publication bias, the focal effect sizes that we emphasize throughout the main text—effect sizes for academic interventions delivered to focal groups and with high-implementation fidelity—are not extremely sensitive to publication bias.

[13] Parameter estimates for type of control group: $b = -0.048$, 95% CI [−0.17, 0.08], $SE = .063$, $p = .443$. To further explore the type of control group as a possible moderator, we ran the same metaregression model as above, but omitted the publication status variable. The conclusion was unchanged: $b = -0.061$, 95% CI [−0.16, 0.04], $SE = 0.051$, $p = .234$.

Figure 5
Metaregression for Academic Performance



Note. See the online article for the color version of this figure.

Model Specification

We focus here on assessing the sensitivity of the results previously depicted in Table 3, which demonstrated that the intervention effect on academic performance was significantly moderated by the focal group while controlling for implementation fidelity. That earlier moderation model estimated the intervention effect to be approximately 0.14 *SD*s on average. The difference in the intervention effect between focal versus nonfocal group, moreover, was estimated to be 0.10 *SD*s.

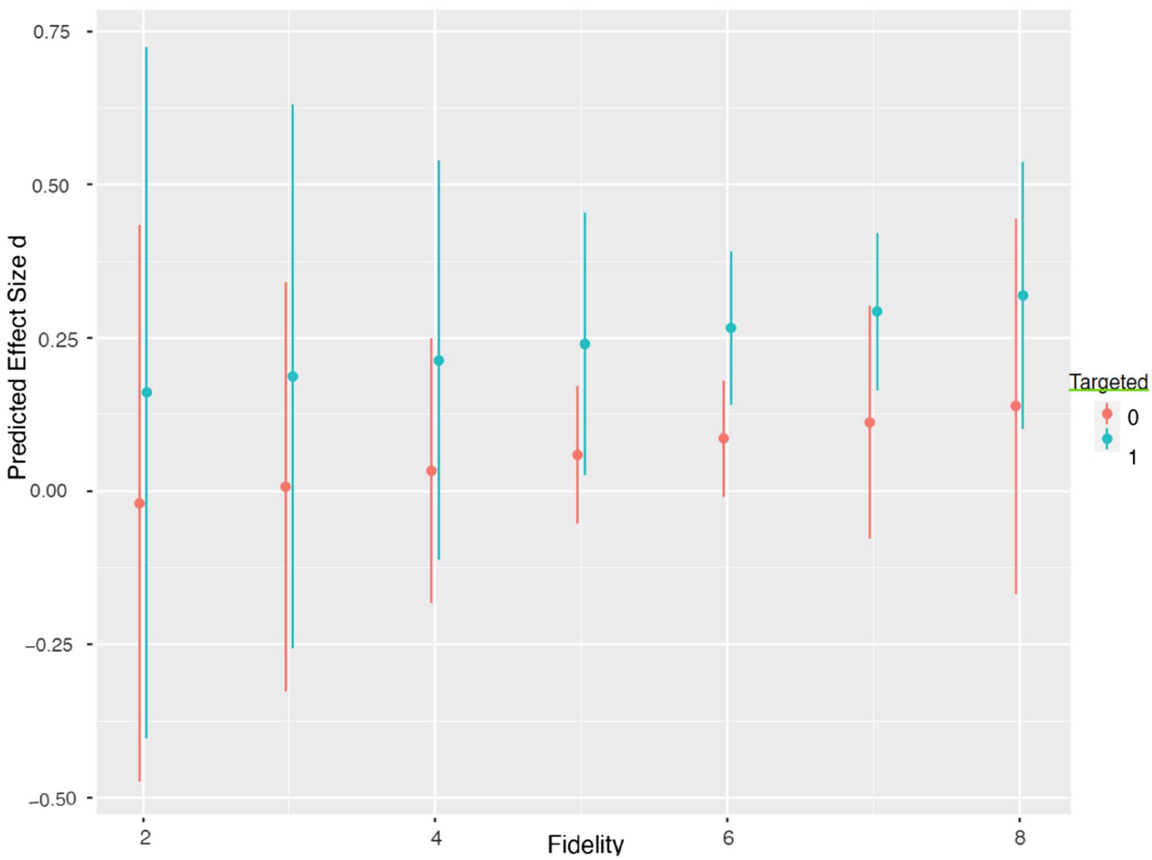
To determine whether the above results are robust to model specification, we estimated a CHE multilevel metaregression model in R using the metafor package (Viechtbauer, 2010), with effect sizes nested in studies as demonstrated by Pustejovsky and Tipton (2022), and with focal group and (median-centered) implementation fidelity as the predictors. Following Pustejovsky and Tipton (2022), this model was estimated based on the assumption of a common correlation (ρ) among effects sizes within studies. As an initial test, we assumed a ρ of 0.6, then demonstrated that conclusions remain unchanged with ρ equal to 0.2, 0.4, or 0.8.

Results from the initial estimation of the CHE model are provided in Table 8. Parameters show both the estimated effect size for interventions with a targeted focal group, and, separately, for interventions without a targeted focal group, controlling for implementation fidelity. In this model, as shown in the upper section of the panel, the intervention effect on targeted focal groups is estimated to be .13 *SD*s—comparable to what we found in our main model. But the standard errors and 95% confidence intervals extracted from that model are not (yet) robust to model misspecification. We then calculated standard errors and 95% confidence intervals for parameter estimates using robust variance estimation, computed with the ClubSandwich package (Pustejovsky, 2020) in R. These robust standard errors, along with the confidence intervals they yield, appear in the lower section of the table. The key result is that even with these (wider) 95% confidence intervals, the intervention effect on focal groups remained significantly greater than zero even when standard errors were computed using robust variance estimation, $b = 0.13$, $SE = 0.033$, 95% CI [0.06, 0.21]. Moreover, the difference in the intervention effect across the two levels of the focal group (0.117) remained statistically significant when standard errors were calculated using robust methods (p value for the Wald test = .011). Together, these dual results from robust variance estimation help safeguard conclusions against the possibility of model misspecification. Additionally, as noted earlier, this key result persisted across a full range of assumed values for ρ.

Exclusion Decisions

Finally, we ran sensitivity analyses related to exclusion decisions. Here, we focus on the focal effect for academic achievement and

Figure 6
Metaregression for Mental Health



Note. See the online article for the color version of this figure.

mental health (for a table of all excluded effects, see files on OSF). We first examined the effect when we included side effects and the train the teacher model effect. The focal effect for academic achievement was virtually unchanged, ranging from 0.14 to 0.15. Likewise, for mental health, effects ranged from 0.34 to 0.35. Second, the ITT versus LATE decision only pertained to academic achievement outcomes. When analyzing available ITT rather than LATE effects, the focal effect was 0.13, 95% CI [.06, .19]. Overall, these findings suggest that effects—at least the focal effect of interest in the present work—are robust to these exclusion decisions.

Table 8
Sensitivity Analysis: Model Specification

Correlated and hierarchical effects model			
Moderator	Estimate	*SE*	95% CI
Without robust standard errors			
Focal group: no	0.015	0.026	[−0.036, 0.066]
Focal group: yes	0.132	0.027	[0.079, 0.186]
Fidelity	0.003	0.013	[−0.023, 0.028]
With robust standard errors			
Focal group: no	0.015	0.020	[−0.027, 0.057]
Focal group: yes	0.132	0.033	[0.057, 0.207]
Fidelity	0.003	0.009	[−0.020, 0.025]

Note. *SE* = standard error; 95% CI = 95% confidence interval.

Discussion

Creating powerful interventions that can be implemented effectively requires a clear understanding of the relevant outcomes, desired end results, populations who benefit most, implementation strategies, contexts, mechanisms of change, and benchmarks for adequate ROI. We offered a framework for starting to understand each of these issues as they relate to growth MIE. More specifically, we integrated mindset theory, the organizational training approach (Figure 1), and a careful analysis of heterogeneity to address seven issues that drove our theoretical overview and analyses (Table 1; Figure 2). In the following sections, we review findings, conclusions, and future research directions organized by issue (for a summary, see Table 9).

