AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Bulletin

ISSN: 0033-2909

2023, Vol. 149, Nos. 3–4, 242–258
https://doi.org/10.1037/bul0000394

REPLY

A Spotlight on Bias in the Growth Mindset Intervention Literature: A Reply to Commentaries That Contextualize the Discussion (Oyserman, 2023; Yan & Schuetze, 2023) and Illustrate the Conclusion (Tipton et al., 2023)

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

Two meta-analyses examined the effects of growth mindset interventions. Burnette et al. (2023) tested two moderators and found that effects ranged from negative to positive. We (Macnamara & Burgoyne, 2023) tested 11 preregistered moderators and examined the evidence according to a well-defined set of best practices. We found major areas of concern in the growth mindset intervention literature. For instance, 94% of growth mindset interventions included confounds, authors with a known financial incentive were two and a half times as likely to report positive effects, and higher quality studies were less likely to demonstrate a benefit. Yan and Schuetze (2023) contextualized these findings by describing problems with mindset theory and its measurement. Likewise, Oyserman (2023) discussed how growth mindset is a culturally fluent idea; papers supportive of growth mindset are widely embraced, whereas papers taking a skeptical approach are challenged. In another commentary, Tipton et al. (2023) challenged our results, claiming to produce positive effects by reanalyzing our data set using Burnette et al.'s (2023) approach. However, in addition to changing the approach, Tipton et al. changed effect sizes, how moderators were coded, and which studies were included, often without explanation. Though we appreciate the discussion of multiple meta-analytic approaches, we contend that meta-analytic decisions should be a priori, transparently reported, and consistently applied. Tipton et al.'s analysis illustrated our (Macnamara & Burgoyne's, 2023) conclusion: Apparent effects of growth mindset interventions on academic achievement may be attributable to inadequate study design, reporting flaws, and bias.

Public Significance Statement
This reply highlights frequent bias in the growth mindset intervention literature quantified in a recent meta-analysis. Two commentaries contextualized this bias, one by describing broader problems with growth mindset theory and its measures, the other by describing the reason for bias associated with intuitively appealing ideas like growth mindset. A third commentary argued for the benefits of growth mindset interventions by reanalyzing data from the meta-analysis after making substantive post hoc and inconsistent changes to the data set, thereby illustrating the conclusion of the original meta-analysis: Apparent benefits of growth mindset interventions on academic achievement may be due to inadequate study designs, flawed reporting, and bias.

Keywords: mindset, meta-analysis, educational interventions, best practices, academic achievement

Supplemental materials: https://doi.org/10.1037/bul0000394.supp

Brooke N. Macnamara https://orcid.org/0000-0003-1056-4996

The authors' data set, associated with the target article (Macnamara & Burgoyne, 2023), can be found at https://osf.io/ajhxv/. The authors have not previously disseminated data or interpretations unique to this article. We worked with the accepted versions of the three commentaries. These versions of the commentaries were prior to copyedits but should remain substantively the same as the final published versions. Likewise, we worked with the supplemental materials Tipton et al. (2023) posted on November 7, 2022. As of this writing, those files have not been modified.

Correspondence concerning this article should be addressed to Brooke N. Macnamara, Department of Psychological Sciences, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, United States. Email: brooke.macnamara@case.edu

Two meta-analyses examining the effects of growth mindset interventions were independently submitted to *Psychological Bulletin* around the same time. Our meta-analysis (Macnamara & Burgoyne, 2023) focused on the effects of growth mindset interventions on academic achievement. Burnette et al.'s (2023) meta-analysis examined the effects of growth mindset interventions on multiple outcomes, one of which was academic achievement.

The two meta-analyses both sought to examine the efficacy of growth mindset interventions. Growth mindset interventions are implemented in classrooms around the world, but the efficacy of these interventions is not well-established. Sometimes effects are found in samples of interest (e.g., Yeager et al., 2019) and sometimes they are not (e.g., Foliano et al., 2018). Meta-analyses offer a way to test hypotheses about intervention efficacy using a body of research. The two meta-analyses used different approaches, offered different amounts of transparency and justification for their decisions, and reached different conclusions about the efficacy of growth mindset interventions.

Burnette et al. (2023) focused on subgroup differences. They acknowledged that "null and even negative (in the case of academic achievement) effects are to be expected in growth mindset interventions" (p. 200). However, a key takeaway was that positive effects were more likely to be observed in subsamples that the original study authors had selected as their focal groups.

We (Macnamara & Burgoyne, 2023) examined the evidence according to a well-defined set of best practice criteria (Appelbaum et al., 2018; Boot et al., 2013; Simons et al., 2016). We found (a) that authors with a known financial incentive to report positive effects were more likely to report positive effects, (b) evidence of publication bias, and (c) that most growth mindset intervention studies contain major threats to internal validity (e.g., not isolating the effect of growth mindset from other treatment factors). Furthermore, higher quality studies were less likely to demonstrate a benefit of growth mindset interventions. We concluded that apparent effects of growth mindset interventions on academic achievement are likely attributable to inadequate study design, reporting flaws, and bias.

Three Commentaries

Three commentaries on the two meta-analyses also appear in this issue. Although this reply is primarily in response to Tipton et al. (2023), we briefly review all commentaries to provide a broader context.

Yan and Schuetze (2023)

Yan and Schuetze (2023) identified problems with mindset theory, its measurement, and the study designs used in growth mindset interventions. They noted that recent work suggests that the primary measure of mindset (Dweck, 1999) lacks response process validity (Limeri et al., 2020). That is, respondents differ in how they interpret the measure's items, and these differing interpretations are associated with different patterns of responses reflecting "fixed" or "growth" mindsets. They also highlighted our finding that 94% of the included growth mindset interventions had differences between treatment and control conditions other than training a growth mindset. Taken together with previous findings that intervention effects are numerically larger when they *fail* to influence students' growth mindset than when they succeed (Sisk et al., 2018), Yan and Schuetze suggest that either growth mindset may not be the critical ingredient in growth mindset interventions, or that results may be due to measurement problems. Either way, the internal validity of growth mindset interventions is undermined, rendering the mechanism driving effects unclear.

Yan and Schuetze (2023) described how we (Macnamara & Burgoyne, 2023) explored theoretically meaningful moderators and found no significant moderation effects. In contrast, they described how Burnette et al. (2023) "took a less theoretically-driven approach in examining moderators" (p. 212) by creating focal groups based on the original study authors' identification of key subsamples. Yan and Schuetze cautioned that the authors of the original studies may have identified these groups post hoc, and therefore that the subgroup effects Burnette et al. (2023) report are potentially inflated.

Oyserman (2023)

Oyserman (2023) explained how growth mindset is a culturally fluent idea, meaning that it is intuitively appealing because it fits with culture-based assumptions. She explained how the culturally fluent approach comes at the risk of researchers and reviewers having the sense that growth mindset "feels right" and adopting a less critical lens when evaluating the theory.

In contrast, a culturally disfluent approach is intuitively unappealing because it counters culture-based assumptions. Oyserman explained how a culturally disfluent approach is often met with disbelief and suspicion. As such, compared with researchers evaluating a culturally fluent idea that is readily accepted by readers and reviewers, researchers taking a culturally disfluent approach must often be more rigorous in their evaluation of a theory.

Oyserman described how Burnette et al. (2023) took a culturally fluent approach, whereas we (Macnamara & Burgoyne, 2023) took a culturally disfluent approach. A culturally disfluent approach is more likely to align with metascience best practices including increased disclosure, greater coverage, and an emphasis on transparency. Oyserman explained how the two meta-analyses converged on the evidence and how the culturally fluent versus culturally disfluent approaches led to different methodological and reporting decisions.

Oyserman (2023) further explained that the culturally disfluent approach leads to more systematic evaluations and draws attention to gaps that are often missed. This approach improves scientific inquiry. Despite the greater service to science, Oyserman (2023) pointed out that researchers who highlight gaps in a culturally fluent idea may face an uphill battle against proponents of the idea.

Tipton et al. (2023)

The commentary by Tipton et al. (2023) differed substantially from the other two commentaries. The authors of the other two commentaries generally agreed that our theoretical and methodological approach was more rigorous than in Burnette et al. (2023). In contrast, the authors of this commentary implied that our approach led to inaccurate conclusions and concluded that Burnette et al.'s (2023) approach offers a promising example of best practices. In their commentary, Tipton and colleagues made five main arguments, which we address in turn:

1. They asserted that a primary reason the conclusions differed between the two meta-analyses was due to different analytical approaches, and that Burnette et al.'s (2023) approach of clustering effects and conducting simultaneous meta-regression was superior.
2. They contended that when they applied Burnette et al.'s (2023) analytical approach to our data set, our data set demonstrated positive effects.
3. They criticized many of our preregistered best practice criteria, claiming that they should not be best practice criteria and that we made errors in coding criteria. They further claimed that when they corrected these errors, financial incentives did not significantly moderate effects, and the results of our model examining the best available evidence demonstrated positive effects.
4. They argued that researchers should prioritize focusing on heterogeneity.
5. They claimed that benefits of growth mindset interventions are well-established.

