Supplemental Online Materials

The purpose of this document is to provide supplemental methods and results.

Table of Contents

Table S1

Full-text Papers that Met Most Inclusion Criteria but Were Excluded with Reasons

Reference	Reason
Arroyo, I., Muldner, K., Schultz, S. E., Burleson, W., Wixon, N., & Woolf, B. P. (2016, January). Addressing affective states with empathy and growth mindset. In *6th International Workshop on Personalization Approaches in Learning Environments (Extended Proceedings)*.	No results reported comparing treatment and control.
Bagès, C. Verniers, C., & Martinot, D. (2016). Virtues of a hardworking role model to improve girls' mathematics performance. *Psychology of Women Quarterly, 40*(1), 55-64.	Ability attributed to effort, but did not teach that a human attribute was malleable.
Cole, J. A. (2017). Overcoming stereotypes that hinder academic performance through psychological priming. *Journal of the South Carolina Junior Academy of Science*, *15*(2), 12.	The intervention is described as teaching students that intelligence is malleable, but the paper students read is provided and there is no mention of malleability of any human trait.
Dallianus, M. (2018). *The relationship between mindset and developmental math at a community colllege*. Doctoral dissertation. Ferris State University.	Only reported a dichotomous outcome measure.
El-Abd, M., Callahan, C., & Azano, A. (2019). Predictive factors of literacy achievement in young gifted children in rural schools. *Journal of Advanced Academics*, *30*(3), 298-325.	Treatment group received two interventions (control group received neither).
Frey, R. F., Fink, A., Cahill, M. J., McDaniel, M. A., & Solomon, E. D. (2018). Peer-led team learning in general chemistry I: Interactions with identity, academic preparation, and a course-based intervention. *Journal of Chemical Education*, *95*(12), 2103-2113.	Mindset intervention results from this study are reported in Fink et al. (2018), which is included in our meta-analysis.
Glerum, J., Loyens, S. M., Wijnia, L., & Rikers, R. M. (2019). The effects of praise for effort versus praise for intelligence on vocational education students. *Educational Psychology*, 1-17.	Treatment group received praise for effort, but intervention did not teach that a human attribute was malleable.

Halliwell, B., Cohen, T., Cruz, T., Gallen, I., Mullarkey, F., & Petrozzino, J. (2017). *The effects of growth mindset intervention on vocabulary skills in first to third grade children.* Presentation at academic festival.	Not enough information to calculate an effect size.
Hong, H. Y., & Lin-Siegler, X. (2012). How learning about scientists' struggles influences students' interest and learning in physics. *Journal of Educational Psychology, 104*(2), 469–484.	Ability attributed to effort, but did not teach that a human attribute was malleable.
Jones, L. T. (2019). *Mindfulness, motivation, and mindset: The effects of positive language on students with deficiencies success*. Unpublished doctoral dissertation. Trevecca Nazarene University.	Effort encouragement, but did not teach that a human attribute was malleable.
Karumbaiah, S., Lizarralde, R., Allessio, D., Woolf, B., Arroyo, I., & Wixon, N. (2017). *Addressing student behavior and affect with empathy and growth mindset.* Proceedings of the 10th annual conference of the International Educational Data Mining Society.	No results comparing treatment and control groups.
Lin-Siegler, X., Ahn, J. N., Chen, J., Fang, F.-F. A., & Luna-Lucero, M. (2016). Even Einstein struggled: Effects of learning about great scientists' struggles on high school students' motivation to learn science. *Journal of Educational Psychology, 108*, 314–328.	Ability attributed to effort, but did not teach that a human attribute was malleable.
Mendoza-Denton, R., Kahn, K., & Chan, W. (2008). Can fixed views of ability boost performance in the context of favorable stereotypes? *Journal of Experimental Social Psychology, 44*, 1187–1193.	Ability attributed to effort, but did not teach that a human attribute was malleable.
Sachs, T. L. (2017). *Growth mindset: How does it affect math achievement in second grade.* Unpublished thesis. University of Central Missouri	Not enough information to calculate an effect size. Could not contact author.
Schrodt, K. E., Elleman, A. M., FitzPatrick, E. R., Hasty, M. M., Kim, J. K., Tharp, T. J., & Rector, H. (2019). An examination of mindset instruction, self-regulation, and writer's workshop on kindergarteners' writing performance and	Not primarily a growth mindset intervention.

motivation: A mixed-methods study. *Reading & Writing Quarterly*, *35*(5), 427-444.	
Schuman, C. L. (2017). *The impacts of teaching growth mindset strategies to students in inquiry science 2 at Ferndale High School.* Unpublished master's thesis. Montana State University.	No control group.
Wang, A. Y., Fuchs, L. S., Fuchs, D., Gilbert, J. K., Krowka, S., Abramson, R. (2019). Embedding self-regulation instruction within fractions intervention for third graders with mathematics difficulties. *American Psychologist*, *74*(9), 1086-1102.	Not primarily a growth mindset intervention.
Worrall, L. K. (2016). *Building academic tenacity by promoting growth mindsets in English.* Unpublished doctoral dissertation. Cardiff Metropolitan University.	No control group.
Zhao, Q., Wichman, A., & Frishberg, E. (2019). Self-doubt effects depend on beliefs about ability: Experimental evidence. *The Journal of General Psychology*, *146*(3), 299-324.	Intervention was not with students.

Note. This table does not include the published studies that were replaced with their unpublished versions due to discrepancies in sample sizes.

Additional Patterns of Significant Effects

When describing the patterns of significant effects in the State of the Literature section, we are referring to the effects for which we coded, i.e., the difference between treatment and an active control group (if available, passive or fixed mindset if not) on the most comprehensive measure of academic achievement, controlling for prior achievement if available, without other covariates, using the longest interval within the same academic context.

There were many more cases where authors reported at least one significant effect that was not the primary outcome or otherwise did not meet our criteria for inclusion (e.g., $p < .05$ if comparing treatment to a fixed mindset condition (when an active control was available); $p < .05$ if not accounting for prior achievement; $p < .05$ for one subgroup if including all covariates). If we consider *any* significant effect on academic achievement reported in the paper, then 44% of the papers report at least one significant effect on academic achievement. Authors with financial incentives reported a significant effect nearly twice as often as authors without financial incentives (67% vs. 35%), χ^2 (1, $N = 61$) = 5.20, $p = .023$.

Additional Publication Bias Analysis Information

PEESE Results

As noted in the main text, when the PET estimate of the true effect after accounting for publication bias is not statistically significant, the PET estimate from the conditional PET-PEESE analysis is used (Stanley & Doucouliagos, 2014). To reiterate, the PET analysis revealed a non-significant estimated true effect after accounting for publication bias, $d = 0.01$, 95% CI [-0.03, 0.05], $SE = .02$, $p = .67$. Furthermore, the PET analysis indicated the presence of publication bias, $b = .42$, 95% CI [.07, .76], $SE = .18$, $p = .02$. Here, we report the results of the PEESE analysis. The PEESE estimate of the true effect (i.e., the intercept) after correcting for publication bias and other small-study effects was $d = 0.04$, 95% CI [0.01, 0.07], $SE = .02$, $p = .02$. The PEESE analysis revealed a non-significant slope, $b = .61$, 95% CI [-0.03, 1.25], $SE = .33$, $p = .06$. Below, we present the R output from the conditional PET-PEESE analyses.