Issue 1: Proximal Outcomes and Final Intervention End Results

The first issue we raised related to understanding and evaluating effectiveness across levels of outcomes and end results as well as an understanding of the degree of the heterogeneity of these effects.

Table 9
Discussion and Conclusions

Question	Findings	Conclusions	Future research directions
1. Can interventions improve important proximal outcomes, along with desired end results? Is there a great deal of heterogeneity in effects?	The larger effects are for the most proximal outcomes, with effects getting smaller when progressing from Level 1 (mindsets) to Level 4 (end results). Heterogeneity was high, especially for psychological and behavioral processes, where *k* was lower.	Average effect sizes are not representative based on the wide prediction intervals. Moderators must be considered.	• Routinely measure proximal intervention targets • At a minimum, always measure and report impact of intervention on growth mindsets • Test effects of interventions on meaningful outcomes beyond academic achievement
2. Are results stronger when interventions and analyses focus on subsamples expected to benefit most?	Effects of targeting interventions and analyses to specific subgroups were stronger than for universal analyses. This replicates past work.	Interventions that target interventions to subgroups with indicators of risk show larger effects.	• Identify and test relevant moderators in all intervention studies • Seek to assess the psychological construct rather than purely equating demographics or categorical descriptions with risk or vulnerability
3. Do differences in implementation practices impact effectiveness?	Our measure of implementation fidelity did account for some heterogeneity but not as much as we expected.	Higher fidelity improves effectiveness for mental health outcomes but had little impact on academic achievement.	• Identify growth mindset best practices and "gold standards" for fidelity • Include comprehensive reporting of intervention implementation and fidelity in future intervention studies
4. In what contexts are mindsets interventions most likely to yield results?	Not enough data to investigate the psychological and systemic characteristics of contexts that are associated with stronger intervention effects.	More data are needed to draw conclusions.	• Using theory, identify, test, and report potential contextual affordances as moderators
5. What is the effect when implemented with focal groups and at high fidelity. What degree of heterogeneity is accounted for?	Including moderators in the model helps to account for the observed heterogeneity.	Intervention effects are best understood with moderators in the model.	• Plan studies a priori to test relevant moderators • Include a combination of key boundary conditions that capture for whom, how, and where these interventions work best
6. Why do interventions work? What are the processes of change?	Relationships between levels (e.g., stronger relationships between adjacent levels) were not as strong as expected.	More theory and data are needed to draw conclusions.	• Specify and assess cognitive and behavioral mediating processes of change • Apply the Operating Conditions Framework (Rothman & Sheeran, 2021) to distinguish moderators that enhance *target engagement* and *target validity*
7. What is a meaningful effect?	To interpret findings, related and intuitive benchmarks must be considered.	When targeting subsamples, results are promising for important societal outcomes.	• Develop meaningful benchmarks for growth mindset interventions and when doing so, make sure to consider the return on investment (ROI)

In line with the organizational training framework, we found that effects were the largest for the most proximal intervention target—mindsets—and were smaller for the more distal end results outcomes (Table 2).

For end results, the average effects replicate past cumulative analyses. For example, the effect size for academic end results, when averaged across studies, is similar to past work examining the links between growth mindsets and goal achievement across contexts (e.g., Burnette et al., 2013; $r = .10$) as well as analyses examining average growth mindset intervention effects on academic achievement specifically (Sisk et al., 2018; $d = 0.08$). Yet, due to substantial heterogeneity in effects for all levels, as indicated by both large τ^2 values and wide prediction intervals, such summary effects are an opaque indicator that can lead to conclusions that insufficiently attend to heterogeneity. Thus, we defer interpretation regarding the practical significance of the end results until we consider relevant moderators—that is, we do not interpret average effects, especially in the presence of such large prediction intervals. Rather, we save the discussion of benchmarks and ROI for the focal effect, which incorporates a multivariate analysis of critical moderators.

We also sought to highlight the importance of examining effectiveness at each level. The inclusion of additional outcomes, such as persistence, is critical for understanding the potential value of these interventions, and to our knowledge, this is the first meta-analysis seeking to understand these additional relevant outcomes related to growth mindset interventions. Additionally, we highlighted the importance of extending existing syntheses to not only consider different levels of effects but also to investigate end results that go beyond academic performance. Although we found larger effects of mindset interventions on mental health and social functioning than on academic outcomes, the wide and varying prediction intervals again indicate substantial heterogeneity. Furthermore, the stronger effects could be due to differences in the extent to which each end result can be influenced by a growth mindset approach and/or due to the particular type of growth mindset that is manipulated (e.g., mindset of intelligence for academic outcomes vs. mindset of personality for social functioning). In addition, there are fewer studies examining mental health and social functioning, which inflates the potential influence of biases.

Our findings related to proximal and distal intervention outcomes suggest several future research directions. Overall, we encourage the ongoing investigation of growth mindset interventions that seek to replicate the application of mindset theory to additional outcomes and that address pressing social issues beyond academic contexts. For example, although changes in mindsets are often measured in intervention studies as a manipulation check, this is not a universal practice. At a minimum, all growth mindset intervention studies should be required to measure and report these effects. More broadly, however, researchers designing mindset interventions should carefully specify the proximal cognitive and goal-directed behavioral processes that they believe link changes in growth mindsets to changes in the final end result and should measure these processes in their research studies. We suggest starting with the end goal that is most meaningful to stakeholders and then working backward through the presumed change processes.

Issue 2: Focal Groups—Subsamples Expected to Benefit Most

Next, we asked the following question: Are intervention effects stronger when particular groups of participants are analyzed or targeted in comparison to when interventions or analyses are universally applied? The simple answer is yes. Our results mirror and extend the existing meta-analysis of growth mindset intervention effects conducted by Sisk et al. (2018). Although the average effect they obtained for academic outcomes was $d = 0.08$, heterogeneity was present, some of which was accounted for by subsamples. Namely, in their work, for students from low-SES backgrounds and those considered to be at risk, effects were larger than when aggregated across samples. In the present work, we find a similar pattern. The average effect for academic achievement is $d = 0.09$. Yet, our multilevel metaregression analyses replicated the importance of focal groups in accounting for heterogeneity. Like Sisk et al. (2018), this moderator was particularly strong for academic outcomes. Overall, the pattern of results—small average effects with wide prediction intervals as well as larger effects for targeted subsamples—mirrors past meta-analyses on growth mindset interventions. Thus, despite our use of different analytic approaches, inclusion decisions, and more, we largely replicated the main empirical conclusions from the existing synthesis (Sisk et al., 2018). We suggest that a key takeaway is that it is critical to understand *for whom* these interventions work. Interventions that are applied to the people who need them most, based on their potential to improve, are more likely to make an impact, whereas those applied more universally are less effective.

In future research, we suggest that simply labeling a particular social group "at risk" and assuming a mindset intervention will work is not the most powerful approach. For example, more theoretical clarity can be applied to understanding the nuance related to those expected to benefit most. That is, is it the participants with the most potential for developing stronger growth mindsets or for those with the most room for change in the end result? Or is it likely some combination of both? Furthermore, mindset interventions are sometimes targeted toward students with lower socioeconomic status (SES) or racial minority status. This method of targeting is merely a descriptive blunt instrument. The use of group characteristics, such as race, as indicators of vulnerability, in particular, reifies these categories as though they are causal variables—a practice that runs the risk of reinforcing negative group stereotypes while at the same time yielding less than it otherwise might from a scientific perspective (Helms et al., 2005). Subsequent studies need to move beyond simple "at-risk" categories to focus on the extent to which interventions target the underlying potential vulnerabilities. The major issue is that risk is not binary. What we really need is an assessment that can help us understand, for example, what being from a low-SES background means in the context of a student's own life, understanding that each lived experience is unique and cannot be captured categorically. Although difficult, if possible, researchers should try to delineate and measure the psychological constructs that they believe actually account for that group's relative vulnerability (Helms et al., 2005). Better measurement and reporting will help to continue to identify for *whom* mindset interventions work, and focusing on these meaningful sources of heterogeneity, rather than on average effects, will contribute to a more nuanced understanding of potential growth MIE.