Differences Between Macnamara and Burgoyne (2023) and Burnette et al. (2023)

Tipton et al. (2023) argued that a primary reason the two meta-analyses differed was our analytical approaches. For example, they claimed that we only included one effect size per study, but our open-access data (https://osf.io/ajhxv/) and Figures 5, 7, and 8 in Macnamara and Burgoyne (2023) show that we included multiple effect sizes per study.[1] We included all relevant within-study effects when enough information was available so as not to exclude participants. In cases where original study authors selectively reported relevant subsamples, we included those subsample effects in relevant moderator analyses.

The commentary authors also claimed that we split the data into small subgroups for moderator analyses. All samples were coded for every moderator and all relevant subgroups were entered into the moderator analyses; see our open data (https://osf.io/ajhxv/) and Tables 4–6 and 10–12 in Macnamara and Burgoyne (2023). In some cases, few studies were available to contribute to a level of a moderator (e.g., elementary school children when examining developmental stage of participants) leading to small subgroups. Thus, any small subgroups in our moderator analyses reflected the body of literature.

Tipton et al. (2023) put forth a universal statement for meta-analysts working with large heterogeneous literatures: moderators should be "tested simultaneously in meta-regression" (p. 232). They stated that Burnette et al.'s (2023) approach of simultaneously analyzing moderators is an example of best practices and implied that our approach of separately examining moderators led to false conclusions. Why did the two meta-analyses' approaches to moderator analysis differ? Burnette et al. (2023) focused on two moderators: the degree to which the intervention was administered as intended and whether the subsample was considered a "focal group" in the original study. Using meta-regression, they simultaneously entered them into the analysis. This approach was appropriate given the number of moderators they examined.

Not including our analyses of bias, we examined 11 theoretical and methodological moderators (e.g., student level of risk for poor grades, socioeconomic status, type of control group, student developmental stage). Simultaneous meta-regression was inappropriate for our data set: due to the large amount of sampling error in estimating meta-regression weights, over 200 effect sizes (around double the number available) would be needed to simultaneously analyze 11 moderators and obtain appropriate predictive value from the regression weights (Schmidt, 1971, 2017). We, therefore, reported the correlations among the moderators and conducted the moderator analyses separately for adequate power to observe potential moderation effects.

Meta-analysts should conduct the analyses that are most appropriate to answer their research questions and for the data set in question. The best approach may or may not entail analyzing moderators simultaneously. For example, in a recent meta-analysis, Tipton and her coauthor estimated separate models for each moderator (Pustejovsky & Tipton, 2022). Although Pustejovsky and Tipton (2022) do not explain why they separately examined moderators in their meta-analysis, as opposed to running the moderator analyses simultaneously, we presume that this approach was most appropriate for their research questions and data set. Likewise, separately examining moderators was appropriate for our data set. Indeed, in Tipton et al.'s (2023) reanalysis of our data, they selected zero, one, or two moderators to analyze, depending on the model. In other words, Tipton et al. did not estimate all moderators simultaneously using our data set. Rather than issuing a universal statement that does not universally apply, we encourage future meta-analysts to preregister moderators and their analytical approach to best answer their research questions given the attributes of the data set.

We suggest there are three main differences between the two meta-analyses that the authors of the other commentaries also noted. For example, Oyserman (2023) pointed out that the two meta-analyses are not directly comparable because Burnette et al. (2023) only included a subset of the relevant studies of academic achievement included in Macnamara and Burgoyne (2023). We (Macnamara & Burgoyne, 2023) included 63 studies, whereas Burnette et al. (2023) included half as many: 32 studies of academic achievement.[2] The different data sets are a main difference between the two meta-analyses.

A second major difference between the meta-analyses was in the examinations of student subgroups (Oyserman, 2023; Yan & Schuetze, 2023). We specified characteristics of subgroups a priori that were hypothesized by mindset theory to demonstrate greater or lesser treatment effects (e.g., developmental stage, level of academic challenge, socioeconomic status), and preregistered these subgroup characteristics of interest. We examined each characteristic independently to test whether a particular sample characteristic consistently demonstrated a relatively larger effect,

[1] For example, we included multiple effects from Wilson (2009), a study with a 2 × 2 design where the treatment effect was hypothesized to vary across the levels of the other factor. Oddly, Tipton et al. (2023) explicitly—and incorrectly—cite Wilson (2009) as an example of us combining effects into a single effect.

[2] Though the two meta-analyses had different inclusion criteria, we counted at least 23 studies (both published and unpublished) that met both meta-analyses' inclusion criteria that we (Macnamara & Burgoyne, 2023) included but Burnette et al. (2023) did not.

with the goal of identifying who might benefit most from growth mindset interventions.

In contrast, Burnette et al. (2023) examined "focal" subgroups that had been identified by the original study authors. Effects were coded as pertaining to a focal group or not. Samples coded as being a focal group included a mix of characteristics that differed from study to study, from women in a laboratory-based stereotype threat manipulation[3] to students from low-socioeconomic households to students holding fixed mindsets. Subgroup, or focal group, analyses are often selected post hoc without correction, resulting in low credibility and limited likelihood of being corroborated by other studies (Buyse, 1989; Moher et al., 2012; Sun et al., 2010, 2014; Wang et al., 2007). Thus, although focal subgroup status explained some degree of heterogeneity, this finding may have little influence on theory development: It suggests that larger effects are more likely to occur in subgroups that original study authors identified as focal subgroups, not that any particular sample characteristic consistently demonstrated a larger effect (Sun et al., 2014).

A third major difference is that we aimed to reduce bias in our analyses and to evaluate bias in the growth mindset intervention literature (Oyserman, 2023; Yan & Schuetze, 2023). We (Macnamara & Burgoyne, 2023) preregistered our hypotheses, search protocol, theoretical moderators, methodological moderators, and planned analyses. We followed clear rules for how to calculate effect sizes when multiple options were available and how to form subgroups. We additionally preregistered a set of best-practice criteria in study design, reporting, and avoiding bias (see https://osf.io/ga9jk). These best-practice criteria were informed by recommendations for examining psychological interventions (Boot et al., 2013; Simons et al., 2016) and reporting standards for quantitative research (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008; Appelbaum et al., 2018). We also preregistered our methods for coding the original studies' adherence to these best practices.

Differing from Macnamara and Burgoyne (2023), Burnette et al. (2023) had no preregistration. They stated, "we did not preregister the search protocol or decisions outlined below" (p. 185). It is unclear which decisions by Burnette et al. were made a priori and which were made post hoc. Post hoc moderator selection contributes to highly inflated effects (Schmidt, 2017; Schmidt & Hunter, 2015; Thompson & Higgins, 2002).

Yet, Tipton et al. (2023) argued that our preregistered meta-analysis is an example of poor methodological practices, whereas Burnette et al.'s (2023) nonpreregistered meta-analysis is an example of best practices. As Oyserman (2023) described, culturally fluent ideas like growth mindset tend to be easily accepted without scrutiny, whereas culturally disfluent perspectives are likely to be heavily criticized by proponents of an idea.

Tipton and colleagues are proponents of growth mindset interventions, for example, Tipton and several of her coauthors are part of an initiative to implement growth mindset interventions in classrooms around the world ("The Global Mindset Initiative," e.g., Tipton et al., 2021). Multiple authors of the commentary also have a financial incentive to report benefits of growth mindsets. For example, Yeager is registered with a speakers' bureau where he can be hired to give keynote speeches on how growth mindset facilitates resilience and achievement (see supplemental Tables S1 and S2). There is a human tendency to give greater weight to information that supports existing beliefs and to seek to discredit contradictory information (Nickerson, 1998). Burnette et al.'s (2023) results support the idea that growth mindset interventions are worthwhile; our results question this conclusion.

Different Conclusions From a Changed Data Set

Tipton et al. (2023) claimed that applying Burnette et al.'s (2023) approach to our data leads to a different conclusion "from the exact same data set" (p. 231), namely, a significant effect for at-risk groups. However, Tipton et al. did not use the exact same data set: rather, they made multiple changes beyond disaggregating dependent effects. These changes ranged from substituting effect sizes that did not take into account baseline differences to changing risk statuses without explanation to coding effects from different studies as coming from the same study.

In total, Tipton et al. (2023) altered information beyond disaggregating effects in more than a third of the studies (22 of 63). In Table 1, we focus on the changes they made to the effect sizes and at-risk statuses in our data set. In a footnote, Tipton et al. noted the criteria for making changes to the data set; however, they did not apply these criteria consistently. They also made other changes outside these criteria without explanation.

For instance, Tipton et al. (2023) changed the effect size for Bostwick (2015), which, accounting for baseline differences, demonstrated that the treatment group improved numerically less than the control group from pre- to postintervention: $d = -0.16$. They replaced this effect size with postintervention group differences that erroneously suggest that the treatment group improved more than the control group: ds = 0.10 and 0.24. However, Tipton et al. did not consistently apply this rule. For example, they did not change the effect size for Coates (2016), where replacing the effect size accounting for baseline differences to postintervention group differences would have lowered the effect size from $d = -0.01$ to $d = -0.18$. We (Macnamara & Burgoyne, 2023) consistently accounted for baseline differences when possible, regardless of the effect size produced.