R Code for PET-PEESE Analyses and Output

Source: https://github.com/Joe-Hilgard/PETPEESE

```
PET=function(dataset, error = "additive") {
            if (error == "additive") {
               petOut = rma(yi = Cohen.d,
                       sei = Std.Err,
                  mods = ~Std.Err,
                  data=dataset,
                  method = "REML")
     }
     if (error == "multiplicative") {
       petOut = lm(Cohen.d ~ Std.Err,
                 weights = 1/Std.Err^2,
                 data=dataset)
     }
     return(petOut)
   }

PET(dataset)

Mixed-Effects Model (k = 79; tau^2 estimator: REML)

tau^2 (estimated amount of residual heterogeneity):     0.0031 (SE = 0.0019)
tau (square root of estimated tau^2 value):             0.0561
I^2 (residual heterogeneity / unaccounted variability): 27.34%
H^2 (unaccounted variability / sampling variability):   1.38
R^2 (amount of heterogeneity accounted for):            17.35%

Test for Residual Heterogeneity:
QE(df = 77) = 131.6293, p-val = 0.0001

Test of Moderators (coefficient 2):
QM(df = 1) = 5.5199, p-val = 0.0188

Model Results:

         Estimate      se    zval    pval    ci.lb   ci.ub
intrcpt    0.0097  0.0225  0.4303  0.6670  -0.0344  0.0538
Std.Err    0.4155  0.1769  2.3495  0.0188   0.0689  0.7622  *

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



PEESE=function(dataset, error = "additive") {
  if (error == "additive") {
    peeseOut = rma(yi = Cohen.d,
              sei = Std.Err,
              mods = ~I(Std.Err^2),
              data=dataset,
              method = "REML")
  }
  if (error == "multiplicative") {
    peeseOut = lm(Cohen.d ~ I(Std.Err^2),
                  weights = 1/Std.Err^2,
                  data=dataset)
  }
  return(peeseOut)
}

PEESE(dataset)
```

```
Mixed-Effects Model (k = 79; tau^2 estimator: REML)

tau^2 (estimated amount of residual heterogeneity):     0.0033 (SE = 0.0020)
tau (square root of estimated tau^2 value):             0.0570
I^2 (residual heterogeneity / unaccounted variability): 28.37%
H^2 (unaccounted variability / sampling variability):   1.40
R^2 (amount of heterogeneity accounted for):            14.64%

Test for Residual Heterogeneity:
QE(df = 77) = 133.6459, p-val < .0001

Test of Moderators (coefficient 2):
QM(df = 1) = 3.4920, p-val = 0.0617

Model Results:

              Estimate      se    zval    pval    ci.lb   ci.ub
intrcpt         0.0379  0.0166  2.2767  0.0228   0.0053  0.0705  *
I(Std.Err^2)    0.6118  0.3274  1.8687  0.0617  -0.0299  1.2534  .

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Best Practices Criteria Removed

Changes from the preregistration are outlined in the Appendix of the main text. Table S2 shows the original 18 best practices criteria (i.e., those included in the preregistration) and indicates which are included in the main text. We then provide the rationale for why the eight removed criteria were initially included and a brief rationale for their exclusion (further details can be found in the Appendix). Supplemental results including these original criteria for meta-analytic models are reported in this document.

Table S2

Original Best Practices Criteria and Whether or Not They Are Included in Model 3 in the Main Text

Best Practices Criteria	Included in Model 3
The intervention is compared to an active control condition	Yes
The intervention has no obvious confounds	Yes
Equivalent control group at baseline	No
A priori power analysis conducted	Yes
Individual-level random assignment to condition	Yes
Blinding	Yes
Including a manipulation check	Yes
Preregistered	Yes
Reporting results of those who participated	Yes
Reporting the relevant results of all subsamples	Yes
Comparing relevant sub-samples	No
Controlling the familywise error rate	No
Reporting confidence intervals or variance statistics	No
Reporting effect sizes for key outcomes	No
Conducting theory-driven analyses	No
Conducting and reporting analyses interpretable to most	No
Appropriately interpreting results	No
Having no financial incentive for a particular outcome	Yes

Equivalent Control Group at Baseline

We describe the importance of this study characteristic in the main text. Fifty-eight percent of the samples (37% of the total students) were from studies where the samples' groups did not have significantly different academic achievement at baseline.

Rational for Exclusion

Though non-equivalent baseline differences can make interpretation difficult, equivalent baselines are not a best *practice*. Baseline differences can occur by chance and, with randomization, should be unbiased in the long-run.

Controlling the Familywise Error Rate

We describe the importance of this study characteristic in the main text. Fifty-eight percent of the samples (23% of the total students) were from studies where either only one comparison was conducted on the sample (and therefore did not need to correct the family-wise error rate) or the authors corrected for the family-wise error rate for the sample.

Rationale for Exclusion

Not controlling for the family-wise error rate increases the chance of Type I error, but it does not impact the size of the effect entered into the meta-analysis.

Testing for Differences Between Subgroups

We describe the importance of this study characteristic in the main text. Of the studies that included a claim of subgroup differences, 81% of the samples (82% of the total students) were from studies where such a test was conducted.

Rationale for Exclusion

Researchers making claims about differences should test and confirm these differences. However, failure to test for group differences does not impact the size of the effect entered into the meta-analysis.

Reporting Confidence Intervals or Variance Statistics