Issue 3: Implementation Fidelity

Recognized indicators of implementation fidelity should account for heterogeneity across intervention effects. This was the case, at least to a small degree, for mental health but not for academic performance. The lack of moderation by fidelity is counter to what we would have anticipated, given the theoretical importance of delivering interventions "as intended." However, the available approach based on existing data likely missed important elements of implementation fidelity for growth mindset interventions, particularly because the field has not yet established an overall set of best practices for mindset interventions or an assessment of fidelity for such practices. In addition, and critically, our analyses were severely limited by sparsity in reporting. Few studies report all of the information necessary to code indicators of implementation fidelity, resulting in a smaller than an ideal number of included studies.

Future research in this area should be aimed at establishing a "gold standard" of implementation for growth mindset interventions, along with best practices for reporting the information necessary to assess fidelity. These goals are particularly important given the relevance of fidelity for both direct and conceptual replications as well as for extensions to new domains and outcomes. Researchers must understand *how* to deliver a growth mindset intervention most effectively, especially as these interventions are extended to novel outcomes. This research should also provide a valid and reliable assessment of fidelity that could then be used as a moderator.

Issue 4: Contextual Affordances

Although we were unable to include affordances in our statistical analyses, it is critical to discover the contexts in which growth mindsets—and psychological interventions in general—can have the most favorable effects. Future research should draw upon the theory of how mindsets are assumed to have their effects to identify additional types of "soil" in which the "seed" of the growth mindset is most likely to yield fruit. For example, in the academic domain, recognizing the importance of teachers' own mindsets in the efficacy of mindset interventions (Yeager et al., 2022), a recent large-scale intervention was effective at increasing struggling students' grades by involving teachers in the delivery of the growth mindset messages (Porter et al., 2020). Or, in the domain of mental health, if growth mindsets of emotions are assumed to improve mental health via increased use of approach-focused regulation strategies (see Burnette, Knouse, et al., 2020), then studies should test whether growth mindset interventions yield stronger effects when delivered in the context of skills-based mental health treatments, such as cognitive–behavioral therapy, compared to other types of mental health interventions. Indeed, any number of therapist characteristics and practices could be conceptualized as consistent with a growth mindset orientation and may predict enhanced effects. Overall, the field still needs a better understanding of the context in which these interventions work best—what culture or environment provides a rich soil ready for growth and development? Is this soil dependent on the end goal?

Issue 5: Focal Effects

Combining moderators gives a sense of the potential of well-targeted and well-implemented growth mindset interventions. These analyses helped to parse some of the heterogeneity, especially the focal group moderator. The focal effect analyses are better estimates than average effects, and results revealed, $d = 0.14$ for academic performance and $d = 0.32$ for mental health (Table 5). In the case of academic performance, we conducted a number of sensitivity analyses and found limited evidence that our conclusions regarding growth mindset intervention effects upon targeted focal groups are sensitive to publication bias, study-level variation in research quality, or model specification.

Yet, even when accounting for implementation fidelity and the use of focal groups, residual heterogeneity was still substantial. Indeed, residual heterogeneity in the estimated distribution of effects for academic outcomes was large enough (95% prediction intervals ranged from −.08 to .35) that null and negative results are sometimes to be expected, even when interventions are targeted to focal groups and implemented with fidelity. This residual heterogeneity calls for additional study of moderators that will further elucidate for whom and in what contexts mindsets interventions have their strongest effects. Importantly, the ability of meta-analyses to answer these questions is limited by the moderators included in past studies and reports of relevant information.

Nonetheless, future mindset work can be carefully designed to test ideas about the conditions under which interventions have their effects. We invite mindset researchers to consider the Operating Conditions Framework (OCF), which was recently developed by *health psychology* researchers Rothman and Sheeran (2021) to foster a more nuanced understanding of precisely where in the causal process moderators have their effects. Specifically, the model requires researchers to specify whether moderators affect *target engagement*, the extent to which the intervention changes the presumed mediator, versus *target validity*, or the extent to which changes in the mediator affect the intended outcome. For example, a mindset researcher might ask whether the effect of a mindset intervention on students' growth mindsets depends on the mindset of their teacher (target engagement moderator). The researcher might then ask whether the effect of the students' changed mindset on academic outcomes depends on the availability of academic support services (target validity moderator). Using the OCF, researchers can pose and test more nuanced questions about the boundary conditions relevant to all the levels outlined by our MIE model.

Issue 6: Mediators

Focusing on multiple levels using the organizational framework helped us to examine questions related to mechanisms of change. Yeager and Walton (2011) contend that wise interventions, such as growth mindset interventions, boost academic achievement "by setting into motion recursive social, psychological, and intellectual processes in school" (p. 286). Effects may be stronger when predicting the psychological processes, which then in turn may promote behaviors that contribute to better desired final results. We find a pattern that theoretically supports this assertion (Table 6). Namely, when moving from proximal to more distal outcomes, effects decrease in magnitude, with the largest effects at Level 1. Yet, we see less support for hypothesized links between levels. This may be due to the nature of the available data to test only one primary mediator at this level (i.e., expectations) and one that is not necessarily the most central to mindset theorizing. Additionally, the wide prediction intervals for all our findings indicate boundary conditions across the outcomes. Because fewer studies reported these intermediary levels, our ability to detect effects, examine all relevant mediators, and test important moderators at each level and link is limited.

In terms of future inquiry, the integration of mindset theory with the organizational training literature revealed substantial gaps in existing empirical tests of theoretically driven mechanisms of change. Based on the current state of the literature, testing an overall process model was not feasible in the present analysis. Too few studies included all levels or tested the specific mediated paths outlined in the MIE model. Additionally, given the small number of studies pertaining to psychological and behavioral processes, and the inclusion of interventions that reported on a potential process variable but not an end result, the field rarely considers each of these driving mechanisms. Furthermore, most of the studies that included process-relevant outcomes were in the context of academic performance, which in turn suggests that in the domains of mental health and social functioning there is much more work to be done in devoting attention and empirical resources to the intermediate influences. The organizational training framework reminds researchers to carefully consider each level starting with the end result and working back to consider the most optimal outcome at each level. For example, when seeking to improve social functioning (e.g., reduce aggression, foster reconciliation), the value individuals place on the relationship and their evaluations of the potential for future exploitation is likely a more influential psychological process (Billingsley et al., 2019; Burnette et al., 2012; Forster et al., 2021) than expectations about their own capacity to engage with the transgressor.

The growth mindset meaning system framework also includes a number of additional theoretically rich psychological processes. For example, in a review of two eras of work on growth mindsets, goals and attributions are highlighted as critical processes (Dweck & Yeager, 2019). Individuals with growth, relative to fixed, mindsets set goals focused on process rather than on performance and view failure as an opportunity for development rather than as proof of lack of aptitude. Additional work highlights the potential importance of being in learning mode and the role that growth mindsets may play in fostering such an orientation (Heslin et al., 2020). Learning mode facilitates an individual's willingness and ability to gain insight from setbacks and ultimately persist and succeed. Overall, more work is needed that investigates *why* growth mindset interventions exert effects on end results.