As another example, Tipton et al. (2023) included a low-socioeconomic status (SES) subsample along with the whole sample in the Churches and Educational Development Trust (2016) studies, where the low-SES subsample effects were generally larger (ds = 0.05 to 0.15) than the whole sample effect sizes (ds = −0.13 to 0.08). With this decision, students who were from low SES backgrounds contributed double the number of effect sizes compared with students who were not from low SES backgrounds in these studies, increasing the size of the meta-analytic estimated effects. However, Tipton et al. did not consistently include effects from low-SES subsamples. For example, in Foliano et al. (2018), they excluded the low-SES subsample effect sizes from their analysis; the low-SES subsample effects were nearly identical to the whole sample effect size—all within 0.01 of $d = 0.00$. In this case, including the low-SES subsample would have lowered both the overall meta-analytic estimate of the effect and the meta-analytic estimate for the

[3] The study contributing this focal group (Aronson et al., 2002) in Burnette et al.'s (2023) meta-analysis explicitly violated Burnette et al.'s inclusion criteria: studies must be "an intervention in an applied setting" (p. 183) and cannot be "lab-based studies that featured an experimental manipulation" (p. 183).

Table 1

Effect Sizes and At-Risk Status Codes From Our Data Set That Tipton et al. (2023) Changed for Their Reanalysis Despite Claiming It Was the Same Data Set

	Macnamara and Burgoyne's (2023) data set		Tipton et al.'s (2023) changes to our data set	
Reference	Effect sizes by at-risk status	Justifications	Effect sizes by at-risk status	Description of changes
Anderson et al. (2016)	No risk status assigned: **$d = 0.11$**	As we described in our open data, we were unable to obtain detailed sample information for this study	Medium: **$d = 0.11$**	Tipton et al. provided no explanation for this change of risk status
Bostwick (2015)	Low: **$d = -0.16$**	Following our preregistered criteria, we used the effect size for the final exam rather than the midterm exam We followed our rule of accounting for baseline differences when possible and calculated group differences in change scores. Here, the treatment group improved numerically less than the control group from pre- to postintervention	Low: **$d = 0.24$** **$d = 0.10$**	Tipton et al. changed the effect size to those reported in Burnette et al. (2023): • Burnette et al. included the effects for both the midterm and the final exam • Burnette et al. used between-group posttreatment effect sizes, failing to account for baseline performance. As can be seen, this calculation erroneously suggests that the treatment had a positive effect, when in fact, the control group improved numerically more than the treatment group
Broda (2015) and Broda et al. (2018)	Medium: **$d = 0.05$**	Following our preregistered criteria, students were coded as medium risk because they were transitioning to a new school Following our preregistered criteria, we used the earlier unpublished version (Broda, 2015) because participants were excluded for the publication (Broda et al., 2018) We followed our rule of using the whole sample effect size when the original study authors created subgroups that did not correspond to preregistered moderators. Here, race/ethnicity was not a preregistered moderator (and most studies did not report effects by this factor), we therefore used the whole sample effect size In the unpublished version we used, Fall GPA was the only GPA reported	Low: **$d = 0.03$** **$d = -0.03$** **$d = -0.01$** High: **$d = 0.46$** **$d = 0.33$** **$d = 0.42$** **$d = 0.05$** **$d = -0.04$** **$d = 0.00$**	Despite all students transitioning to a new school and Tipton et al. coding other samples transitioning to a new school as medium risk, Tipton et al. coded all White students as low risk (not facing a situational challenge). Tipton et al. provided no explanation for this change in risk status Tipton et al. coded all Latino/a and Black students as high-risk of low achievement. Tipton et al. provided no explanation for this change in risk status and did not code risk by ethnicity in other studies Tipton et al. used effect sizes from a version of the study report that excluded multiple subgroups from their subgroup results. By using this version of the study with select subgroup effects, Tipton et al. excluded half of the subgroups and excluded 680 total students Tipton et al. counted GPA multiple times for each student: they included effects for Fall GPA and Spring GPA, *and* the effect for the average of Fall and Spring GPA
Churches and Education Development Trust (2016)	Low: **$d = -0.13$** **$d = 0.05$** $d = 0.08_{\text{dep. (0.05)}}$ $d = 0.01_{\text{dep. (0.05)}}$ $d = 0.05^{s}$ $d = 0.14^{s}$ $d = 0.15^{s}_{\text{dep. (0.14)}}$ $d = 0.12^{s}_{\text{dep. (0.14)}}$	Following our preregistered criteria, students were coded as low risk because they were neither transitioning to a new school nor low achieving Following our rule, dependent effect sizes were adjusted for dependency Following our rule, subsamples relevant for moderators (here, low SES: ds = 0.05 and 0.14) replaced the whole sample in the relevant moderator analysis	Low: **$d = -0.13$** **$d = 0.08$** **$d = 0.01$** High: **$d = 0.05$** **$d = 0.15$** **$d = 0.12$**	Tipton et al. coded the low-socioeconomic status (SES) subsamples as high-risk in this study They did not consistently code low-SES subsamples as high risk across studies. Tipton et al. provided no explanation for this change in risk status Tipton et al. indicated that we coded the low-SES subsamples as high risk. We did not. We examined SES as its own factor. Tipton et al. separated dependent effect sizes, but overweighted low-SES students by including the whole sample and the subsample together in the same analysis
Fabert (2014)	Medium: **$d = 0.09$**	Following our preregistered criteria, students were coded as medium risk because they were transitioning to a new school We followed our rule of using the whole sample effect size when the original study authors created subgroups that did not	Low: **$d = -0.10$** High: **$d = 0.20$**	Despite all students transitioning to a new school and Tipton et al. coding other samples transitioning to a new school as medium risk, Tipton et al. coded all male students transitioning to a new school as low risk (not facing a situational challenge). Tipton et al. provided no explanation for this change in risk status

(*table continues*)

Table 1 (*continued*)

Reference	Macnamara and Burgoyne's (2023) data set		Tipton et al.'s (2023) changes to our data set	
	Effect sizes by at-risk status	Justifications	Effect sizes by at-risk status	Description of changes
		correspond to preregistered moderators. Here, gender was not a preregistered moderator (and most studies did not report effects by this factor), we therefore used the whole sample effect size		Tipton et al. coded all women in the sample as high risk (low achieving) despite nearly identical academic achievement between women and men in the sample Tipton et al. stated that the study was designed to help women and women were the targeted group. This is misleading. Fabert (2014) hypothesized gender effects on self-efficacy and intention but had no hypotheses about the intervention being more effective for women's than men's academic achievement and did not target women for the academic achievement measure
Foliano et al. (2018)	Low: **$d = -0.003$** $d = -0.01_{\text{dep. } (-0.003)}$ $d = 0.00_{\text{dep. } (-0.003)}$ $d = 0.00_{\text{dep. } (-0.003)}$ $d = 0.00^{s}$ $d = 0.00^{s}_{\text{dep. } (0.00)}$ $d = 0.01^{s}_{\text{dep. } (0.00)}$ $d = 0.00^{s}_{\text{dep. } (0.00)}$	Following our rule, dependent effect sizes were adjusted for dependency Following our rule, subsamples relevant for moderators (here, low SES: $d = 0.00$) replaced the whole sample in the relevant moderator analysis	Low: **$d = -0.01$** **$d = 0.00$** **$d = 0.00$**	Tipton et al. separated the dependent effect sizes, but did not include the low-SES subsamples as they did with other studies
Outes et al. (2017) and Outes-León et al. (2020)	No risk status assigned: **$d = 0.02$** $d = 0.03_{\text{dep. } (0.02)}$ $d = 0.01_{\text{dep. } (0.02)}$	As we described in our open data, demographic information for students or schools was not reported in the 2017 study report we used. (The 2020 report Burnette et al. and Tipton et al. used became available after our search stop date) Following our rule, dependent effect sizes were adjusted for dependency We followed our rule of using the whole sample effect size when the original study authors created subgroups that did not correspond to preregistered moderators. Here, "region of Peru" was not a preregistered moderator (and no other studies reported effects by this factor), we therefore used the whole sample effect size We used effect sizes for students who received the treatment	Low: **$d = 0.00$** **$d = 0.02$** **$d = 0.01$** **$d = -0.01$** **$d = -0.02$** High: **$d = 0.35$** **$d = 0.23$** **$d = 0.25$** **$d = 0.31$** **$d = 0.26$**	Tipton et al. indicated that we coded the sample as medium risk. We did not Tipton et al. separated dependent effect sizes by short-term effects, midterm effects, and school subject. They separated independent effects by region of Peru Tipton et al. coded risk status by region of Peru in this study. They misidentified urban schools outside of Lima as "rural." They coded Lima schools as low risk and urban schools outside of Lima as high risk. This coding is questionable. According to the 2020 study report, all samples were high risk (6%–12% satisfactory achievement) Tipton et al. used effect sizes that included treatment schools that never received the treatment materials (40% of treatment schools did not administer the treatment). Oddly, in this study using effects from students who never received the treatment substantially increased the effect of treatment
Paunesku et al. (2015)	Medium: **$d = -0.01$** High: $d = 0.20$	Following our preregistered criteria, the majority of students were coded as medium risk because they were transitioning to a new school The high risk subsample was at risk of failing. Following our rule, subsamples relevant for moderators (here, high risk: $d = 0.20$) replaced the whole sample in the relevant moderator analysis	Low: **$d = -0.01$** High: **$d = 0.11$**	Despite all students transitioning to a new school and Tipton et al. coding other samples transitioning to a new school as medium risk, Tipton et al. coded the majority of students as low risk in this study. Tipton et al. provided no explanation for this change in risk status Tipton et al. used Burnette et al.'s effect sizes for the low-risk and high-risk subsamples, which Burnette et al. estimated from a figure. These estimates appear incorrect. Burnette et al. misattributed the sample size of three separate interventions (mindset, belonging, and mindset + belonging) to the mindset intervention