Confidence intervals provide an estimated range of population parameter values that are compatible with the data. Reporting confidence intervals provides more information to readers about the range of values a replication might produce (e.g., a 95% confidence interval will include the value from a same-sized replication ~83% of the time; Cumming & Maillardet, 2006). Alternatively, providing the standard deviation, standard error of the mean, or another variance statistic allows readers to interpret the variability of scores, which is not possible when a study only reports group means.

Ninety percent of the samples (99% of the total students) were from studies where variance around the sample's group means was reported.

Rationale for Exclusion

Failure to provide clear information about the variability of group means does not impact the size of the effect entered into the meta-analysis. (Effect sizes can be calculated from test statistics when means and associated variances are not provided).

Reporting Effect Sizes for Key Results

Effect sizes—the magnitude of the effect—are the most important outcome of empirical studies, in part because they communicate the practical importance of results (Lakens, 2013). While researchers might disagree about whether a particular effect size is trivial or meaningful, reporting effect sizes provides readers the opportunity to make those judgments.

Fifty-two percent of the samples (77% of the total students) were from studies where the treatment-control difference effect size for the sample was reported.

Rationale for Exclusion

Failure to provide the effect size does not impact the size of the effect entered into the meta-analysis. (Effect sizes can be calculated from reported descriptive or test statistics).

Conducting Theory-driven, Rather than Data-driven, Analyses

When testing a hypothesis, analyses should be theory-driven confirmatory analyses. Data-driven analyses are acceptable for exploratory purposes, however, they can enable researchers to attempt to derive significant or large effects by combining or separating groups or measures. Such analytical flexibility is likely to increase the rate of false positives (Simmons et al., 2011) and should therefore be subsequently confirmed with a new dataset.

One hundred percent of samples (100% of students) were from studies where analyses designed for hypothesis testing—as opposed to (only) analyses designed to be data driven (e.g., machine learning techniques)—were reported.

Rationale for Exclusion

We cannot know if any of the analyses intended for hypothesis testing were selected in a data-driven way (i.e., *p*-hacking).

Conducting and Reporting Analyses Interpretable to Most Readers

Researchers should endeavor to communicate their findings using the most appropriate analyses to answer their research question. This includes consideration of the methodological design, complexity of the data, and interpretability of the results. If analyzing data in a way that is unknown or likely uninterpretable to most readers, efforts should be made to also provide interpretable information. Without such information, it can be unclear how analysis choices impacted the size of the effect.

For example, the primary analysis reported by Yeager et al. (2018, see also Yeager et al., 2019) to test whether the mindset intervention improved students' grades was described as a "cluster-robust fixed-effects linear regression model that used weights provided by the research firm to make coefficients generalizable" (p. 10-11). The weighting scheme was not provided. Without such information, readers cannot evaluate the appropriateness of the research firm's weighting decisions, and therefore cannot evaluate the appropriateness of the study's effect size. (Note: Employees of the research firm were also study authors.)

Four psychology professors who teach graduate statistics courses rated each unique analysis as understandable to most readers or not understandable to most readers. One rater noted several in which she believed most readers would not have in-depth understanding, but coded these differently from ones in which she believed readers would not be able to understand. We coded the former as understandable. Inter-rater reliability was good (Fleiss et al., 2003): Percent agreement = 84.31%, Fleiss-Kappa = .69, 95% CI [.45, .93]. Only one case resulted in an even split between understandable and not understandable. For all others, three or all four raters agreed. For all cases where three raters agreed, the majority rating was understandable. Three analyses had uniform agreement among raters that they were not understandable to most readers. Studies using these three analyses were rated as not understandable; all others were rated as understandable.

Ninety-two percent of samples (75% of students) were from studies where analyses for the sample were interpretable to most readers.

Rationale for Exclusion

There are multiple reasons why researchers might conduct analyses uninterpretable to most readers, including the complexity of the study design. We cannot know how these decisions impacted effect sizes.

Appropriately Interpreting Effects

Researchers should only conclude the intervention was successful if there was a significant difference between the treatment and control groups *and* the intervention successfully influenced students' mindsets (i.e., successful manipulation check).

Seventy percent of samples (44% of students) were from studies where analyses for that sample were appropriately interpreted.

Table S3

Moderator Correlation Matrix

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.Developmental Stage	----									
2. Academic Challenge Status	V = .337, *p* = .008	----								
3. Socioeconomic Status	V = .477, *p* = .071	V = .250, *p* = .380	----							
4. Intervention-Measure Interval	V = .404, *p* = .003	V = .338, *p* = .043	V = .362, *p* = .194	----						