Issue 7: ROI Benchmarks

As emphasized in the organizational training literature and our introduction, the key question here is as follows: Do focal effects offer a reasonable ROI for end results? Evaluating ROI requires identifying a target effect size at or above which the intervention is deemed worth the overall cost—that is, the time, effort, and money invested. The value of the effect of a given intervention increases if that intervention can be delivered at relatively low cost relative to other viable options. For educational interventions designed to enhance academic performance, this meta-analysis offers some grounds for optimism—particularly given that academic achievement is a notoriously "hard-to-move" outcome and considering that growth mindset interventions are particularly low cost. For example, a year's worth of instruction in ninth grade or a year with a good versus average teacher produced average effect sizes around, $d = 0.20$ (Yeager et al., 2019). Although this particular comparison effect is slightly larger than the $d = 0.14$ found in the present work for interventions directed toward a targeted focal group, growth mindset interventions typically involve a substantially lower investment. Furthermore, growth mindset interventions designed with the goal of improving mental health outcomes seem promising. The focal intervention effect upon mental health ($d = 0.32$) compares favorably with the effect of antidepressant medication versus placebo on mild-to-moderate depression ($d = 0.20$; Fournier et al., 2010).

Finally, we obtained the largest effect on end results when the goal was to improve social functioning, such as responding to peer aggression, but here we note that this is simply the average—the literature is too nascent for moderator analyses. The effect size of $d = 0.36$ suggests promising initial effects for these interventions and the potential for considerable ROI, especially once more work is done that outlines the who, how, and why of these types of interventions.

Summary

We argued that questions about growth MIE should incorporate wide-ranging outcomes, consider critical boundary conditions, understand the mechanisms, and include a discussion of meaningful benchmarks for ROI. We find that growth mindset interventions can and do foster stronger growth mindsets and to some degree also impact psychological and behavioral processes related to end results. However, more work is needed here that focuses on moderators of these effects. Additionally, using multilevel metaregression, we shed light on focal groups as a key contributor of heterogeneity in effects for both academic achievement and mental health. Although fidelity failed to account for much heterogeneity, we encourage the field to continue to consider how best to assess this variable and to report mindset intervention implementation practices clearly. We also provide an initial, although tentative, first look at overall effects for social functioning-related outcomes. The framework and model provided offer a path forward for future scientific inquiry that continues to examine growth MIE—a path that must be sensitive to heterogeneity. The limitations listed below also provide a springboard for such future work.

Limitations

The findings are not without limitations that need to be considered before drawing decisive conclusions. Many of these limitations apply more broadly to meta-analytic reports, whereas others pave the way for future inquiry as they relate to existing gaps in the current state of the literature.

First, any meta-analysis is limited by the quality of the data included in analyses. In order to be comprehensive, we did not include or exclude studies on the basis of predetermined indicators of study quality. As for quality, many of the interventions included in the analysis were published prior to the credibility revolution (Vazire, 2018), and thus results should be interpreted with this in mind. We coded for research quality and included this variable in models that uncovered no significant evidence of impact, but definitions of research quality are often subjective. Rather than relying solely on findings from cumulative analyses, any effect reported herein should be considered in concert with results from large-scale, preregistered interventions using agreed-upon best research practices (van Elk et al., 2015).

Second, in addition to limitations related to methodological rigor in primary studies included, generalizability is also restrained by the nature of primary study samples and the search procedures. The samples were nearly 100% WEIRD. Yet, behavioral scientists consistently find substantial variability across populations (Henrich et al., 2010). As noted by Johnson (2021) in his recent editorial discussing how syntheses can be a corrective social force, *Science risks advancing only the status quo—and potentially perpetuating falsehoods—unless the dialogue incorporates those with divergent backgrounds* (p. 7). A similar issue arises when considering search terms. Although studies written in languages other than English were not excluded a priori from consideration, the forward literature search was conducted using English search terms only. Likewise, the legacy search began with only English-language publications, and outreach to relevant professional organizations for unpublished studies was similarly done in English. The analytic sample therefore may reflect mono-language bias (Johnson, 2021). Overall, future work should seek to address generalizability by investigating the robustness of conclusions across previously ignored groups.

Third, the literature applying growth mindset research to interventions is too nascent for a statistical analysis of all of the relevant mediating mechanisms. For example, we limited our Level 2 analyses to expectations, as this literature was prolific enough to examine cumulatively. Although there is some evidence of a link between growth mindsets and expectations in past correlational studies (e.g., van Aalderen-Smeets & Walma van der Molen, 2018) and meta-analyses (e.g., Burnette et al., 2013, where the link between growth mindsets and expectations for success was $r = .16$), empirical evidence from the present meta-analysis suggests that interventions are not reliably shifting this psychological process and this process is not a relevant predictor of the end result. This may be, in part, because growth mindsets are also postulated to enhance challenge-seeking and a willingness to try new things, including engaging in tasks and activities where one lacks the ability currently.

There are many postulated social-cognitive mechanisms that may better capture how these interventions work. For example, in a nationally representative sample, mindsets predicted academic performance and challenge-seeking via attributions, goals, and effort beliefs (Dweck & Yeager, 2019). In other work, growth mindset interventions fostered greater value placed on learning and interest in the field (Burnette, Hoyt, et al., 2020). Overall, future research needs to consider and test the cognitive processes driving intervention effects.

Fourth, for Level 3, goal-directed behavior, we took the opposite approach from what we took for Level 2, broadening the scope and definition. This choice, similar to our reasoning for focusing on expectations, was based on the available data within the studies included in analyses. Namely, work to date rarely includes behavior, and the smaller k available for this level meant we needed to combine types of goal-directed behaviors. However, this choice led to greater variability in the nature of the behaviors. For example, it is possible that growth mindset interventions improve help-seeking behaviors but do not as strongly impact persistence. Alternatively, these interventions may be best situated for fostering challenge-seeking behaviors such as taking more difficult math courses (e.g., Yeager et al., 2019) or taking a full course load in college (e.g., Yeager, Walton, et al., 2016). Put simply, a limitation is that Level 3 is restricted in terms of delineating precise behaviors that may be most impacted by growth mindset interventions. Future work, drawing on the organizational training approach offered here, needs to delineate the most powerful behavioral processes driving effects for each desired end goal.

Fifth, we obtained rather large prediction intervals and indicators of heterogeneity at all levels. These wide prediction intervals, even when accounting for meaningful heterogeneity, suggest additional boundary conditions. Just as we could not model all potential mediators, we could not model or test all relevant moderators. There are abundant sources of heterogeneity ranging from methodological (e.g., measurement) to demographic differences in samples (e.g., age). In considering the level of Type I error that may exist when the number of moderator variables becomes very large (e.g., Hedges & Olkin, 1985), we limited our focus to the theoretically derived moderators outlined in our model. Yet, there are important additional boundary conditions that need to be tested in future work. For example, questions remain regarding the impact of time of assessment—were outcomes measured immediately postintervention or at follow-up? Perhaps, a more practical question is this: Is the impact of the intervention durable? This question is best answered with a continuous moderator—something not practical in the present work due to lack of reporting on time between assessments (e.g., authors often noted the end of semester as the testing point), the use of multiple assessments (e.g., 3, 6, and 9 months) and limited immediate posttest evaluations of impact for most end results (e.g., depression is only assessed at follow-up). Furthermore, although it was beyond the scope to model and test all moderators or boundary conditions for all levels of effects (e.g., mindsets, motivation, behavior), as described earlier, the OCF (Rothman & Sheeran, 2021) can be used in future inquiry to develop a more nuanced understanding of precisely where in the causal process moderators are impactful.

Additionally, and related, it is important to keep in mind that the results of moderator analyses in meta-analysis are observational, not experimental. The characteristics could be, and likely are, confounded with other features of the interventions that may account for observed effects. Indeed, moderator analyses are rough estimates across samples, rather than estimates garnered from experimental studies designed to specifically parse effects and are often based on smaller samples of studies. An important next step will be preregistered, large-scale, collaborative lab efforts that test both theoretically driven moderators and methodological artifacts of intervention effectiveness.