(*table continues*)

Table 1 (*continued*)

Reference	Macnamara and Burgoyne's (2023) data set: Effect sizes by at-risk status	Macnamara and Burgoyne's (2023) data set: Justifications	Tipton et al.'s (2023) changes to our data set: Effect sizes by at-risk status	Tipton et al.'s (2023) changes to our data set: Description of changes
Rienzo et al. (2015)	Low: **$d = 0.05$** $d = 0.02_{\text{dep. (0.05)}}$ $d = 0.07_{\text{dep. (0.05)}}$ $d = 0.15^{s}$ $d = 0.07^{s}_{\text{dep. (0.15)}}$ $d = 0.22^{s}_{\text{dep. (0.15)}}$ High: $d = 0.33$ $d = 0.29_{\text{dep. (0.33)}}$ $d = 0.36_{\text{dep. (0.33)}}$	Following our rule, subsamples relevant for moderators (here, high risk: $d = 0.33$ and low-SES: $d = 0.15$) replaced the whole sample in the relevant moderator analysis Following our rule, dependent effect sizes were adjusted for dependency We followed our rule of accounting for baseline differences when possible and calculated group differences in change scores for all groups	Low: **$d = 0.02$** **$d = 0.07$** High: **$d = 0.38$** **$d = 0.34$**	Tipton et al. separated the dependent effect sizes, but overweighted the high-risk subsample by including the whole sample and the subsample together in the same analysis Tipton et al. did not include the low-SES subsample as they did with other studies. No explanation is given for this inconsistency Tipton et al. used Burnette et al.'s effect sizes for the high-risk subsample. Burnette et al. used between-group posttreatment effect sizes without accounting for baseline performance
Yeager, Romero et al. (2016)	Medium: **$d = -0.04$** High: **$d = 0.11$**	Following our preregistered criteria, the non-high-risk subsample of students was coded as medium risk because they were transitioning to a new school The high-risk subsample was lower achieving Following our rule, we separated relevant subsamples for the main analysis when enough information was available to do so without excluding students	Low: **$d = -0.04$** **$d = 0.01$** High: **$d = 0.11$** **$d = 0.13$**	Despite all students transitioning to a new school and Tipton et al. coding other samples transitioning to a new school as medium risk, Tipton et al. coded half the students as low risk in this study. Tipton et al. provided no explanation for this change in risk status Tipton et al. added two effect sizes that violated our inclusion criteria because they were dichotomized measures. Oddly, Tipton et al. argued against dichotomized measures (see p. 237), then added these dichotomized measures to the data set. Tipton et al. included effects for GPA and a dichotomized measure of GPA (% of students with low GPA) together for the same sample
Yeager et al. (2018) and Yeager et al. (2019)	Medium: **$d = -0.01$** High: **$d = 0.08$**	Following our preregistered criteria, the non-high-risk subsample of students was coded as medium risk because they were transitioning to a new school The high-risk subsample was lower achieving Following our preregistered criteria, we used the 2018 unpublished study report because participants were excluded for the 2019 publication	Low: **$d = -0.01$** High: **$d = 0.11$**	Despite all students transitioning to a new school and Tipton et al. coding other samples transitioning to a new school as medium risk, Tipton et al. coded half the students as low risk in this study. Tipton et al. provided no explanation for this change in risk status Tipton et al. stated that we should not have used the effect sizes from the posted preprint because those effect sizes were incorrect. Yeager et al. posted six versions of the preprint all with the same effect sizes and sample sizes. The published version excludes details of results from >6,000 participants (results are found in the extended data, in a note for a table, without indicators of which covariates are included), has a different high-risk sample size, and a larger effect size than the previous six versions of the study

Note. Low = effect sizes associated with samples coded as low risk of poor grades; Medium = effect sizes associated with samples coded as medium risk of poor grades (transitioning to a new school or under a stereotype threat manipulation); High = effect sizes associated with samples coded as high risk of poor grades. Bolded effect sizes are effect sizes included in the main analysis. Nonbolded effect sizes labeled as "High" replaced the whole sample when we tested for moderation by risk status. $_{\text{dep.}}$ = dependent effect that contributed to the dependency-adjusted effect size (in parentheses) that was entered in the analysis; ^{s} = low-socioeconomic status subsamples that replaced the whole sample when we tested for moderation by socioeconomic status. GPA = grade point average.

focal at-risk group they created for their analysis. Tipton et al. provided no explanation for the inconsistent inclusion/exclusion of low-SES subsample effects.

Tipton and colleagues (2023) also made changes to the at-risk status of samples. We coded samples where the majority of students were low achieving (e.g., at risk of failing) as "high risk" of poor grades, samples where the majority of students were facing a situational challenge (e.g., transitioning to a new school) as "medium risk" of poor grades, and samples where the majority of students were neither low achieving nor facing a situational challenge as "low risk." Tipton et al. made changes to these risk statuses, often without explanation. Likewise, their selection of subgroup characteristics was inconsistent and unexplained.

For instance, Tipton et al. (2023) changed the risk status of several samples yielding small effect sizes (ds = −0.04 to 0.03) from medium risk to low risk, but did not change the risk status of other subgroups with the same characteristics, without explanation. As a case in point, Yeager (an author of Tipton et al., 2023) contributed five studies, each with 9th-graders transitioning to a new school (Yeager et al., 2014 [two studies]; Yeager et al., 2019; Yeager, Lee, & Jamieson, 2016; Yeager, Romero et al., 2016). In each case, Yeager previously argued that these students were facing a situational challenge that increased their academic risk. In the reanalysis, Tipton et al. changed the risk status of two of the samples, whose average effect size was $\bar{d}$ = −0.02, from medium risk to low risk; they maintained the medium-risk status of the other samples in these studies, whose average effect size was $\bar{d}$ = 0.31.

After making these changes to our data set and changing who was coded as at-risk, Tipton and colleagues (2023) concluded that our data set demonstrated an effect for at-risk students. These changes precisely illustrate the conclusion of Macnamara and Burgoyne (2023): benefits of growth mindset interventions appear more likely to be observed when researchers make post hoc decisions and selectively report data.

Reanalysis of Our Data Set Using Burnette et al.'s (2023) Approach

Burnette et al.'s (2023) approach, like all meta-analytic approaches, has pros and cons. A benefit of simultaneous meta-regression is that, if one has enough power for the number of moderators under examination, controlling for the effects of other moderators can help to control for potential confounding from correlated moderators. A benefit of clustering effects is that all effects are available for all analyses.

A disadvantage of clustering is that it is not designed for partially dependent samples, which appeared in just over 10% of the studies in our data set. For example, Good et al. (2003) reported the treatment effect on reading scores for all students in the sample, but reported the treatment effect on math scores separately for boys and girls. In this case, the sample contributing to the effect for reading overlaps with each of the two subsamples contributing to the effects for math, but the two effects for math are independent of one another. Confusingly, in Tipton et al.'s (2023) reanalysis, they coded the whole sample as independent from its constituent subsamples, and coded the separate subsamples as though they were made up of the same students. In contrast, with the approach we used in Macnamara and Burgoyne (2023), we were able to adjust for these types of partially dependent samples.

In other cases, partially dependent effects came from studies where the original study authors provided effects for the whole sample and one or more subsamples (e.g., high-risk students), but not the remainder of the sample. Including both the whole sample and a subsample of the whole sample means that some students in the sample are contributing more effects than others. In these cases, the subsample completely overlaps with the whole sample and the whole sample partially overlaps with the subsample. Tipton et al. (2023) sometimes included subsamples along with the whole samples, but coded subsamples as having no overlap with the whole sample. Burnette et al. (2023) often only included the subgroup and excluded the whole sample effects. In Macnamara and Burgoyne (2023), we included the whole sample, and when non-independent subsamples were relevant for particular moderator analyses, we presented the results of the moderator analyses first with the whole samples and then with the relevant subsamples replacing the whole samples. We chose this approach because it included all data without treating dependent effects as independent or double counting effects from some students but not others.

Tipton et al. (2023) also coded several different studies as the same study, introducing errors into their cluster-adjusted approach. For example, they coded Peterson (2018), a study with elementary school children, as the same study as Paunesku et al. (2015), a study with high-school students. As another example, they coded Schubert (2017), a study with college students, as the same study as Saunders (2013), a study with middle school students. These studies were independent and varied in important aspects, which should have contributed to between-study heterogeneity rather than within-study heterogeneity. In total, Tipton et al. incorrectly coded 13% of the studies (8 of 63) as being the same as another study, erroneously increasing within-study heterogeneity.