5. Intervention Type	V = .281, *p* = .056	V = .194, *p* = .215	X	V = .325, *p* = .055	----					
6. Number of Sessions	Θ = .467	Θ = .270	r_{pb} = -.412, *p* = .026	Θ = .308	Θ = .134					
7. Intervention Delivery Mode	V = .275, *p* = .039	V = .315, *p* = .018	V = .230, *p* = .649	V = .373, *p* = .003	V = .358, *p* = .003	Θ = .344	----			
8. Administrator	V = .355, *p* = .180	V = .366, *p* = .187	V = .056, *p* = .997	V = .256, *p* = .852	V = .145, *p* = .923	Θ = .482	V = .316, *p* = .380	----		
9. Context	V = .452, *p* = .001	V = .506, *p* < .001	V = .205, *p* = .253	V = .246, *p* = .358	V = .218, *p* = .154	r_{pb} = .424, *p* < .001	V = .378, *p* = .010	V = .616, *p* = .003	----	
10. Achievement Measure	V = .407, *p* = .010	V = .412, *p* = .024	V = .502, *p* = .166	V = .443, *p* < .001	V = .375, *p* = .075	Θ = .316	V = .300, *p* = .443	V = .487, *p* = .055	V = .575, *p* < .001	----
11. Laboratory test status	V = .151, *p* = .620	V = .331, *p* = .015	X	V = .616, *p* < .001	V = .069, *p* = .828	r_{pb} = -.111, *p* = .336	V = .106, *p* = .828	V = .214, *p* = .742	V = .282, *p* = .012	V = .544, *p* = .001

Note. V = Cramer's V (correlation between categorical variables). r_{pb} = point biserial correlation (correlation between one continuous variable and one dichotomous variable). Θ = Freeman's Theta (correlation between one continuous variable and one non-dichotomous nominal variable). We are unaware of any method available to calculate *p*-values for Freeman's Theta. Confidence intervals can be calculated, but because the statistic cannot be negative, depending on the method of calculation, a confidence interval will never include zero. X = could not compute correlation (in both cases socioeconomic status information was only available for one level of the other variable).

All Studies that Met At Least Half of the Original Eighteen Best Practices Criteria

This model is identical to Model 1: No Quality Control in the main text except it excludes the 15 samples associated with studies that met fewer than half the original eighteen best practices criteria. See Figure S1 for the studies included in this model.

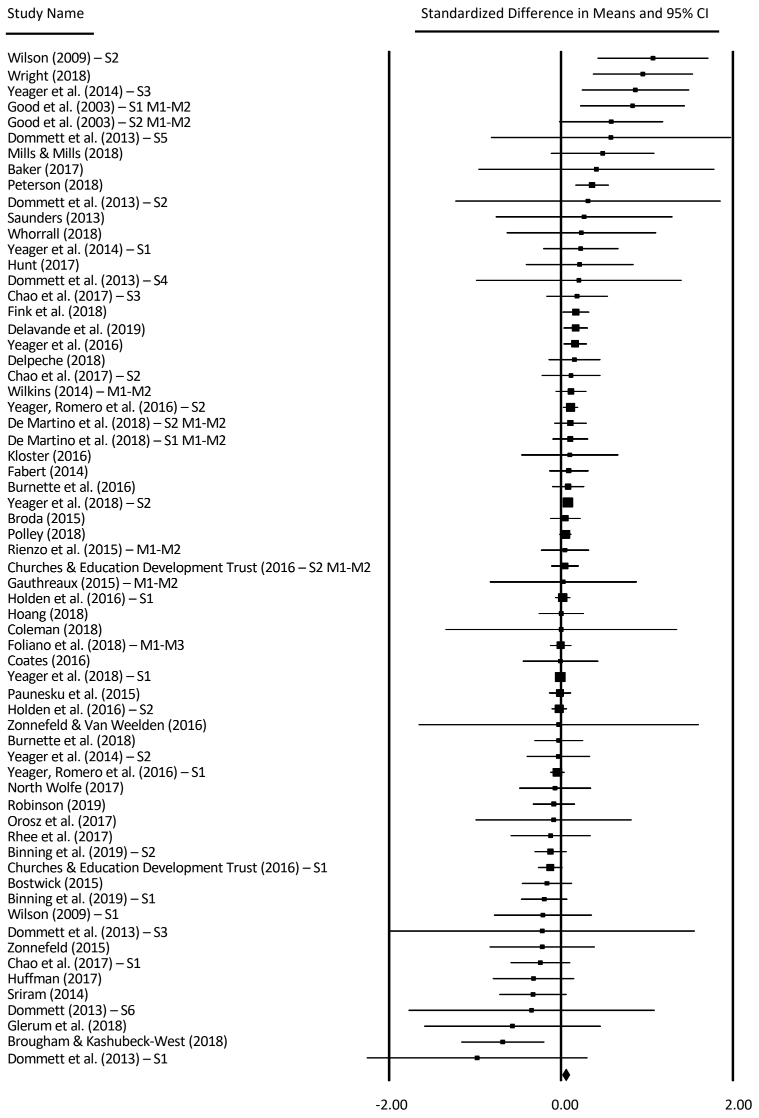
Results

Overall Results

The overall meta-analytic average standardized mean difference in academic achievement between students receiving a growth mindset intervention and students in a control group was $\bar{d}$ = 0.05, 95% CI = [0.01, 0.08], p = .020. See Figure S1.

Figure S1

Each Sample's Effect Size and 95% Confidence Interval for All Studies that Met At Least Half the Original Eighteen Best Practices Criteria



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

Moderator Analyses

The between-study variability in effect sizes due to heterogeneity rather than random error was moderate, $I^2 = 42.42$ ($\tau^2 = .006$). We investigated the source of this heterogeneity through moderator analyses. We conducted moderator analyses when there were at least five effect sizes contributing to a subgroup (Williams, 2012). The results largely follow the patterns observed in Meta-analysis 1 in the main text. See Table S4.

Table S4

Moderator Results for Model 1A: All Studies that Met At Least Half the Original Eighteen Best Practices Criteria

Moderator and Levels	Result		
		Theoretical Factors	
Developmental Stage		**$Q(2) = 0.30, p = .859$**	
Adults	$\bar{d} = 0.05$	95% CI [-0.02, 0.13]	$p = .184$
Adolescents	$\bar{d} = 0.04$	95% CI [-0.01, 0.10]	$p = .090$
Children	$\bar{d} = 0.09$	95% CI [−0.06, 0.23]	$p = .240$
Academic Challenge Status [a]		**$Q(2) = 2.63, p = .269$**	
High challenge level (e.g., low grades)	$\bar{d} = 0.09$	95% CI [0.05, 0.13]	$p < .001$
Situational challenge (e.g., new school)	$\bar{d} = 0.05$	95% CI [-0.02, 0.12]	$p = .200$
Low challenge level	$\bar{d} = 0.03$	95% CI [-0.02, 0.09]	$p = .259$
Socioeconomic status [b]		**$Q(1) = 1.49, p = .223$**	
Middle-high	$\bar{d} = 0.03$	95% CI [-0.02, 0.08]	$p = .273$
Low	$\bar{d} = 0.14$	95% CI [-0.03, 0.30]	$p = .109$
Intervention-Outcome Measure Interval [c]		**$Q(1) = 0.77, p = .379$**	
Short (interval ≤ four months)	$\bar{d} = 0.07$	95% CI [0.01, 0.13]	$p = .027$
Long (interval > four months)	$\bar{d} = 0.03$	95% CI [-0.03, 0.09]	$p = .329$
		Methodological Factors	
Intervention Type		**------------------**	
Interactive (e.g., “saying-is-believing” task)	$\bar{d} = 0.05$	95% CI [0.01, 0.10]	$p = .014$
Passive (e.g., only reading materials)	---------	----------------	------
Number of Sessions		**$Q(1) = 5.04, p = .025$**	
Slope	$b = 0.01$	95% CI [0.0007, 0.01]	$p = .025$
Intervention delivery mode		**$Q(2) = 3.83, p = .147$**	
Reading material	---------	----------------	------

Computer program	$\bar{d}$ = 0.03	95% CI [-0.004, 0.07]	*p* = .081
In person	$\bar{d}$ = 0.03	95% CI [-0.06, 0.11]	*p* = .546
Combination of delivery modes	$\bar{d}$ = 0.34	95% CI [0.03, 0.65]	*p* = .030
Administrator (of in-person delivery)		***Q*(2) = 1.44, *p* = .487**	
Teacher	$\bar{d}$ = 0.07	95% CI [−0.04, 0.19]	*p* = .196
Researcher	$\bar{d}$ = -0.01	95% CI [−0.32, 0.31]	*p* = .965
Teacher who is also the researcher	---------	----------------	------
Other	$\bar{d}$ = 0.23	95% CI [-0.05, 0.52]	*p* = .103
Context		***Q*(1) = 1.52, *p* = .217**	
In the classroom	$\bar{d}$ = 0.09	95% CI [0.01, 0.18]	*p* = .032
Outside the classroom	$\bar{d}$ = 0.03	95% CI [-0.01, 0.08]	*p* = .126
Academic Achievement Measure		***Q*(3) = 2.33, *p* = .506**	
Course exam grade	$\bar{d}$ = 0.05	95% CI [-0.06, 0.16]	*p* = .374
Single course grade	$\bar{d}$ = 0.10	95% CI [-0.001, 0.20]	*p* = .053
Multi-course grade average (e.g., GPA)	$\bar{d}$ = 0.02	95% CI [-0.03, 0.08]	*p* = .367
Standardized test score	$\bar{d}$ = 0.09	95% CI [-0.02, 0.20]	*p* = .106
Laboratory v. actual standardized test		***Q*(1) = 0.18, *p* = .669**	