Sixth, publication bias is an important issue to consider and one that could limit the conclusions that can be drawn from any meta-analytic approach. As one cannot correct for publication bias the way one corrects for measurement error, the primary question is to what extent might an estimate be adjusted based on the sensitivity analysis. This helps to inform the degree to which an overall conclusion may change if bias were present. The selection models we used have the advantage of determining the extent to which effects—including moderated effects under conditions of heterogeneity—remain robust to differing degrees of publication bias, which is uncommon in previous techniques used to detect publication bias (Vevea et al., 2019). Yet, even these advanced approaches do not fully account for the nested nature of the data. Although we took steps to assess how sensitive our conclusions are to publication bias and introduced procedures to account in part for the nested structure of our data, our sensitivity analyses could not fully account for the data's complex, multilevel structure. The field would greatly benefit from extending existing sensitivity analyses into a multilevel context. Furthermore, readers should interpret findings and conclusions through a critical lens that considers the issues related to obtaining unpublished work as well as other biases that are part of the scientific process.

Seventh, and related, our sensitivity analyses focused primarily on the focal effect for academic achievement, as this effect had adequate studies to run multivariate moderator tests. Our approaches revealed that the focal effect was mostly robust to publication bias, research quality, control group type, and the statistical model for handling interdependency. Nonetheless, despite the absence of statistical significance regarding moderators, these tests alone should not be used to draw conclusions about moderation, especially given limited power based on the number of studies in the sample. For example, future research might be especially helpful in clarifying the possible moderating role of research quality. In the present analyses, research quality did not significantly moderate intervention effects on academic performance, but the lack of a significant result does not entail that there is no association of research quality with effect magnitude at the population level. If there is such an association, our results suggest that this association is likely to be negative (at least in the context of academic achievement), such that estimated effects among the highest quality studies would be lower than effects in studies of low or median research quality. Additionally, our exclusion decision sensitivity analyses included both academic achievement and mental health as outcomes. Despite showing that the focal effect for these outcomes is rather robust to exclusion decisions, this sensitivity analysis, and the moderator analyses more generally, are limited by the available data reported in the original articles. Namely, for the heterogeneity revolution to work, authors need to avoid selective reporting and make sure to report all simple effects, regardless of statistical significance. Overall, in addition to more complete reporting, more work is needed that explores sources of heterogeneity, especially research quality. Conclusions need to be considered in concert with other evidence.

Eighth, another potential limitation is the possible presence of expectancy effects—findings that are not a result of "true" effects but instead reflect that the researcher or intervention implementer shaped the outcome via their own expectations (e.g., Rosenthal & Rubin, 1978). In the context of growth mindset interventions, researcher expectancy effects are most problematic for interventions in which teachers both deliver intervention content and assess primary outcomes (e.g., assign course or project grades). In the present work, such concerns are largely mitigated. Namely, in the case of academic achievement, only roughly 20% of interventions were administered by the teacher, and in the majority (63%) of those cases, expectancy bias was avoided entirely by the use of standardized test scores as the primary outcomes. Nonetheless, expectancy bias and demand characteristics as well as other biases should be considered when interpreting effects.

In outlining limitations, we encourage readers to consider the results and conclusions offered here with a critical lens toward the complex issues related to systematic reviews, both in terms of the available data from primary studies as well as the copious decision points and inherent limitations of meta-analytic approaches. The multilevel approach allowed us to retain all data points (rather than picking single effects) and captured crucial within-study variation. The approach also allowed us to simultaneously model multiple moderators. Nonetheless, we could not model all sources of heterogeneity, and causal conclusions cannot be drawn from any moderation test in meta-analysis. Additionally, despite employing multilevel metaregression analyses to account for the nested nature of data (Gooty et al., 2021) and trying to follow best practices in our approach (Tipton et al., 2019a, 2019b), there are multiple perspectives on how best to conduct systematic reviews and inevitably different decisions and conclusions can and will be drawn by different researchers.

Despite the many limitations because we used metaregression driven by empirically tested and theoretically driven choices of moderators, our meta-analysis is potentially more robust than meta-analyses which (a) are not based on experimental data (DeSimone et al., 2021); (b) do not consider the nested nature of data (or whether or not primary study authors considered the nested nature of the data; Gooty et al., 2021); and (c) do not model multiple moderators simultaneously (Gonzalez-Mulé & Aguinis, 2018). Nonetheless, before drawing definitive conclusions, the work presented here needs to be supplemented with large-scale, preregistered, replication studies that use experimental methods to test moderators. We hope that our theoretical model that outlines the potential sources of heterogeneity provides a strong foundation for this necessary follow-up work.

Conclusions

Current trends indicate sustained interest in growth mindset interventions, with applications to a number of pressing social issues that go beyond academic performance. An overarching framework for understanding growth MIE is especially timely and relevant in light of not only the multidisciplinary expansion but also the interest of the lay public as well as scientists and policymakers. We offered a theoretical model that focused on for whom, how, and why might growth mindset intervention work. Grounded in mindset theory and drawing on the training effectiveness literature, we investigated effect sizes across levels, end results, populations, and implementation strategies. In addition to offering theoretical coherence to a mounting literature, we highlighted that average effects hide important differences. Indeed, in line with recent work (IntHout et al., 2016) calling for the inclusion of prediction intervals, it is clear in the present work that null and even negative (in the case of academic achievement) effects are to be expected in growth mindset interventions. The broad prediction intervals point to the importance of moving beyond average effects to understand heterogeneity. Syntheses that code subgroups within studies, model them in a theoretically driven multilevel approach, and focus conclusions on moderated findings, are recommended to help understand such variability (e.g., McShane & Böckenholt, 2018; Tipton et al., 2019b).

We highlighted how effects are stronger to the degree that the analyses and/or interventions are targeted to focal groups, and we outlined the potential role of implementation fidelity and context–consciousness. We also discussed several limitations that can spark discussion and future inquiry. Overall, we hope that this multilevel meta-analytic approach with an eye toward heterogeneity can offer researchers across disciplines a roadmap to follow when conducting research syntheses. We also hope that the theoretical foundation paves the way for further replication efforts and encourages additional applications of growth mindset interventions to end results that go beyond academic performance.

References

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

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

Altman, D. G., & Bland, J. M. (2011). How to obtain the confidence interval from a *p* value. *BMJ*, *343*, Article d2090. https://doi.org/10.1136/bmj.d2090

*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

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. The Psychological Corporation.

*Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F., & Yeager, D. S. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior & Organization*, *146*, 1–15. https://doi.org/10.1016/j.jebo.2017.11.032

Billingsley, J., Burnette, J. L., & McCullough, M. E. (2019). An evolutionary perspective on forgiveness. In E. L. Worthington & N. G. Wade (Eds.), *Handbook of forgiveness* (pp. 52–62). Taylor & Francis. https://doi.org/10.4324/9781351123341-6

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

*Blau, I., & Benolol, N. (2016). Can designing self-representations through creative computing promote an incremental view of intelligence and enhance creativity among at-risk youth? *Interdisciplinary Journal of E-Skills and Lifelong Learning*, *12*, 267–278. https://doi.org/10.28945/3577

Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 207–243). Sage Publications. https://doi.org/10.7758/9781610448864.14

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (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* [Master's thesis]. Oregon State University. https://ir.library.oregonstate.edu/downloads/k0698c85d

*Brez, C., Hampton, E. M., Behrendt, L., Brown, L., & Powers, J. (2020). Failure to replicate: Testing a growth mindset intervention for college student success. *Basic and Applied Social Psychology*, *42*(6), 460–468. https://doi.org/10.1080/01973533.2020.1806845