Given the issues in handling partially dependent samples and the effects-to-moderators ratio, we believe the approach we used in Macnamara and Burgoyne (2023) was the best approach for our data set. There are, of course, multiple ways to analyze a data set. We subjected our data to the approach used by Burnette et al. (2023) and advocated by Tipton et al. (2023) where effects were clustered within samples, which were clustered within studies, and entered all moderators simultaneously in a metaregression. We coded subsamples and the whole sample they originated from as belonging to the same sample, as there is more dependence than independence in these cases.

When we analyze all effects using this approach, regardless of study quality, we find a small overall effect on academic achievement as we did in Macnamara and Burgoyne (2023): $\bar{d}$ = 0.07, 95% CI [0.02, 0.12], p = .005, τ^2 = .03 (see Macnamara & Burgoyne, 2023: $\bar{d}$ = 0.05, 95% CI [0.02, 0.09], p = .004, τ^2 = .01). Unsurprisingly, with this approach, we were unable to simultaneously analyze all moderators; there were insufficient data for the model to estimate all moderator effects simultaneously using cluster-robust standard errors.

We next reanalyzed the second model presented in Macnamara and Burgoyne (2023). Here, we tested whether growth mindset is the critical ingredient of growth mindset interventions. We included effects from all studies that demonstrated the treatment changed students' mindsets as intended, meeting the minimal standard of evidence to attribute treatment effects to growth mindset.

When we analyze these effects clustered within samples within studies, we do not observe a significant treatment effect on academic

achievement: $\bar{d} = 0.03$, 95% CI [−0.07, 0.13], $p = .522$, $\tau^2 = .03$. Similarly, we did not observe a significant treatment effect in Macnamara and Burgoyne (2023): $\bar{d} = 0.04$, 95% CI [−0.01, 0.10], $p = .146$, $\tau^2 < .01$.

Finally, we reanalyzed the third model presented in Macnamara and Burgoyne (2023). Here, we sought to evaluate the best available evidence—studies that influenced students' mindsets and met the most best practice criteria in study design, reporting, and avoiding bias. When we analyze these effects clustered within samples within studies, we do not observe a significant treatment effect on academic achievement: $\bar{d} = -0.003$, 95% CI [−0.11, 0.10], $p = .939$, $\tau^2 < .01$. Similarly, we did not observe a significant treatment effect in Macnamara and Burgoyne (2023): $\bar{d} = 0.02$, 95% CI [−0.06, 0.10], $p = .666$, $\tau^2 = .01$.

As can be seen, when evaluating the same data set using a different approach, the results are quite similar. In line with our previous report, an apparent small effect emerges when including all studies, regardless of quality control. Limiting inclusion to studies that demonstrated they changed students' mindsets as intended should produce a stronger effect if growth mindset is the critical ingredient in growth mindset interventions. Instead, in line with our previous report, there is no significant evidence of a treatment effect on academic achievement in this case. Likewise, when including only studies of the highest caliber available, in line with our previous report, the effect is nonsignificant.

Debate and Conclusions With Respect to Best-Practices Criteria and Designations

We (Macnamara & Burgoyne, 2023) listed 10 best-practices criteria essential for drawing causal conclusions. These best-practices criteria pertained to study design, reporting, and avoiding bias, and were informed by recommendations for psychological interventions (Boot et al., 2013; Simons et al., 2016) and reporting standards (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008). Each study was coded for each of these 10 best practices.

Many growth mindset intervention studies failed to follow best practices: 42% of samples failed to compare their treatment to an active control group. Ninety-four percent of samples failed to isolate the key treatment variable of interest. Seventy-two percent of samples failed to blind students, study administrators, and teachers to condition. Tipton and colleagues (2023) made no comment on these best practices. They did not express concern over the large number of studies that failed to adhere to these best practices, the threat to internal validity of those studies, or how results from those studies might skew the interpretation of effects in the growth mindset intervention literature.

We also included conducting an a priori power analysis as a best practice criterion and found that 75% of the growth mindset intervention study samples failed to conduct and report an a priori power analysis. Tipton et al. (2023) argued that conducting an a priori power analysis should not be included as a best practice criterion because (a) power analyses are not uniformly required by journals and (b) a priori power analyses are not a requirement in other guidelines. We note that a goal of best practices is to move beyond minimal requirements in order to produce better science. We also note that a priori power analyses are found in other study quality guidelines and have been since 2008 (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008; see also Appelbaum et al., 2018; Grant et al., 2018). Ninety-four percent of the studies in our meta-analysis were produced after this standard was implemented.

We also included random assignment to condition at the student level as a best practice criterion and found that nearly half of the studies (49% of samples) did not randomly assign students to condition. Tipton et al. (2023) argued that random assignment at the individual level should not be a best practice criterion because (a) assigning at the group level is common in education research and (b) existing standards require that random assignment to condition match the level of analysis, but do not state it must be at the student level. We note that though a practice might be common in education research, that does not mean it is a best practice. We also note that, as Tipton et al. acknowledged, the level of assignment to condition should match the level of analysis. Otherwise, sampling variance is underestimated, producing highly misleading significance test results (Hox, 1998; McCoach & Adelson, 2010). Researchers examining students' academic achievement analyze at the student level, therefore, they should randomize at the student level.

Moreover, if the level of assignment differs from the level of analysis, statistical adjustments need to be made to account for the design effect. As an example, in Blackwell et al.'s (2007) highly cited growth mindset intervention study, Blackwell et al. randomly assigned to condition at the group level but analyzed at the individual student level without correction. When adjusting for the design effect, the treatment effect in Blackwell et al.'s (2007) study is no longer statistically significant.

Another best practice criterion was checking whether the intervention changed students' mindsets as intended by testing for significant pre- to postintervention increases in growth mindset in the treatment group. For 41% of the samples, the researchers did not report whether the intervention changed treatment students' mindsets as intended. Tipton et al. (2023) argued that postintervention group differences between treatment and control on mindset should suffice as a manipulation check. We disagree. Postintervention differences can lead to erroneous conclusions if the groups differed preintervention (which is more likely to occur in small samples and when not randomly assigning to condition at the individual level). Ideally, researchers would ensure that the treatment students' mindsets shifted toward a growth mindset from pre- to postintervention and that the control students' mindsets did not shift during the same period. Few growth mindset intervention studies examined changes in mindset in this way. Our best practice criterion was more lenient than this standard while still providing a more robust test of whether the intervention changed treatment students' beliefs to more of a growth mindset than Tipton et al.'s suggested approach.

We also included preregistration as a best practice criterion. Tipton and colleagues (2023) appeared to argue against preregistration as a best practice criterion. That is, they described the quality measure as "anachronistic" (p. 236) for studies that were conducted before preregistration was introduced to psychology. However, the vast majority (89%) of the study reports were produced after preregistration was introduced to psychology. In total, very few studies (3% of samples) were preregistered. In fact, there were more nonpreregistered studies that claimed to be preregistered (six studies) than there were actual preregistered studies (two studies).

Despite seeming to argue against preregistration as a best practice criterion, Tipton et al. (2023) implied that preregistered studies offer the best evidence that growth mindset effects are well-established. They cited multiple "preregistered" studies as support for the benefits of growth mindset. However, none of those studies were actually preregistered. For example, Tipton et al. (2023) cited their own work, Yeager et al. (2019), as an example of a preregistered study. The "preregistration" is a document Yeager et al. (2019) wrote after analyzing a portion of the data to help "inform" the preregistration (see p. 3 of the document). Analyzing data to inform a preregistration violates the fundamental purpose of a preregistration.

Likewise, Tipton and colleagues (2023) claim that a "preregistered replication" by Zhu et al. (2019) verified Yeager et al.'s (2019) results. However, this study was a reanalysis of Yeager et al.'s (2019) data set that used the same poststudy document from Yeager et al. (2019) as its "preregistration." Claiming nonpreregistered studies are preregistered is a problem in the growth mindset intervention literature (Macnamara & Burgoyne, 2023). Likewise, the claim that Zhu et al. (2019) verified the results of Yeager et al. (2019) is questionable. Using the same data set, Zhu et al. (2019) found an effect almost half the size of what Yeager et al. (2019) reported for the focal group of lowerachieving students: 0.06 versus 0.10 average grade points.

Best Practices in Avoiding Conflicts of Interest

We (Macnamara & Burgoyne, 2023) also included a best practice criterion that authors of growth mindset intervention studies should not have a financial incentive to demonstrate benefits of growth mindset interventions. We did not code a study having grant funding as having a financial incentive. Rather, financial incentives were coded based on whether a study author received additional personal income, outside their faculty/researcher salary, aligned with reporting a particular outcome. We coded studies as having one or more authors with a perceived financial conflict of interest if any author was registered with a speakers' bureau to give motivational, inspirational, or keynote speeches on the benefits of growth mindset; if any author cofounded, was employed by, or served as a consultant to an organization or for-profit company that promotes or sells growth mindset services or products; and/or if any author profits from book royalties claiming benefits of growth mindset.