Laboratory-based standardized test	$\bar{d}$ = 0.16	95% CI [-0.22, 0.54]	*p* = .413
Actual standardized test score	$\bar{d}$ = 0.07	95% CI [-0.04, 0.18]	*p* = .210

[a] Two studies provided information for high challenge-level subsamples. When replacing the whole samples with these subsamples, the pattern of results is unchanged. [b] Studies not reporting student-level socioeconomic status were not included in this moderator analysis. Four studies provided information for low socioeconomic subsamples. When replacing the whole samples with these subsamples, the pattern of results is unchanged. [c] For seven samples, a longer interval was available beyond the academic context in which the intervention was administered. When replacing effects with those from these longer intervals, the moderator and effect for long intervals remain non-significant and the effect for short intervals becomes non-significant.

Financial incentives. For studies where an author had a financial incentive to find positive effects, the effect was significant, $\bar{d}$ = 0.06, 95% CI [0.01, 0.11], *p* = .027. For studies where no authors had financial conflicts of interest, the effect of growth mindset interventions on academic achievement was not significant, $\bar{d}$ = 0.03, 95% CI [-0.03, 0.09], *p* = .323. However, whether or not one or more authors had a financial incentive to find positive effects was not a significant moderator, *Q*(1) = 0.40, *p* = .527.

Financial incentives in the published literature. Within the published literature, there was a significant difference between studies where an author had a financial incentive to find positive effects, $\bar{d}$ = 0.14, 95% CI [0.04, 0.24], p = .008, and studies where no authors had financial conflicts of interest, $\bar{d}$ = -0.12, 95% CI [-0.27, 0.03], p = .122; Q(1) = 7.66, p = .006.

All Studies that Met At Least Half of the Ten Best Practices Criteria

This model is identical to Model 1: No Quality Control in the main text except it excludes the 34 samples associated with studies that met fewer than half of the ten best practices criteria. See Figure S2 for the studies included in this model.

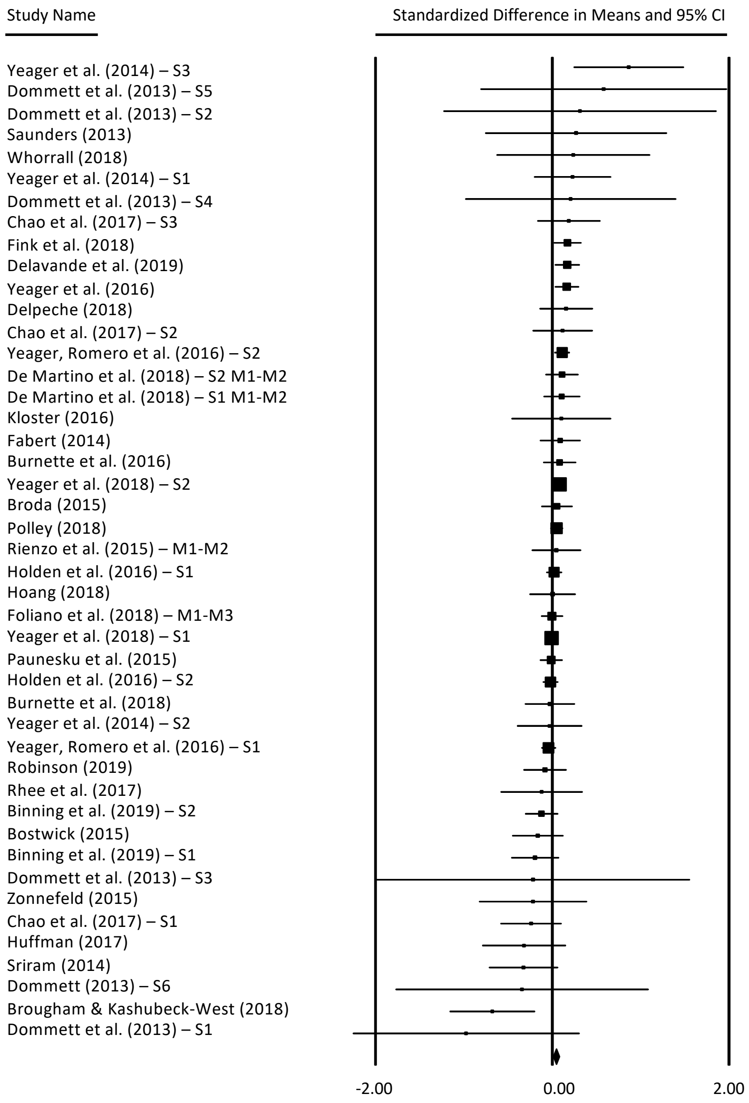
Results

Overall Results

The overall meta-analytic average standardized mean difference in academic achievement between students receiving a growth mindset intervention and students in a control group was not significant: $\bar{d}$ = 0.03, 95% CI = [-0.02, 0.07], p = .068. See Figure S2.

Figure S2

Each Sample's Effect Size and 95% Confidence Interval in Meta-analysis 1B: All Studies that Met At Least Half of the Ten Best Practices Criteria



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

Moderator Analyses

The between-study variability in effect sizes due to heterogeneity rather than random error was low, $I^2 = 28.16$ ($\tau^2 = .003$). We conducted moderator analyses when there were at least five effect sizes contributing to a subgroup (Williams, 2012). The results largely follow the patterns observed in Meta-analysis 1 in the main text. See Table S5.

Table S5
Moderator Results for Model 1B: All Studies that Met At Least Half of the Ten Best Practices Criteria

Moderator and Levels	Result		
Theoretical Factors			
Developmental Stage		**$Q(2) = 0.18$, $p = .914$**	
Adults	$\bar{d} = 0.03$	95% CI [-0.02, 0.09]	$p = .249$
Adolescents	$\bar{d} = 0.03$	95% CI [-0.02, 0.08]	$p = .190$
Children	$\bar{d} = 0.01$	95% CI [−0.10, 0.11]	$p = .866$
Academic Challenge Status [a]		**$Q(2) = 5.82$, $p = .054$**	
High challenge level (e.g., low grades)	$\bar{d} = 0.09$	95% CI [0.04, 0.13]	$p < .001$
Situational challenge (e.g., new school)	$\bar{d} = 0.02$	95% CI [-0.04, 0.08]	$p = .505$
Low challenge level	$\bar{d} = 0.02$	95% CI [-0.03, 0.06]	$p = .459$
Socioeconomic status [b]		**$Q(1) = 0.08$, $p = .775$**	
Middle-high	$\bar{d} = 0.03$	95% CI [-0.01, 0.08]	$p = .147$
Low	$\bar{d} = 0.06$	95% CI [-0.10, 0.21]	$p = .461$
Intervention-Outcome Measure Interval [c]		**$Q(1) = 0.001$, $p = .979$**	
Short (interval ≤ four months)	$\bar{d} = 0.03$	95% CI [-0.02, 0.08]	$p = .262$
Long (interval > four months)	$\bar{d} = 0.03$	95% CI [-0.03, 0.09]	$p = .329$
Methodological Factors			
Intervention Type		**------------------**	
Interactive (e.g., “saying-is-believing” task)	$\bar{d} = 0.04$	95% CI [-0.0003, 0.08]	$p = .052$
Passive (e.g., only reading materials)	---------	----------------	------
Number of Sessions		**$Q(1) = 0.26$, $p = .608$**	
Slope	$b = -0.004$	95% CI [-0.02, 0.01]	$p = .608$
Intervention delivery mode		**$Q(1) = 0.46$, $p = .498$**	
Reading material	---------	----------------	------
Computer program	$\bar{d} = 0.03$	95% CI [-0.004, 0.07]	$p = .083$
In person	$\bar{d} = -0.002$	95% CI [-0.06, 0.11]	$p = .546$

Combination of delivery modes	---------	----------------	------
Administrator (of in-person delivery)		**Q(1) = 0.55, p = .460**	
Teacher	$\bar{d}$ = 0.007	95% CI [−0.10, 0.12]	p = .906
Researcher	$\bar{d}$ = -0.12	95% CI [−0.44, 0.20]	p = .458
Teacher who is also the researcher	---------	----------------	------
Other	---------	----------------	------
Context		**Q(1) = 0.11, p = .736**	
In the classroom	$\bar{d}$ = 0.02	95% CI [-0.04, 0.08]	p = .508
Outside the classroom	$\bar{d}$ = 0.03	95% CI [-0.007, 0.07]	p = .109
Academic Achievement Measure		**Q(2) = 0.21, p = .898**	
Course exam grade	---------	----------------	------
Single course grade	$\bar{d}$ = 0.03	95% CI [-0.04, 0.10]	p = .348
Multi-course grade average (e.g., GPA)	$\bar{d}$ = 0.02	95% CI [-0.03, 0.08]	p = .380
Standardized test score	$\bar{d}$ = 0.01	95% CI [-0.08, 0.10]	p = .876
Laboratory v. actual standardized test		**Q(1) = 0.53, p = .467**	
Laboratory-based standardized test	$\bar{d}$ = 0.10	95% CI [-0.17, 0.38]	p = .461
Actual standardized test score	$\bar{d}$ = -0.004	95% CI [-0.10, 0.09]	p = .930