*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

*Brougham, L., & Kashubeck-West, S. (2017). 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

Burkhardt, J., & Lenhard, W. (2022). A meta-analysis on the longitudinal, age-dependent effects of violent video games on aggression. *Media Psychology*, *25*(3), 499–512. https://doi.org/10.1080/15213269.2021.1980729

Burnette, J., Billingsley, J., & Banks, G. C. (2022, July 8). *Growth mindset intervention meta-analysis*. https://osf.io/2fgey

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., Forsyth, R. B., Desmarais, S. L., & Hoyt, C. L. (2019). Mindsets of addiction: Implications for treatment intentions. *Journal of Social and Clinical Psychology*, *38*(5), 367–394. https://doi.org/10.1521/jscp.2019.38.5.367

*Burnette, J. L., Hoyt, C. L., Russell, V. M., Lawson, B., Dweck, C. S., & Finkel, E. (2020). 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., Knouse, L. E., Vavra, D. T., O'Boyle, E., & Brooks, M. A. (2020). Growth mindsets and psychological distress: A meta-analysis. *Clinical Psychology Review*, *77*, Article 101816. https://doi.org/10.1016/j.cpr.2020.101816

Burnette, J. L., McCullough, M. E., Van Tongeren, D. R., & Davis, D. E. (2012). Forgiveness results from integrating information about relationship value and exploitation risk. *Personality and Social Psychology Bulletin*, *38*(3), 345–356. https://doi.org/10.1177/0146167211424582

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., Pollack, J. M., Forsyth, R. B., Hoyt, C. L., Babij, A. D., Thomas, F. N., & Coy, A. E. (2020). A growth mindset intervention: Enhancing students' entrepreneurial self-efficacy and career development. *Entrepreneurship Theory and Practice*, *44*(5), 878–908. https://doi.org/10.1177/1042258719864293

*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

*Calvete, E., Fernández-Gonzalez, L., Orue, I., Echezarraga, A., Royuela-Colomer, E., Cortazar, N., Muga, J., Longa, M., & Yeager, D. S. (2019). The effect of an intervention teaching adolescents that people can change on depressive symptoms, cognitive schemas, and hypothalamic-pituitary-adrenal axis hormones. *Journal of Abnormal Child Psychology*, *47*(9), 1533–1546. https://doi.org/10.1007/s10802-019-00538-1

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

*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, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292. https://doi.org/10.3102/0013189X16656615

Coburn, K. M., & Vevea, J. L. (2019). *Estimating weight-function models for publication bias* [R package weightr Version 2.0.2]. Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=weightr

Cook, P. J., Dodge, K. A., Farkas, G., Fryer, R. G., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., & Steinberg, L. (2015). *Not too late: Improving academic outcomes for disadvantaged youth*. [Institute for Policy Research Northwestern University Working Paper WP-15-01]. https://www.ipr.northwestern.edu/our-work/working-papers/2015/ipr-wp-15-01.html

Costa, A., & Faria, L. (2018). Implicit theories of intelligence and academic achievement: A meta-analytic review. *Frontiers in Psychology*, *9*, Article 829. https://doi.org/10.3389/fpsyg.2018.00829

Cury, F., Da Fonseca, D., Zahn, I., & Elliot, A. (2008). Implicit theories and IQ test performance: A sequential mediational analysis. *Journal of Experimental Social Psychology*, *44*(3), 783–791. https://doi.org/10.1016/j.jesp.2007.07.003

Da Fonseca, D., Cury, F., Fakra, E., Rufo, M., Poinso, F., Bounoua, L., & Huguet, P. (2008). Implicit theories of intelligence and IQ test performance in adolescents with Generalized Anxiety Disorder. *Behaviour Research and Therapy*, *46*(4), 529–536. https://doi.org/10.1016/j.brat.2008.01.006

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45. https://doi.org/10.1016/S0272-7358(97)00043-3

Davis, D. E., Choe, E., Meyers, J., Wade, N., Varjas, K., Gifford, A., Quinn, A., Hook, J. N., Van Tongeren, D. R., Griffin, B. J., & Worthington, E. L. (2016). Thankful for the little things: A meta-analysis of gratitude interventions. *Journal of Counseling Psychology*, *63*(1), 20–31. https://doi.org/10.1037/cou0000107

*Derr, S., & Morrow, M. T. (2020). Effects of a growth mindset of personality on emerging adults' defender self-efficacy, moral disengagement, and perceived peer defending. *Journal of Interpersonal Violence*, *35*(3–4), 542–570. https://doi.org/10.1177/0886260517713716

DeSimone, J. A., Brannick, M. T., O'Boyle, E. H., & Ryu, J. W. (2021). Recommendations for reviewing meta-analyses in organizational research. *Organizational Research Methods*, *24*(4), 694–717. https://doi.org/10.1177/1094428120967089

Dickens, L. R. (2017). Using gratitude to promote positive change: A series of meta-analyses investigating the effectiveness of gratitude interventions. *Basic and Applied Social Psychology*, *39*(4), 193–208. https://doi.org/10.1080/01973533.2017.1323638

Dobson, D., & Cook, T. J. (1980). Avoiding type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning*, *3*(4), 269–276. https://doi.org/10.1016/0149-7189(80)90042-7

Dodge, K. A. (2020). Annual Research Review: Universal and targeted strategies for assigning interventions to achieve population impact. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *61*(3), 255–267. https://doi.org/10.1111/jcpp.13141

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Phillips Smith, E., & Prinz, R. J. (2001). Promoting intervention fidelity. Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine*, *20*(1 Suppl. 1), 38–47. https://doi.org/10.1016/S0749-3797(00)00272-5

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237–256. https://doi.org/10.1093/her/18.2.237

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

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

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

Dworkin, E. R., Menon, S. V., Bystrynski, J., & Allen, N. E. (2017). Sexual assault victimization and psychopathology: A review and meta-analysis. *Clinical Psychology Review*, *56*, 65–81. https://doi.org/10.1016/j.cpr.2017.06.002

*Fabert, N. S. (2014). *Growth mindset training to increase women's self-efficacy in science and engineering: A randomized-controlled trial (UMI No. 3632583)* [Doctoral dissertation]. ProQuest Dissertations and Theses database.

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

Forster, D. E., Billingsley, J., Burnette, J. L., Lieberman, D., Ohtsubo, Y., & McCullough, M. E. (2021). Experimental evidence that apologies promote forgiveness by communicating relationship value. *Scientific Reports*, *11*(1), Article 13107. https://doi.org/10.1038/s41598-021-92373-y

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., & Fawcett, J. (2010). Antidepressant drug effects and depression severity: A patient-level meta-analysis. *JAMA*, *303*(1), 47–53. https://doi.org/10.1001/jama.2009.1943

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

*Ganimian, A. J. (2020). Growth-mindset interventions at scale: Experimental evidence from Argentina. *Educational Evaluation and Policy Analysis*, *42*(3), 417–438. https://doi.org/10.3102/0162373720938041

*Gillen, M. J. (2014). *Altering shyness mindset: Enhancing treatment for social performance anxiety (UMI No. 3681933)* [Doctoral dissertation]. ProQuest Dissertations and Theses database.