Having financial incentives to demonstrate benefits does not preclude rigorous research. Researchers with financial incentives undoubtedly view their own work as unbiased (Simons et al., 2016). Despite beliefs about objectivity, investigations from fields such as medicine suggest that researchers' decisions during the study design and reporting processes are influenced by financial incentives that are aligned with reporting a particular outcome (Simons et al., 2016; e.g., see Bekelman et al., 2003; Garg et al., 2005; Perlis et al., 2005).

Tipton et al. (2023) argued that several researchers we coded as having a financial incentive did not have a financial incentive. For example, they argued that Carol Dweck does not have a financial incentive to report benefits of growth mindsets. We coded studies authored by Dweck as having one or more authors with a financial incentive for multiple reasons: Dweck is registered with several speakers' bureaus to give motivational speeches to corporations about growth mindset, she cofounded a for-profit company that sells growth mindset products, and she profits from sales of her bestselling self-help book "Mindset." Tipton et al. argued that because Dweck divested from the for-profit company she cofounded that she does not have a financial incentive. They neither comment on Dweck's lucrative income from speaking engagements (Chivers, 2019) nor do they comment on her income from her book royalties. We hold that Dweck has financial incentives to demonstrate benefits of growth mindsets and that studies authored by Dweck are properly coded as having one or more authors with a financial incentive.

Tipton and colleagues (2023) also claimed that we defined financial conflicts of interest as any subsequent financial success by an author of a growth mindset intervention study. They described two cases in reference to this claim. In one case, they implied that coding Orvidas as having a financial incentive to find positive effects for her 2018 article is inappropriate because she only formed a consulting company 2 years later, in 2020. Critically, however, at the time Orvidas was conducting and publishing her mindset research, she was also working as a health and growth mindset coach (see supplemental Table S1).

The second case Tipton et al. describe is in regard to McDaniel, an author of Tipton et al. (2023). We (Macnamara & Burgoyne, 2023) coded a 2018 growth mindset intervention study authored by McDaniel (Fink et al., 2018) as having one or more authors with a financial incentive to report positive effects because McDaniel had previously written a popular press book in 2014 with a section describing benefits of growth mindset. Surprisingly, Tipton and colleagues argued that McDaniel should not count as having a financial incentive, claiming he wrote his popular press book after authoring his 2018 article. Contrary to their claim, the 2014 book was for sale and had been for years when the 2018 article was first submitted to the journal where it was later published. Thus, for neither Orvidas nor McDaniel was the financial incentive subsequent to the growth mindset intervention study as Tipton et al. claimed.[4]

Tipton et al. (2023) further stated that when they reanalyzed the data, authors with a financial incentive did not contribute significantly larger effect sizes than authors without a financial incentive. Yet, Tipton et al. made multiple changes to the financial incentive statuses in the data set before conducting their analysis. Further, these changes were inconsistently applied. For example, despite arguing that neither McDaniel nor Orvidas had a financial incentive at the time of their research, they changed one financial incentive status, but not the other. Specifically, they changed the financial incentive status of McDaniel's study ($d = 0.17$) to having no authors with a financial incentive—but maintained the financial incentive status of Orvidas' study ($d = -0.03$). For the sake of transparency, it would have been beneficial to provide an explanation for this inconsistency.

Of the studies Tipton et al. (2023) argued were incorrectly coded, the average effect size of those where they then changed the financial incentive status from having a financial incentive to "no authors with a financial incentive" for their analysis was $\bar{d} = 0.80$, whereas the average effect size of those where they maintained the financial incentive status for their analysis was $\bar{d} = 0.38$. By recoding some authors' financial incentives but not others, the overall effect size associated with having a financial incentive was reduced, as was the

[4] Tipton et al. (2023) also claimed that Mark A. McDaniel has not been involved with mindset research apart from one study in 2018. However, McDaniel, an author of Tipton et al. (2023), authored a second growth mindset intervention study less than 1 year ago (Fink et al., 2022).

likelihood of the financial incentive status moderator being statistically significant. Tipton and colleagues did not explain why they changed some financial incentive statuses, but not others, when they applied the same argument to both. Supplemental Table S1 provides (a) the rationale for why each study was originally coded as having one or more authors with a financial incentive; (b) Tipton et al.'s argument, if any, and whether or not they changed the status; and (c) comments on Tipton et al.'s argument and why the study's original financial incentive status is warranted. We hold that each study's original financial incentive status in Macnamara and Burgoyne (2023) was correctly coded.

Best Practices Model

We (Macnamara & Burgoyne, 2023) reported a model of studies that met the minimal standard of evidence (i.e., demonstrated the intervention influenced students' mindsets) and met at least 60% of the best practice criteria. The model yielded a nonsignificant effect. Tipton et al. (2023) argued that dichotomizing a continuous variable (number of best practices met) requires an arbitrary cutoff and is not a best practice. They also noted that testing small subgroups reduces power.

Rather than an arbitrary decision for a cutoff, we sought to include only studies that demonstrated the intervention changed students' mindsets and met 100% of the best-practices criteria. There were no such studies available. The lack of quality studies in the growth mindset intervention literature required us to lower our threshold to 60% to have enough studies to analyze. We preregistered this model, including the contingency plan for lowering the threshold in case few high-quality studies were available in the growth mindset intervention literature.

Additionally, rather than only testing one small subgroup, we explored the entire best practices space across the growth mindset intervention literature: we tested over 200 models for every combination and number of best-practices criteria met, with and without the criterion that the study demonstrated the intervention influenced students' mindsets, when at least five studies were available. Thus, we did not simply dichotomize the number of best-practices criteria met nor did we rely on a single model with a limited number of studies and power. We found that as the number of best practices adhered to increased, the number of significant models decreased.

Tipton et al. (2023) claimed to reanalyze this model using our criteria but with Burnette et al.'s (2023) approach. From their model, they reported a significant effect of treatment on academic achievement. Tipton and colleagues made multiple changes to the model by changing the inclusion criteria, reverse coding one of the best practice criteria, and claiming to fix errors. Their model has almost no overlap with the model we reported in Macnamara and Burgoyne (2023).

First, Tipton et al. (2023) used different inclusion criteria for their reanalysis. We included studies that demonstrated the intervention changed students' mindsets and met at least 60% best practice criteria. Tipton et al. did not use the first criterion, resulting in the inclusion of an additional nine studies (average $d = 0.12$) that failed to demonstrate that the intervention changed students' mindsets. We are unclear why they used different inclusion criteria for their reanalysis. They did so without explanation.

Second, Tipton and colleagues (2023) reverse-coded the financial incentive status for this model. In doing so, they counted all studies where one or more authors had a financial incentive to report positive effects as meeting the best practice that no authors had a financial incentive, and studies where no authors had a financial incentive as failing to meet this best practice criterion. In other words, studies where authors had a perceived financial conflict of interest were awarded higher quality ratings, whereas studies where no authors had a perceived financial conflict of interest were docked in their study quality rating (i.e., reducing their proportion of best practices met). This resulted in Tipton et al. excluding two studies, Hoang (2018), $d = 0.00$, and Sriram (2014), $d = -0.33$. These authors had no financial incentives but because of Tipton et al.'s reverse coding of the financial incentive variable, they were reported as having met fewer best practices and were excluded from the model. Tipton and colleagues provided no explanation for reverse coding this criterion for this model.[5]

Finally, Tipton et al. (2023) stated that the results we presented for this model were "an artifact of one study being coded erroneously and then included in the high-quality group, and two other studies that were erroneously excluded" (p. 236) and that when they "corrected these errors" (p. 236) they found a significant overall effect. Specifically, they excluded a negative effect size ($d = -0.68$) from Brougham and Kashubeck-West (2018), claiming that we coded it as having an a priori power analysis when it did not and that changing this status lowered it to below threshold for inclusion in this model. In our (Macnamara & Burgoyne's, 2023) open data, we directed readers to the location of additional information for this study, which includes the calculation of its a priori power analysis. Thus, we (Macnamara & Burgoyne, 2023) correctly coded this study as having an a priori power analysis, and it met the threshold to be included in our model of best available evidence.

Next, Tipton et al. (2023) added two nonpreregistered studies (Yeager, Lee, & Jamieson, 2016; Yeager, Romero et al., 2016), ds = −0.04, 0.01, 0.11, 0.13, and 0.16, claiming that they were preregistered and that changing these statuses meets the inclusion criteria for this model. The first study, Yeager, Lee, and Jamieson (2016), did not demonstrate the intervention changed students' mindsets, so it was not eligible for inclusion in the model. Furthermore, these two studies were not preregistered, see Table 2 (see also Macnamara & Burgoyne, 2023, Table 2). In short, one study has a wiki statement with no methods and was posted after study authors had processed the data, and the other study's document contained no hypotheses, no methods, and no planned analyses for the impact of a growth mindset on academic achievement. We (Macnamara & Burgoyne, 2023) coded these studies as nonpreregistered, and they failed to meet the threshold of best practices met to be included in our model of best available evidence.