[a] Two studies provided information for high challenge-level subsamples. When replacing the whole samples with these subsamples, the moderator reaches significance, Q(2) = 6.52, p = .038. The patterns of significance at each level of the moderator are unchanged. [b] Studies not reporting student-level socioeconomic status were not included in this moderator analysis. Two studies provided information for low socioeconomic subsamples. When replacing the whole samples with these subsamples, the pattern of results is unchanged. [c] For seven samples, a longer interval was available beyond the academic context in which the intervention was administered. When replacing effects with those from these longer intervals, the pattern of results is unchanged.

Financial incentives. For studies where an author had a financial incentive to find positive effects, the effect was marginally significant, $\bar{d}$ = 0.05, 95% CI [0.001, 0.09], p = .045. For studies where no authors had financial conflicts of interest, the effect of growth mindset interventions on academic achievement was not significant, $\bar{d}$ = 0.01, 95% CI [-0.04, 0.06], p = .699. However, whether or not one or more authors had a financial incentive to find positive effects was not a significant moderator, Q(1) = 1.03, p = .310.

Financial incentives in the published literature. Within the published literature, there was a significant difference between studies where an author had a financial incentive to find positive effects, $\bar{d}$ = 0.11, 95% CI [0.02, 0.20], p = .017, and studies where no authors had financial conflicts of interest, $\bar{d}$ = -0.14, 95% CI [-0.28, -0.0006], p = .049; Q(1) = 8.65, p = .003.

Minimal Standard of Evidence that Met At Least Half of the Best Practices Criteria

This model is identical to Model 2: Minimal Standard of Evidence in the main text except it further excludes three samples associated with studies that met fewer than half of the best practices criteria. The same three samples are excluded when applying the 50% threshold to the original eighteen best practices criteria and when applying the 50% threshold to the final ten best practices criteria, yielding identical models. See Figure S3 for studies included in this model.

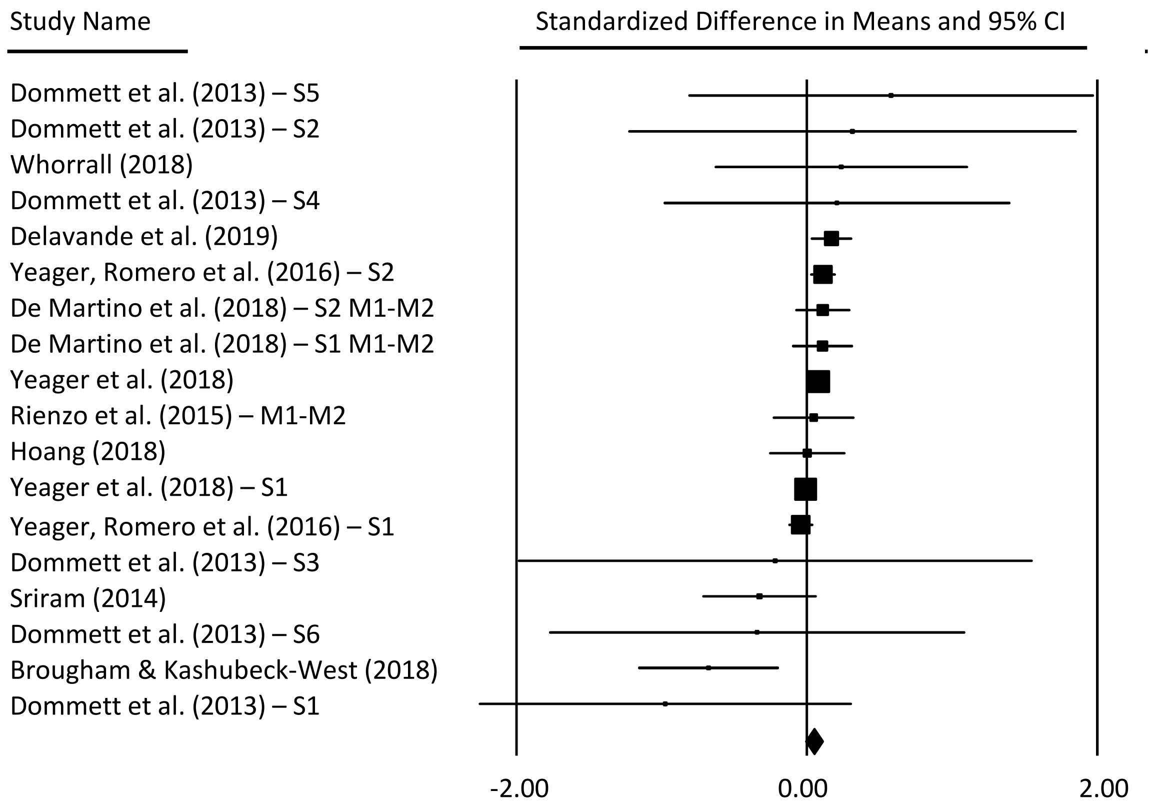
Results

Overall Results

The overall meta-analytic average standardized mean difference between treatment and control groups in academic achievement was not significant, $\bar{d}$ = 0.04, 95% CI = [-0.02, 0.10], p = .192. See Figure S3.

Figure S3

Each Sample's Effect Size and 95% Confidence Interval for Studies that Met the Minimal Standard of Evidence and At Least Half the Best Practices Criteria



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

Moderator analyses. The between-study variability in effect sizes due to heterogeneity rather than random error was moderate, $I^2 = 44.27$ ($\tau^2 = .004$). We investigated the source of this heterogeneity through moderator analyses. We conducted moderator analyses when there were at least five effect sizes contributing to a subgroup (Williams, 2012). The relatively small number of studies that met the criteria for this meta-analysis limited the moderator analyses we could conduct. See Table S6.

Table S6

Model 2A: Minimal Standard of Evidence that Met At Least Half of the Best Practices Criteria Moderator Results

Moderator and Levels	Result		
Theoretical Factors			
Developmental Stage		**------------------**	
Adults	---------	----------------	------
Adolescents	$\bar{d}$ = 0.03	95% CI [-0.03, 0.10]	*p* = .284
Children	---------	----------------	------
Academic Challenge Status [a]		***Q*(1) = 2.99, *p* = .084**	
High challenge level (e.g., low grades)	$\bar{d}$ = 0.08	95% CI [0.04, 0.13]	*p* < .001
Situational challenge (e.g., new school)	$\bar{d}$ = -0.01	95% CI [−0.10, 0.09]	*p* = .909
Low challenge level	---------	----------------	------
Socioeconomic status [b]		***Q*(1) = 0.63, *p* = .427**	
Middle-high	$\bar{d}$ = 0.05	95% CI [-0.01, 0.11]	*p* = .119
Low	$\bar{d}$ = 0.11	95% CI [-0.03, 0.25]	*p* = .111
Intervention-Outcome Measure Interval [c]		***Q*(1) = 0.09, *p* = .759**	
Short (interval ≤ four months)	$\bar{d}$ = 0.02	95% CI [-0.11, 0.15]	*p* = .779
Long (interval > four months)	$\bar{d}$ = 0.04	95% CI [-0.02, 0.10]	*p* = .205
Methodological Factors			
Intervention Type		**------------------**	
Interactive (e.g., “saying-is-believing” task)	$\bar{d}$ = 0.04	95% CI [-0.02, 0.10]	*p* = .192
Passive (e.g., only reading materials)	---------	----------------	------
Number of Sessions		***Q*(1) = 0.03, *p* = .858**	
Slope	*b* = -0.003	95% CI [-0.03, 0.03]	*p* = .858
Intervention delivery mode		***Q*(1) = 0.71, *p* = .400**	
Reading material	---------	----------------	------
Computer program	$\bar{d}$ = 0.05	95% CI [-0.01, 0.10]	*p* = .100
In person	$\bar{d}$ = -0.15	95% CI [-0.60, 0.30]	*p* = .519
Combination of delivery modes	---------	----------------	------
Administrator (of in-person delivery)		**------------------**	
Teacher	---------	----------------	------
Researcher	---------	----------------	------
Teacher who is also the researcher	---------	----------------	------

Other	---------	----------------	------
Context		**------------------**	
In the classroom	---------	----------------	------
Outside the classroom	$\bar{d}$ = 0.04	95% CI [-0.02, 0.10]	p = .192
Academic Achievement Measure		**Q(1) = 0.43, p = .513**	
Course exam grade	---------	----------------	------
Single course grade	---------	----------------	------
Multi-course grade average (e.g., GPA)	$\bar{d}$ = 0.03	95% CI [-0.05, 0.10]	p = .492
Standardized test score	$\bar{d}$ = -0.03	95% CI [-0.20, 0.13]	p = .689
Laboratory v. actual standardized test		**------------------**	
Laboratory-based standardized test	$\bar{d}$ = -0.08	95% CI [-0.66, 0.49]	p = .780