*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

Goldenberg, A., Endevelt, K., Ran, S., Dweck, C. S., Gross, J. J., & Halperin, E. (2017). Making intergroup contact more fruitful: Enhancing cooperation between Palestinian and Jewish-israeli adolescents by fostering beliefs about group malleability. *Social Psychological & Personality Science*, *8*(1), 3–10. https://doi.org/10.1177/1948550616672851

Gonzalez-Mulé, E., & Aguinis, H. (2018). Advancing theory by assessing boundary conditions with metaregression: A critical review and best-practice recommendations. *Journal of Management*, *44*(6), 2246–2273. https://doi.org/10.1177/0149206317710723

*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

Gooty, J., Banks, G. C., Loignon, A. C., Tonidandel, S., & Williams, C. E. (2021). Meta-analyses as a multi-level model. *Organizational Research Methods*, *24*(2), 389–411. https://doi.org/10.1177/1094428119857471

Hecht, C., Yeager, D. S., Dweck, C., & Murphy, M. (2021). Beliefs, affordances, and adolescent development: Lessons from a decade of growth mindset interventions. In J. J. Lockman (Ed.), *Advances in child development and behavior* (Vol. 61, pp. 169–197). Elsevier Academic Press. https://doi.org/10.1016/bs.acdb.2021.04.004

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press. https://idostatistics.com/hedges-olkin-1985-statistical-methods-for-meta-analysis/

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Helmreich, R., & Stapp, J. (1974). Short forms of the Texas Social Behavior Inventory (TSBI), an objective measure of self-esteem. *Bulletin of the Psychonomic Society*, *4*(5), 473–475. https://doi.org/10.3758/BF03334260

Helms, J. E., Jernigan, M., & Mascher, J. (2005). The meaning of race in psychology and how to change it: A methodological perspective. *American Psychologist*, *60*(1), 27–36. https://doi.org/10.1037/0003-066X.60.1.27

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Heslin, P. A., Keating, L. A., & Ashford, S. J. (2020). How being in learning mode may enable a sustainable career across the lifespan. *Journal of Vocational Behavior*, *117*, Article 103324. https://doi.org/10.1016/j.jvb.2019.103324

*Heslin, P. A., Latham, G. P., & VandeWalle, D. (2005). The effect of implicit person theory on performance appraisals. *Journal of Applied Psychology*, *90*(5), 842–856. https://doi.org/10.1037/0021-9010.90.5.842

*Heslin, P. A., Vandewalle, D., & Latham, G. P. (2006). Keen to help? Managers' implicit person theories and their subsequent employee coaching. *Personnel Psychology*, *59*(4), 871–902. https://doi.org/10.1111/j.1744-6570.2006.00057.x

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

*Hoang, T. V. (2018). *Growth mindset and task value interventions in college algebra* [Doctoral dissertation]. Texas State University. https://digital.library.txstate.edu/handle/10877/7423

*Im, T. (2012). *Effects of emotional support and cognitive motivational messages on math anxiety, self-efficacy, and math problem solving* [Doctoral dissertation]. Florida State University. https://diginole.lib.fsu.edu/islandora/object/fsu%3A182927/

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, *6*(7), Article e010247. https://doi.org/10.1136/bmjopen-2015-010247

Ivie, E. J., Pettitt, A., Moses, L. J., & Allen, N. B. (2020). A meta-analysis of the association between adolescent social media use and depressive symptoms. *Journal of Affective Disorders*, *275*, 165–174. https://doi.org/10.1016/j.jad.2020.06.014

Johnson, B. T. (2021). Toward a more transparent, rigorous, and generative psychology. *Psychological Bulletin*, *147*(1), 1–15. https://doi.org/10.1037/bul0000317

Johnson, B. T., & Huedo-Medina, T. B. (2011). Depicting estimates using the intercept in meta-regression models: The moving constant technique. *Research Synthesis Methods*, *2*(3), 204–220. https://doi.org/10.1002/jrsm.49

Kepes, S., Banks, G., & Oh, I.-S. (2014). Avoiding bias in publication bias research: The value of "null" findings. *Journal of Business and Psychology*, *29*(2), 183–203. https://doi.org/10.1007/s10869-012-9279-0

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *The Journal of the American Society of Training Directors*, *13*(11), 3–9.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publishers.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, *2*(1), 61–76. https://doi.org/10.1002/jrsm.35

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Lenhard, W., & Lenhard, A. (2016). *Computation of effect sizes* [Data set]. Psychometrica. https://doi.org/10.13140/RG.2.2.17823.92329

Limpo, T., & Alves, R. A. (2014). Implicit theories of writing and their impact on students' response to a SRSD intervention. *The British Journal of Educational Psychology*, *84*(Pt. 4), 571–590. https://doi.org/10.1111/bjep.12042

*McCabe, J. A., Kane-Gerard, S., & Friedman-Wheeler, D. G. (2020). Examining the utility of growth-mindset interventions in undergraduates: A longitudinal study of retention and academic success in a first-year cohort. *Translational Issues in Psychological Science*, *6*(2), 132–146. https://doi.org/10.1037/tps0000228

McShane, B. B., & Böckenholt, U. (2018). Multilevel multivariate meta-analysis with application to choice overload. *Psychometrika*, *83*(1), 255–271. https://doi.org/10.1007/s11336-017-9571-z

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Mesurado, B., Guerra, P., Richaud, M. C., & Rodriguez, L. M. (2019). Effectiveness of prosocial behavior interventions: A meta-analysis. In P. Á. Gargiulo & H. L. Mesones Arroyo (Eds.), *Psychiatry and neuroscience update: From translational research to a humanistic approach—Volume III* (pp. 259–271). Springer International Publishing. https://doi.org/10.1007/978-3-319-95360-1_21

Miller, D. T., Dannals, J. E., & Zlatev, J. J. (2017). Behavioral processes in long-lag intervention studies. *Perspectives on Psychological Science*, *12*(3), 454–467. https://doi.org/10.1177/1745691616681645

*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

*Miu, A. S. (2016). *The effects of mindsets on depression (UMI No. 10290120)* [Doctoral dissertation]. ProQuest Dissertations and Theses database.

*Miu, A. S., & Yeager, D. S. (2015). Preventing symptoms of depression by teaching adolescents that people can change: Effects of a brief incremental theory of personality intervention at 9-month follow-up. *Clinical Psychological Science*, *3*(5), 726–743. https://doi.org/10.1177/2167702614548317

Molden, D. C., & Dweck, C. S. (2006). Finding "meaning" in psychology: A lay theories approach to self-regulation, social perception, and social development. *American Psychologist*, *61*(3), 192–203. https://doi.org/10.1037/0003-066X.61.3.192

*Moloy, C. (2016). *Examining growth mindset in an inpatient population: Do youth with significant emotional and behavioral challenges think change is possible?* [Doctoral dissertation]. William James College. https://search.proquest.com/docview/1810421939/abstract/3EAA4E3CBE254B84PQ/1

Murphy, M. C., & Dweck, C. S. (2016). Mindsets shape consumer behavior. *Journal of Consumer Psychology*, *26*(1), 127–136. https://doi.org/10.1016/j.jcps.2015.06.005

Nussbaum, A. D., & Dweck, C. S. (2008). Defensiveness versus remediation: Self-theories and modes of self-esteem maintenance. *Personality and Social Psychology Bulletin*, *34*(5), 599–612. https://doi.org/10.1177/0146167207312960

O'Brien, K., & Lomas, T. (2017). Developing a growth mindset through outdoor personal development: Can an intervention underpinned by psychology increase the impact of an outdoor learning course for young people? *Journal of Adventure Education and Outdoor Learning*, *17*(2), 133–147. https://doi.org/10.1080/14729679.2016.1232199

*O'Rourke, E., Haimovitz, K., Ballweber, C., Dweck, C., & Popović, Z. (2014). Brain points: A growth mindset incentive structure boosts persistence in an educational game. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3339–3348). https://doi.org/10.1145/2556288.2557157

Okonofua, J. A., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(19), 5221–5226. https://doi.org/10.1073/pnas.1523698113

*Outes-Leon, I., Sánchez, A., & Vakis, R. (2020). *The power of believing you can get smarter: The impact of a growth-mindset intervention on academic achievement in peru* (World Bank Policy Research Working Paper No. 9141). Social Science Research Network. https://papers.ssrn.com/abstract=3531336