Despite Tipton and colleagues (2023) claiming they conducted a reanalysis of our model, only three of 13 studies they included in their model are the same as in Macnamara and Burgoyne (2023); see Table 2. Tipton et al. provided no explanation for 10 of their 13 inclusions and exclusions. The three studies they excluded were the studies in our model whose results were most counter to the notion

[5] Tipton et al. (2023) reverse-coded financial incentive status in the best practices model but did not reverse-code financial incentive status when they analyzed financial incentive status as a moderator in a different model.

Table 2

Comparison of Macnamara and Burgoyne's (2023) Model of Best Available Evidence and Tipton et al.'s (2023) Reanalysis of This Model

	Macnamara and Burgoyne (2023)		Tipton et al. (2023)		
Reference	Status and explanation	*d*(s)	Status and explanation	*d*(s)	Comments
Brougham and Kashubeck-West (2018)	Included: this study met the inclusion criteria for this model	−0.68	Excluded: Tipton et al. (2023) stated that we wrongly coded this study as including an a priori power analysis when the article did not include any power calculations and after changing this status that the study no longer met threshold to enter into this model		Tipton et al. (2023) are incorrect. We accurately coded this study. As we noted in our open data (https://osf.io/ajhxv/), we were able to obtain additional study information from the first author's dissertation (Brougham, 2016). There, she presented the a priori power calculations for this study (see Brougham, 2016, p. 84)
Burnette et al. (2016)	Excluded: this study failed to demonstrate that the intervention changed students' mindsets		Included: no explanation given	0.08	Without explanation, Tipton et al. did not use the same inclusion criteria as Macnamara and Burgoyne (2023) in their reanalysis of this model
Burnette et al. (2018)	Excluded: this study failed to demonstrate that the intervention changed students' mindsets		Included: no explanation given	−0.03	Without explanation, Tipton et al. did not use the same inclusion criteria as Macnamara and Burgoyne (2023) in their reanalysis of this model
De Martino et al. (2018)	Included: this study met the inclusion criteria for this model	−0.07 0.24 0.24	Included	−0.07 0.24 0.24	Tipton et al. correctly included this study
Delpéche (2018)	Excluded: this study failed to demonstrate that the intervention changed students' mindsets		Included: no explanation given	0.16	Without explanation, Tipton et al. did not use the same inclusion criteria as Macnamara and Burgoyne (2023) in their reanalysis of this model
Hoang (2018)	Included: this study met the inclusion criteria for this model	0.00	Excluded: no explanation given		Without explanation, Tipton et al. coded studies where no authors had a financial incentive as not meeting the criterion that no authors had a financial incentive, dropping this study to below threshold
Holden et al. (2016)	Excluded: this study failed to demonstrate that the intervention changed students' mindsets		Included: no explanation given	0.02 −0.02	Without explanation, Tipton et al. did not use the same inclusion criteria as Macnamara and Burgoyne (2023) in their reanalysis of this model
Polley (2018)	Excluded: this study did not report a test of whether the intervention changed treatment students' mindsets		Included: no explanation given	0.05	Without explanation, Tipton et al. did not use the same inclusion criteria as Macnamara and Burgoyne (2023) in their reanalysis of this model
Rienzo et al. (2015)	Included: this study met the inclusion criteria for this model	0.02 0.07	Included	0.02 0.07 0.38 0.34	Tipton et al. correctly included this study. They included effects for the whole sample plus the high-risk subsample, but excluded the effects from the low-socioeconomic status subsample without explanation
Robinson (2019)	Excluded: this study did not report a test of whether the intervention changed treatment students' mindsets		Included: no explanation given	−0.08	Without explanation, Tipton et al. did not use the same inclusion criteria as Macnamara and Burgoyne (2023) in their reanalysis of this model

(*table continues*)

Table 2 (*continued*)

Reference	Macnamara and Burgoyne (2023)		Tipton et al. (2023)		Comments
	Status and explanation	*d*(s)	Status and explanation	*d*(s)	
Sriram (2014)	Included: this study met the inclusion criteria for this model	−0.33	Excluded: no explanation given		Without explanation, Tipton et al. coded studies where no authors had a financial incentive as not meeting the criterion that no authors had a financial incentive, dropping this study to below threshold
Yeager et al. (2014), Study 2	Excluded: this study did not report a test of whether the intervention changed treatment students' mindsets		Included: no explanation given	0.23	Without explanation, Tipton et al. did not use the same inclusion criteria as Macnamara and Burgoyne (2023) in their reanalysis of this model
Yeager et al. (2014), Study 3	Excluded: this study failed to demonstrate that the intervention changed students' mindsets		Included: no explanation given	−0.03 0.87	Without explanation, Tipton et al. did not use the same inclusion criteria as Macnamara and Burgoyne (2023) in their reanalysis of this model
Yeager, Lee, and Jamieson (2016)	Excluded: this study did not report a test of whether the intervention changed treatment students' mindsets. Also, this study met <60% best practice criteria		Included. Tipton et al. changed the preregistration status to "preregistered"	0.16	Without explanation, Tipton et al. did not use the same inclusion criteria as Macnamara and Burgoyne (2023) in their reanalysis of this model. Tipton et al. also changed the preregistered status from "no" to "yes," which increased the percent of best practices met to 60%. The preregistration contained no hypotheses, no methods, and no planned analyses for the impact of a growth mindset intervention on academic achievement. For this reason, we (Macnamara & Burgoyne, 2023) did not code it as a preregistered study
Yeager, Romero et al. (2016)	Excluded: this study met <60% best practice criteria		Included: Tipton et al. changed the preregistration status to "preregistered"	−0.04 0.11 0.01 0.13	Tipton et al. claimed that this study was preregistered. However, there is no preregistration associated with this study. A wiki statement with hypotheses but no methods exists. This wiki statement was posted approximately 3 months after data were collected and 3 weeks before the article was submitted. Yeager, Romero et al. (2016) state the study was preregistered before researchers received the data from the third-party research firm, but employees of the third-party research firm were authors of Yeager, Romero et al. (2016)
Yeager et al. (2018)	Included: this study met the inclusion criteria for this model	−0.01 0.08	Included.	−0.01 0.11	Tipton et al. correctly included this study. Tipton et al. changed the effect size for the lower achieving sample from the preprint version that included all participants, to the published version

Note. The inclusion criteria for Macnamara and Burgoyne (2023) for this model were (a) demonstrate that the intervention changed students' mindsets (significant pre- to posttreatment shift toward growth for treatment students) and (b) meet at least 60% of the best practice criteria.

that growth mindset interventions positively affect students' academic achievement, *d*s = −0.68 to 0.00.

Focusing on the Quality of the Evidence

When a literature has high rates of flawed study designs, selective reporting, and bias, researchers should prioritize conducting more rigorous research (Yan & Schuetze, 2023). Currently, the evidence suggests growth mindset intervention effects may be spurious and due to inadequate study design, flawed reporting, and bias. However, we can better assess whether growth mindset interventions are effective and if the effect varies systematically across theory-driven factors by focusing on improving study design, comprehensive reporting, and reducing bias in the growth mindset intervention literature. Systematic reviews and meta-analyses should focus on the quality of the evidence being evaluated.

We also appreciate the need to examine heterogeneity, which is why we reported heterogeneity estimates in multiple ways: I^2, τ^2, and 95% confidence intervals. We additionally tested 11 moderators to examine whether these theoretical and methodological factors could account for some of the moderate heterogeneity we observed. Indeed, examining heterogeneity and applying best practices should not be opposing goals; pitting one against the other is a false dichotomy. We can only establish whether replicable, positive effects in subgroups exist if researchers select characteristics of those subgroups consistently and a priori, and if the study followed best practices in study design, reporting, and avoiding bias. In contrast, inconsistent, post hoc selections of subgroup characteristics hinders scientific understanding of mechanisms. In short, researchers need to apply methodological rigor and evaluate study quality when testing for heterogeneity.

Growth Mindset Is Not Well Established

The mark of a strong theory is one where the evidence for the theory persists when researchers apply methodological rigor. If the mechanism underlying an effect is unclear, investigating causal mechanisms is a prime use of meta-analysis. Researchers conducting meta-analyses should attempt to make sense of conflicting findings and investigate the conditions under which a theory's hypotheses hold and when they do not. We found that some of mindset theory's hypotheses held when examining studies with problematic study designs, reporting flaws, and bias. In contrast, evidence for mindset theory's hypotheses was absent when applying quality control.

Conclusion

Bias and a lack of rigorous study design and reporting are major areas of concern in the growth mindset intervention literature (Burnette et al., 2023; Macnamara & Burgoyne, 2023; Oyserman, 2023; Yan & Schuetze, 2023). In our meta-analysis and systematic review (Macnamara & Burgoyne, 2023), we described common study design and reporting problems, along with likely bias, in the growth mindset intervention literature. Examples include failing to specify decisions a priori; manipulating multiple variables at once, such that the effect of growth mindset cannot be isolated; claiming nonpreregistered studies are preregistered; and having a financial incentive aligned with a particular outcome (Macnamara & Burgoyne, 2023).