Actual standardized test score	---------	----------------	------

------- = not enough studies available to include in analysis. [a] One study was available that provided information for a high-risk subsample. The pattern of results does not change when replacing the whole sample with this subsample. [b] Studies not reporting student-level socioeconomic status were not included in this moderator analysis. Not enough low-SES samples were available for moderation analysis, unless we replaced whole samples with available low-SES subsamples. The results in the table reflect results with sub-samples replacements. [c] For six samples, a longer interval was available beyond the academic context in which the intervention was administered. When replacing effects with those from longer intervals, the pattern of results is unchanged.

Financial incentives. With the increased quality control of this model—studies that met at least half the best practices criteria and provided evidence the intervention influenced students' mindsets—not enough studies by authors with perceived financial conflicts of interest remained to conduct this moderator analysis. The average effect on academic achievement from authors without financial incentives was non-significant, $\bar{d}$ = 0.02, 95% CI [-0.11, 0.15], p = .778.

Combinations of Best Practices Criteria Adherence

Meta-analysis 3 in the main text included the studies that met the most best practices criteria regardless of which best practices they met. However, not all best practices criteria are equally crucial, and researchers might disagree on which combination of the best practices criteria should be considered in a model examining the highest quality studies. Additionally, if measures of mindset are not valid measures of the mindset construct, then we should not require that all studies meet our criterion that the treatment measurably impacted students' mindsets to be included in a model of the best available evidence.

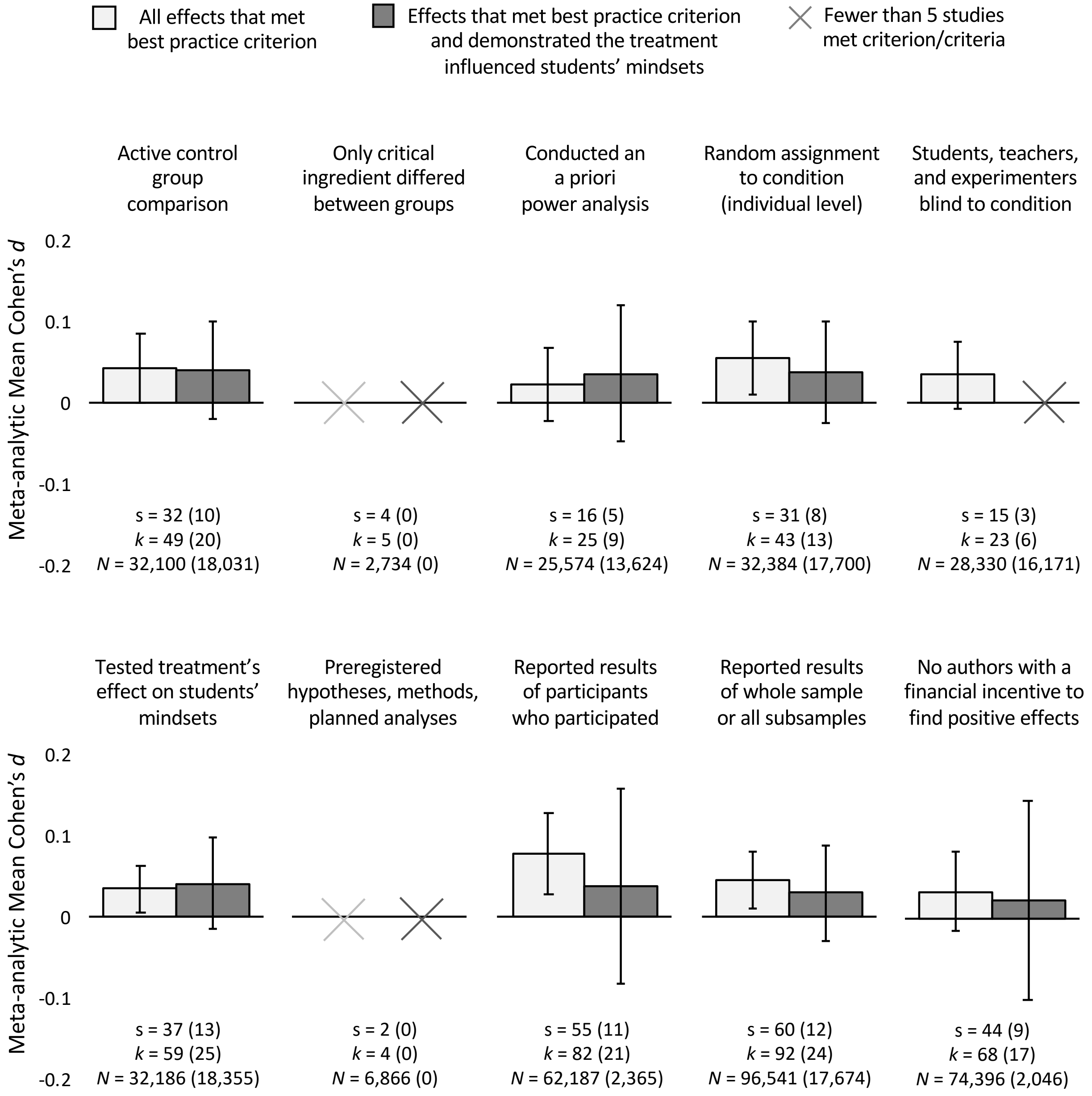
Here, we present models of all possible combinations of the ten best practices criteria when at least five studies meet the criteria. We report results with and without the requirement that the studies demonstrated the intervention influenced students' mindsets. We invite readers to 1) decide which combination of best practices criteria, see Table S7, they consider critical for evaluating the evidence that instilling a growth mindset increases students' academic achievement, then 2) move to the subsection below that matches the number of criteria met in the preferred combination, and 3) examine the meta-analytic results that applies the preferred combination of criteria. If selecting more than one combination as evidence for a treatment effect, readers should adjust the alpha for multiple comparisons. We present the results in order of increasing number of best practices criteria met.

Meta-analyses of Studies that Met at Least One Best Practice Criterion by Criterion

We presume that most researchers would agree that studies should meet more than one best practice criterion to be considered the highest-quality evidence. Nonetheless, for each criterion, we present the results for all studies that met that criterion, regardless of any other criteria met, with and without the requirement that the intervention demonstrate changes in students' mindsets. See Figure S4.

Figure S4

Results of Studies That Met Each Best Practice Criterion



Note. s = number of studies, *k* = number of effects, *N* = total sample size. Numbers in parentheses are for studies that demonstrated the treatment influenced students' mindsets. Error bars represent 95% confidence intervals.

There were not enough studies that isolated the critical ingredient of teaching attribute malleability to include in any models. Likewise, there were not enough preregistered studies to include in any models.

We present the results using a liberal alpha of .05—that is, not adjusting for multiple comparisons. Five models, all lacking the criterion that the treatment measurably influence students' mindsets, were significant: studies using an active control group: $\bar{d}$ = 0.04, 95% CI = [0.001, 0.09], *p* = .046; studies that randomly assigned students to condition, $\bar{d}$ = 0.06, 95% CI = [0.01, 0.10], *p* = .016; studies that tested (regardless of outcome) whether the treatment influenced students' mindsets, $\bar{d}$ = 0.04, 95% CI = [0.01, 0.06], *p* = .016; studies that reported the results of participants who participated in the study, $\bar{d}$ = 0.08, 95% CI = [0.03, 0.13], *p* = .002; and studies that reported the results of the whole sample/all subsamples, $\bar{d}$ = 0.05, 95% CI = [0.01, 0.08], *p* = .009.

However, in each of the significant models, publication bias analyses suggested that publication bias was affecting these models. When these models were adjusted for the inflation due to publication bias, all were non-significant. Model non-significance was consistent regardless of the method used to adjust for inflation (Duval and Tweedie's Trim and Fill or conditional PET-PEESE).

Meta-analyses of Studies that Met at Least Two Best Practices Criteria by Combination

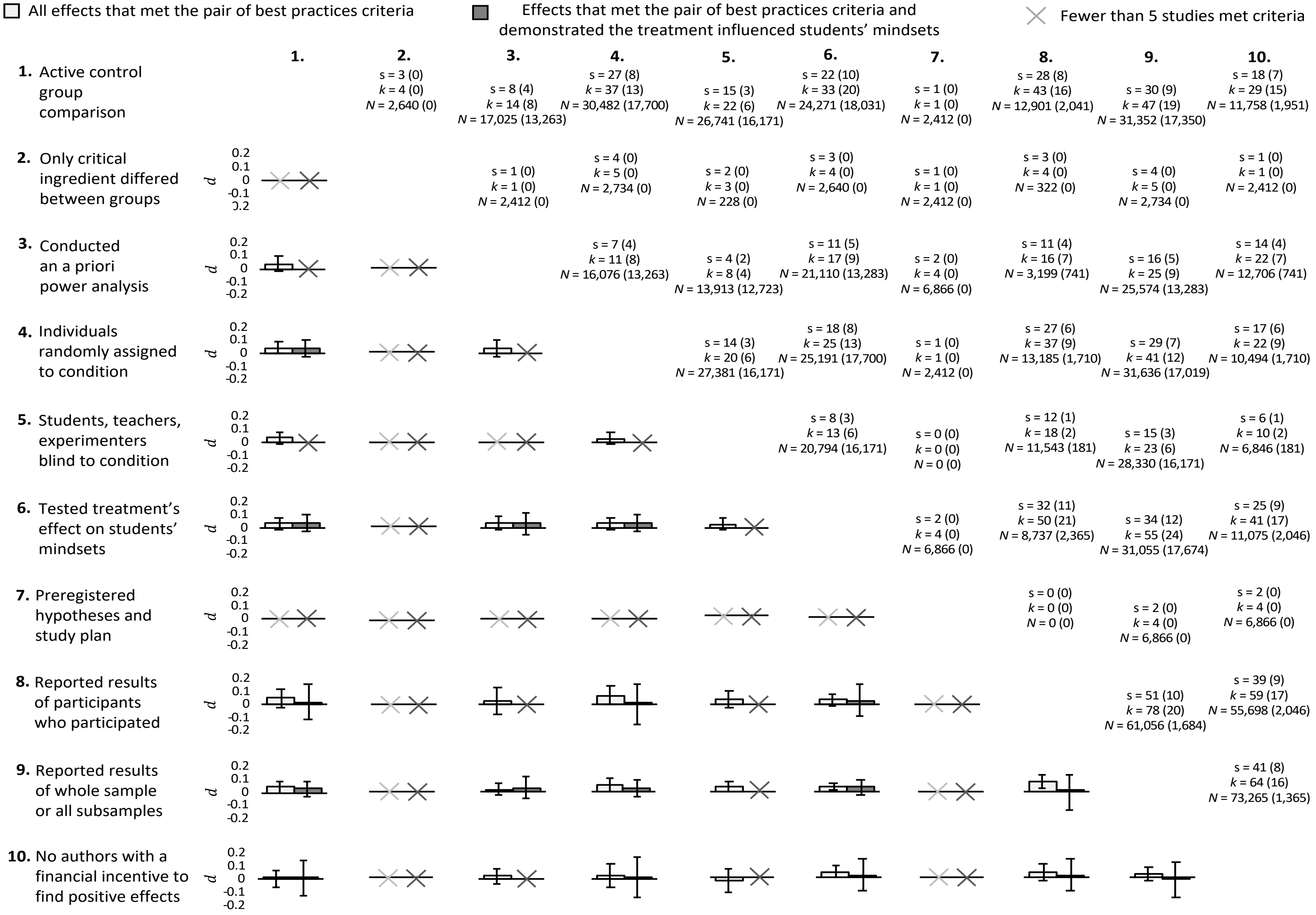