*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

*Porter, T., Martinus, A., Ross, R., Cyster, C. F., & Trzesniewski, K. (2020). Changing learner beliefs in south african townships: An evaluation of a growth mindset intervention. *Social Psychological & Personality Science*, *11*(7), 991–998. https://doi.org/10.1177/1948550620909738

Pustejovsky, J. E. (2020). *Clubsandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections (0.4.2)* [R package]. https://github.com/jepusto/clubSandwich

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical bayes meta-analysis. *Journal of Educational Statistics*, *10*(2), 75–98. https://doi.org/10.3102/10769986010002075

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

*Rienzo, C., Rolfe, H., & Wilkinson, D. (2015). *Changing mindsets: Evaluation report and executive summary*. Education Endowment Foundation. https://eric.ed.gov/?id=ED581132

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

Rothman, A. J., & Sheeran, P. (2021). The operating conditions framework: Integrating mechanisms and moderators in health behavior interventions. *Health Psychology*, *40*(12), 845–857. https://doi.org/10.1037/hea0001026

*Rutledge, E. (2016). *Do implicit personality theories impact parental reactions to child transgressions? A randomized controlled trial* [Doctoral dissertation]. Northern Illinois University. http://search.proquest.com/pqdtglobal/docview/1824357261/abstract/EAD115D2EF624978PQ/1

Sarrasin, J. B., Nenciovici, L., Foisy, L.-M. B., Allaire-Duquette, G., Riopel, M., & Masson, S. (2018). Effects of teaching the concept of neuroplasticity to induce a growth mindset on motivation, achievement, and brain activity: A meta-analysis. *Trends in Neuroscience and Education*, *12*, 22–31. https://doi.org/10.1016/j.tine.2018.07.003

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, *10*, Article 813. https://doi.org/10.3389/fpsyg.2019.00813

*Schleider, J., & Weisz, J. (2018). A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*(2), 160–170. https://doi.org/10.1111/jcpp.12811

Schleider, J. L., Abel, M. R., & Weisz, J. R. (2015). Implicit theories and youth mental health problems: A random-effects meta-analysis. *Clinical Psychology Review*, *35*, 1–9. https://doi.org/10.1016/j.cpr.2014.11.001

Schleider, J. L., Burnette, J. L., Widman, L., Hoyt, C., & Prinstein, M. J. (2020). Randomized trial of a single-session growth mind-set intervention for rural adolescents' internalizing and externalizing problems. *Journal of Clinical Child and Adolescent Psychology*, *49*(5), 660–672. https://doi.org/10.1080/15374416.2019.1622123

*Schleider, J. L., & Weisz, J. R. (2016). Reducing risk for anxiety and depression in adolescents: Effects of a single-session intervention teaching that personality can change. *Behaviour Research and Therapy*, *87*, 170–181. https://doi.org/10.1016/j.brat.2016.09.011

*Schmidt, J. A., Shumow, L., & Kackar-Cam, H. Z. (2017). Does mindset intervention predict students' daily experience in classrooms? A comparison of seventh and ninth graders' trajectories. *Journal of Youth and Adolescence*, *46*(3), 582–602. https://doi.org/10.1007/s10964-016-0489-z

*Schrodt, K. M. (2015). *The relationship among mindset instruction, kindergarteners' performance, and motivation in writer's workshop* [Doctoral dissertation]. Middle Tennessee State University. http://search.proquest.com/pqdtglobal/docview/1682475588/abstract/D80101F7A8504F4DPQ/1

Shaffer, C. T. (2014). *Self-theories of mental skill abilities in collegiate athletes* [Doctoral dissertation]. The Florida State University. http://search.proquest.com/pqdtglobal/docview/1556609127/abstract/82739A9612514D2FPQ/1

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

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

*Skipper, Y. (2015). An evaluation of an intervention to change first-year psychology students' theory of intelligence. *Psychology Teaching Review*, *21*(2), 69–80. https://files.eric.ed.gov/fulltext/EJ1146563.pdf

*Smith, E. N., Romero, C., Donovan, B., Herter, R., Paunesku, D., Cohen, G. L., Dweck, C. S., & Gross, J. J. (2018). Emotion theories and adolescent well-being: Results of an online intervention. *Emotion*, *18*(6), 781–788. https://doi.org/10.1037/emo0000379

*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

Stern, S. K. (2015). *Supporting growth mindset in upper-level high school students: A comparison of two classroom interventions* [Doctoral dissertation]. Northern Illinois University. https://commons.lib.niu.edu/handle/10843/18873

Stevens, C. N. (2018). *The effects of stress and growth mindset integration for students in remediation* [Doctoral dissertation]. Grand Canyon University. http://search.proquest.com/pqdtglobal/docview/2102491396/abstract/11EB5CC7B3144E8BPQ/1

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634. https://doi.org/10.3102/1076998615606099

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019a). A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Research Synthesis Methods*, *10*(2), 161–179. https://doi.org/10.1002/jrsm.1338

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019b). Current practices in meta-regression in psychology, education, and medicine. *Research Synthesis Methods*, *10*(2), 180–194. https://doi.org/10.1002/jrsm.1339

van Aalderen-Smeets, S. I., & Walma van der Molen, J. H. (2018). Modeling the relation between students' implicit beliefs about their abilities and their educational STEM choices. *International Journal of Technology and Design Education*, *28*(1), 1–27. https://doi.org/10.1007/s10798-016-9387-7

van Elk, M., Matzke, D., Gronau, Q. F., Guan, M., Vandekerckhove, J., & Wagenmakers, E.-J. (2015). Meta-analyses are no substitute for registered replications: A skeptical perspective on religious priming. *Frontiers in Psychology*, *6*, Article 1365. https://doi.org/10.3389/fpsyg.2015.01365

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, *13*(4), 411–417. https://doi.org/10.1177/1745691617751884

Vevea, J. L., Coburn, K. M., & Sutton, A. J. (2019). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 383–429). Sage Publications. https://doi.org/10.7758/9781610448864.21

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435. https://doi.org/10.1007/BF02294384

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(1), 1–48. https://doi.org/10.18637/jss.v036.i03

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

*Wanzek, J., Otaiba, S. A., Petscher, Y., Lemons, C. J., Gesel, S. A., Fluhler, S., Donegan, R. E., & Rivas, B. K. (2021). Comparing the effects of reading intervention versus reading and mindset intervention for upper elementary students with reading difficulties. *Journal of Learning Disabilities*, *54*(3), 203–220. https://doi.org/10.1177/0022219420949281

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*, Article 1832. https://doi.org/10.3389/fpsyg.2016.01832

*Wilkins, P. B. B. (2014). *Efficacy of a growth mindset intervention to increase student achievement (UMI No. 3642232)* [Doctoral dissertation]. ProQuest Dissertations and Theses database.

Yeager, D. S., Carroll, J. M., Buontempo, J., Cimpian, A., Woody, S., Crosnoe, R., Muller, C., Murray, J., Mhatre, P., Kersting, N., Hulleman, C., Kudym, M., Murphy, M., Duckworth, A. L., Walton, G. M., & Dweck, C. S. (2022). Teacher mindsets help explain where a growth-mindset intervention does and doesn't work. *Psychological Science*, *33*(1), 18–32. https://doi.org/10.1177/09567976211028984

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., 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., Miu, A. S., Powers, J., & Dweck, C. S. (2013). Implicit theories of personality and attributions of hostile intent: A meta-analysis, an experiment, and a longitudinal intervention. *Child Development*, *84*(5), 1651–1667. https://doi.org/10.1111/cdev.12062

*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., Trzesniewski, K. H., & Dweck, C. S. (2013). An implicit theories of personality intervention reduces adolescent aggression in response to victimization and exclusion. *Child Development*, *84*(3), 970–988. https://doi.org/10.1111/cdev.12003

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. (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

Received September 21, 2020
Revision received July 12, 2022
Accepted July 15, 2022 ■