Yan and Schuetze (2023) extended this discussion by describing how mindset theory is underspecified, how mindset measures lack response process validity, and how growth mindset may not be the critical ingredient in growth mindset interventions. They made specific suggestions the field can take to improve the research in this area, including (a) define measurable constructs, (b) specify testable process models, (c) identify subgroups a priori, and (d) present messaging around growth mindset as nuanced as the empirical evidence. Oyserman (2023) further contextualized the discussion by describing how growth mindset is a culturally fluent idea, meaning that it aligns with already-held beliefs. Culturally fluent ideas are often accepted with relatively little criticism. In contrast, when researchers question a culturally fluent idea like growth mindset, they are likely to be met with increased criticism and counterarguments by proponents.

Tipton et al. (2023) focused on differences in analytic approaches between the two meta-analyses. They argued that growth mindset interventions are well established, as evidenced by preregistered studies, and that by applying Burnette et al.'s (2023) analytical approach to our data set, the results are in line both with Burnette et al.'s (2023) results and other preregistered studies. However, in making these arguments, Tipton and colleagues made several concerning study design and reporting decisions common in the growth mindset intervention literature. For example, the preregistered studies they referenced in support of their argument were not actually preregistered. The decisions they made in their reanalysis were post hoc and appeared to favor positive outcomes. In many cases, they gave no rationale for those decisions. Finally, Tipton and colleagues did not disclose that multiple authors on the team have a financial incentive to report positive effects.

Though we appreciate the discussion of multiple meta-analytic approaches, we contend that meta-analytic decisions should be a priori, transparently reported, and consistently applied. Tipton et al. (2023) commentary and analysis illustrated our (Macnamara & Burgoyne, 2023) conclusion: benefits of growth mindset interventions appear most likely to emerge when authors make problematic design and analysis decisions, engage in selective reporting, and have financial incentives to demonstrate benefits of growth mindset interventions.

References

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

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

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

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping

theories of intelligence. *Journal of Experimental Social Psychology*, *38*(2), 113–125. https://doi.org/10.1006/jesp.2001.1491

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

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

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

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

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

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

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

Broda, M. D., 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. (2016). *Impact of a growth mindset intervention on academic performance of students at two urban high schools* [Unpublished doctoral dissertation]. University of Missouri-St. Louis.

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

Burnette, J. L., Billingsley, J., Banks, G. C., Knouse, L. E., Hoyt, C. L., Pollack, J. M., & Simon, S. (2023). A systematic review and meta-analysis of growth mindset interventions: For whom, how, and why might such interventions work? *Psychological Bulletin*, *149*(3–4), 174–205. https://doi.org/10.1037/bul0000368

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

Burnette, J. L., 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

Buyse, M. E. (1989). Analysis of clinical trial outcomes: Some comments on subgroup analyses. *Controlled Clinical Trials*, *10*(Suppl. 4), 187S–194S. https://doi.org/10.1016/0197-2456(89)90057-3

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

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

Coates, K. (2016). *An evaluation of Growing Early Mindsets (GEM™)* [Unpublished doctoral dissertation]. University of Oregon.

De Martino, S., Porter, T., Martinus, A., Ross, R., Cyster, C. F., Pons, G., & Trzesniewski, K. H. (2018). *Changing learner beliefs in South African townships—An evaluation of a growth mindset intervention* [Unpublished manuscript].

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

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.

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

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

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

Fink, A., Mattson, P. S., Cahill, M. J., Frey, R. F., & McDaniel, M. A. (2022). Growth mindset intervention improves performance of highly nontraditional introductory psychology students at a community college. *Teaching of Psychology*. Advance online publication. https://doi.org/10.1177/00986283211053806

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

Foliano, F., Rolfe, H., Buzzeo, J., Runge, J., & Wilkinson, D. (2018). *Changing mindsets: Effectiveness trial. Evaluation report*. Education Endowment Foundation.

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

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

Grant, S., Mayo-Wilson, E., Montgomery, P., Macdonald, G., Michie, S., Hopewell, S., Moher, D., & on behalf of the CONSORT-SPI Group. (2018). CONSORT-SPI 2018 Explanation and Elaboration: Guidance for reporting social and psychological intervention trials. *Trials*, *19*(1), Article 406. https://doi.org/10.1186/s13063-018-2735-z

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

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

Hox, J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways. Studies in classification, data analysis, and knowledge organization* (pp. 147–154). Springer. https://doi.org/10.1007/978-3-642-72087-1_17

Limeri, L. B., Choe, J., Harper, H. G., Martin, H. R., Benton, A., & Dolan, E. L. (2020). Knowledge or abilities? How undergraduates define

intelligence. *CBE Life Sciences Education*, *19*(1), Article ar5. https://doi.org/10.1187/cbe.19-09-0169

Macnamara, B. N., & Burgoyne, A. P. (2018, August 11). *Preregistration for Do growth mindset interventions impact students' academic achievement? A systematic review and meta-analysis with recommendations for best practices*. Open Science Framework. https://doi.org/10.17605/osf.io/ga9jk

Macnamara, B. N., & Burgoyne, A. P. (2021). *Open data for Do growth mindset interventions impact students' academic achievement? A systematic review and meta-analysis with recommendations for best practices*. Open Science Framework. https://doi.org/10.17605/osf.io/ajhxv

Macnamara, B. N., & Burgoyne, A. P. (2023). Do growth mindset interventions impact students' academic achievement? A systematic review and meta-analysis with recommendations for best practices. *Psychological Bulletin*, *149*(3–4), 133–173. https://doi.org/10.1037/bul0000352

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

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., Altman, D. G., & the CONSORT. (2012). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *International Journal of Surgery*, *10*(1), 28–55. https://doi.org/10.1016/j.ijsu.2011.10.001

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

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

Outes, I., Sanchez, A., & Vakis, R. (2017). *Growth mindset at scale: Impact of a psychosocial intervention on secondary school attainment in Peru*. Project: Growth Mindset at Scale—Increasing school attainment by affecting the mindset of pupils and teachers [Unpublished project report].

Outes-León, 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. *The World Bank*. https://doi.org/10.1596/1813-9450-9141

Oyserman, D. (2023). Culturally fluent theories, metascience and scientific progress: A case example: Commentary on Macnamara and Burgoyne (2023) and Burnette et al. (2023). *Psychological Bulletin*, *149*(3–4), 220–228. https://doi.org/10.1037/bul0000376

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

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

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

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

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

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

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

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

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, *31*(3), 699–714. https://doi.org/10.1177/001316447103100310

Schmidt, F. L. (2017). Statistical and measurement pitfalls in the use of meta-regression in meta-analysis. *Career Development International*, *22*(5), 469–476. https://doi.org/10.1108/CDI-08-2017-0136

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications. https://doi.org/10.4135/9781483398105

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

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

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

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

Sun, X., Briel, M., Walter, S. D., & Guyatt, G. H. (2010). Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ*, *340*, Article c117. https://doi.org/10.1136/bmj.c117

Sun, X., Ioannidis, J. P., Agoritsas, T., Alba, A. C., & Guyatt, G. (2014). How to use a subgroup analysis: Users' guide to the medical literature. *JAMA*, *311*(4), 405–411. https://doi.org/10.1001/jama.2013.285063

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, *21*(11), 1559–1573. https://doi.org/10.1002/sim.1187

Tipton, E. [@stats_tipton]. (2022a, November 7). *Ok, but now for my thoughts. I want to begin by noting that I am a statistician and that I* [Tweet]. Twitter. https://twitter.com/stats_tipton/status/1589690832281829376

Tipton, E. [@stats_tipton]. (2022b, November 7). *The short answer: Skip MA1. Read MA2. The answer isn't "yes" or "no"* [Examining the data in both MA1 and Tweet]. Twitter. https://twitter.com/stats_tipton/status/1589690830495051777

Tipton, E., Bryan, C., Murray, J., McDaniel, M. A., Schneider, B., & Yeager, D. S. (2023). Why meta-analyses of growth mindset and other interventions should follow best practices for examining heterogeneity: Commentary on Macnamara and Burgoyne (2023) and Burnette et al. (2023). *Psychological Bulletin*, *149*(3–4), 229–241. https://doi.org/10.1037/bul0000384

Tipton, E., Hedges, L., Yeager, D., Murray, J., & Gopalan, M. (2021). *Global mindset initiative paper 4: Research infrastructure and study design* [Unpublished manuscript]. SSRN.

Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J., & Drazen, J. M. (2007). Statistics in medicine—Reporting of subgroup analyses in clinical trials. *The New England Journal of Medicine*, *357*(21), 2189–2194. https://doi.org/10.1056/NEJMsr077003

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

Yan, V. X., & Schuetze, B. A. (2023). What is meant by "growth mindset"? Current theory, measurement practices, and empirical results leave much open to interpretation: Commentary on Macnamara and Burgoyne (2023) and Burnette et al. (2023). *Psychological Bulletin*, *149*(3–4), 206–219. https://doi.org/10.1037/bul0000370

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

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

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

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., Lee, H. Y., O'Brien, J., Flint, K., Roberts, A., Trott, J., Greene, D., Walton, G. M., & Dweck, C. S. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, *108*(3), 374–391. https://doi.org/10.1037/edu0000098

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

Zhu, P., Garcia, I., & Alonzo, E. (2019). *An independent evaluation of the growth mindset intervention. Issue Focus.* MDRC.

Received January 27, 2023
Revision received May 3, 2023
Accepted May 5, 2023 ■