We next conducted meta-analyses of studies that met each combination of two best practices criteria when at least five studies were available. Of the 45 possible pairs of best practice criteria, 18 had fewer than five available studies that met both best practice criteria. These included any model requiring that there were no other differences between treatment and control except for teaching attribute malleability and any model requiring preregistration. Using a liberal alpha of .05—that is, not adjusting for multiple comparisons—four of the remaining 27 models were significant: studies that randomly assigned student to condition and reported the results of those who participated, $\bar{d}$ = 0.07, 95% CI = [0.001, 0.14], *p* = .046; studies that randomly assigned student to condition and reported the results of the whole sample/all subsamples, $\bar{d}$ = 0.05, 95% CI = [0.01, 0.10], *p* = .022; studies that tested (regardless of outcome) whether the treatment influenced students' mindsets and reported results of the whole sample/all subsamples, $\bar{d}$ = 0.03, 95% CI = [0.005, 0.06], *p* = .021; and studies that reported the results of participants who participated in the study and reported results of the whole sample/all subsamples, $\bar{d}$ = 0.08, 95% CI = [0.02, 0.13], *p* = .004. See Figure S5.

However, in each of the significant models, publication bias analyses suggested that publication bias was affecting these models. When these models were adjusted for the inflation due to publication bias, all were non-significant. Model non-significance was consistent regardless of the method used to adjust for inflation (Duval and Tweedie's Trim and Fill or conditional PET-PEESE).

When restricting these models to studies that demonstrated the treatment influenced students' mindsets, 26 of the 45 possible pairs had fewer than five available studies that met both criteria. Of the remaining nineteen pairs, none of the meta-analyses yielded significant effects. See Figure S5.

Figure S5

Results of Studies that Met Each Combination of Two Best Practices Criteria



	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Active control group comparison		s = 3 (0) *k* = 4 (0) *N* = 2,640 (0)	s = 8 (4) *k* = 14 (8) *N* = 17,025 (13,263)	s = 27 (8) *k* = 37 (13) *N* = 30,482 (17,700)	s = 15 (3) *k* = 22 (6) *N* = 26,741 (16,171)	s = 22 (10) *k* = 33 (20) *N* = 24,271 (18,031)	s = 1 (0) *k* = 1 (0) *N* = 2,412 (0)	s = 28 (8) *k* = 43 (16) *N* = 12,901 (2,041)	s = 30 (9) *k* = 47 (19) *N* = 31,352 (17,350)	s = 18 (7) *k* = 29 (15) *N* = 11,758 (1,951)
2. Only critical ingredient differed between groups			s = 1 (0) *k* = 1 (0) *N* = 2,412 (0)	s = 4 (0) *k* = 5 (0) *N* = 2,734 (0)	s = 2 (0) *k* = 3 (0) *N* = 228 (0)	s = 3 (0) *k* = 4 (0) *N* = 2,640 (0)	s = 1 (0) *k* = 1 (0) *N* = 2,412 (0)	s = 3 (0) *k* = 4 (0) *N* = 322 (0)	s = 4 (0) *k* = 5 (0) *N* = 2,734 (0)	s = 1 (0) *k* = 1 (0) *N* = 2,412 (0)
3. Conducted an a priori power analysis				s = 7 (4) *k* = 11 (8) *N* = 16,076 (13,263)	s = 4 (2) *k* = 8 (4) *N* = 13,913 (12,723)	s = 11 (5) *k* = 17 (9) *N* = 21,110 (13,283)	s = 2 (0) *k* = 4 (0) *N* = 6,866 (0)	s = 11 (4) *k* = 16 (7) *N* = 3,199 (741)	s = 16 (5) *k* = 25 (9) *N* = 25,574 (13,283)	s = 14 (4) *k* = 22 (7) *N* = 12,706 (741)
4. Individuals randomly assigned to condition					s = 14 (3) *k* = 20 (6) *N* = 27,381 (16,171)	s = 18 (8) *k* = 25 (13) *N* = 25,191 (17,700)	s = 1 (0) *k* = 1 (0) *N* = 2,412 (0)	s = 27 (6) *k* = 37 (9) *N* = 13,185 (1,710)	s = 29 (7) *k* = 41 (12) *N* = 31,636 (17,019)	s = 17 (6) *k* = 22 (9) *N* = 10,494 (1,710)
5. Students, teachers, experimenters blind to condition						s = 8 (3) *k* = 13 (6) *N* = 20,794 (16,171)	s = 0 (0) *k* = 0 (0) *N* = 0 (0)	s = 12 (1) *k* = 18 (2) *N* = 11,543 (181)	s = 15 (3) *k* = 23 (6) *N* = 28,330 (16,171)	s = 6 (1) *k* = 10 (2) *N* = 6,846 (181)
6. Tested treatment's effect on students' mindsets							s = 2 (0) *k* = 4 (0) *N* = 6,866 (0)	s = 32 (11) *k* = 50 (21) *N* = 8,737 (2,365)	s = 34 (12) *k* = 55 (24) *N* = 31,055 (17,674)	s = 25 (9) *k* = 41 (17) *N* = 11,075 (2,046)
7. Preregistered hypotheses and study plan								s = 0 (0) *k* = 0 (0) *N* = 0 (0)	s = 2 (0) *k* = 4 (0) *N* = 6,866 (0)	s = 2 (0) *k* = 4 (0) *N* = 6,866 (0)
8. Reported results of participants who participated									s = 51 (10) *k* = 78 (20) *N* = 61,056 (1,684)	s = 39 (9) *k* = 59 (17) *N* = 55,698 (2,046)
9. Reported results of whole sample or all subsamples										s = 41 (8) *k* = 64 (16) *N* = 73,265 (1,365)
10. No authors with a financial incentive to find positive effects										

Note. s = number of studies, *k* = number of effects, *N* = total sample size. Numbers in parentheses are for studies that demonstrated the treatment influenced students' mindsets. Error bars represent 95% confidence intervals.

Meta-analyses of Studies that Met Three or More Best Practices Criteria by Combination

We conducted a meta-analysis for every combination of three best practices criteria met, four best practices met, five best practices met and so on when at least five studies were available. Using a liberal alpha of .05—that is, not adjusting for multiple comparisons—there were no significant models that adhered to any combination of three or more best practices criteria.

References

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

Cumming G., & Maillardet, R. (2006). Confidence intervals and replication: Where will the next mean fall? *Psychological Methods, 11*, 217–227. doi:10.1037/1082-989X.11.3.217 pmid:16953701

Fleiss, J. L., Levin, B., Cho Paik, M. (2003). *Statistical methods for rates and proportions* (3rd ed.). New York: John Wiley and Sons.

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

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

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 72*, 304–305.

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

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

https://doi.org/10.1177/0956797611417632

Stanley, T. D., & Doucouliagos, H. (2014). Meta‐regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*, 60-78.

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

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

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., 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., Man Yang, S., 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