Received: 29 April 2024 | Accepted: 17 April 2025

DOI: 10.1002/rev3.70066

RESEARCH REPORT

Review of Education BERA

Can growth mindset interventions improve academic achievement? A structured review of the existing evidence

Carolina Gazmuri

Durham University Centre for Evidence in Education, Durham University, Durham, UK

Correspondence
Carolina Gazmuri, Durham University Centre for Evidence in Education, Durham University, Durham, UK.
Email: carolina.gazmuri@durham.ac.uk

Abstract

Encouraging the idea of a growth mindset in which students believe that they can improve their ability, as opposed to a fixed mindset, has been suggested as an effective and relatively cheap approach to improving student attainment at school. This paper offers a comprehensive review of the evidence from growth mindset interventions. After a rigorous search, screening and evaluation, the inclusion criteria led to 24 studies. All were randomised control trials (RCTs) focused on growth mindset of intelligence interventions for school-age children and included output measures for academic performance assessment. Their findings reveal that the strongest studies, characterised by larger sample sizes, minimal missing data and high data quality, exhibit null or very small effect sizes, ranging from Cohen's $d = -0.01$ to $+0.065$. Additionally, certain findings raise concerns about a potential conflict-of-interest bias, suggesting that some negative or null results may remain unpublished. The review identifies four evaluations with a high degree of trustworthiness and non-conflict of interest. Among these, two studies indicate no discernible impact, while the other two show a very small impact. Given these findings, we found evidence that suggest that growth mindset interventions targeted for school-age students, do not have much or any relevant impact in academic achievement. It is therefore not advisable for schools, school districts or governments to allocate significant time or resources to the implementation of growth

mindset interventions for school-age students, as the anticipated outcomes are likely to be either null or very modest. However, if there is an opportunity to implement such interventions at a minimal or negligible cost, or as part of another objective, it might be reasonable to proceed with them, considering the potential for a small positive impact.

KEYWORDS

best evidence, educational interventions, mindset, randomised controlled trials

Context and implications

Rationale for this study and why the new findings matterThe concept of Growth Mindset has gained widespread popularity over the past decade, prompting numerous organisations to offer Growth Mindset interventions and resources. Consequently, schools, districts and governments are allocating substantial financial resources to this area. However, the true impact of these interventions remains a contentious issue among scholars. The findings of this study suggest that more reliable research does not demonstrate a meaningful effect of these interventions on students' academic outcomes.

Implications for schools, school districts and governmentsSchools, school districts and governments should be cautious about dedicating substantial time or resources to the implementation of growth mindset interventions for students, as the anticipated results are likely to be limited or marginal. Nevertheless, if there are chances to execute these interventions at a minimal or negligible expense, it could be worthwhile to proceed, considering the possibility of achieving a slight positive effect.

INTRODUCTION

Over the past century, substantial efforts have been made worldwide to enhance students' academic achievement. The more specific purpose of these efforts is to enhance the impact of investment in school education systems making school attendance worthwhile, contributing to life-long benefits for students and societies in general. Until the 1990s, the primary focus of most research was on the relationship between cognitive skills, student performance and future outcomes. However, since the 2000s, there has been a growing body of research that highlights the influence of non-cognitive skills on students' academic outcomes and long-term prospects.

Some economists have suggested that non-cognitive skills are as relevant as cognitive skills in explaining academic success, personal achievements and future earnings (Elias & Haynes, 2008; Heckman et al., 2006; Heckman & Rubinstein, 2001). What amplifies the relevance of non-cognitive skills in the quest to improve student learning is their apparently greater malleability compared to cognitive skills. While cognitive abilities appear to stabilise after the age of 10, non-cognitive skills appear to remain malleable into adolescence and adulthood (Almlund et al., 2011; Kautz et al., 2014). Moreover, programmes designed for

school-age students with the goal of enhancing socioemotional skills have demonstrated positive impacts on both non-cognitive skills and academic achievement (Durlak et al., 2010, 2011). The malleability of these skills and their potential to influence academic performance make socioemotional interventions particularly interesting for public policy. In response to this potential, there has been a surge in interventions in recent years aimed at enhancing students' socioemotional skills.

As part of this, the concept of Growth Mindset has gained widespread popularity. Coined by psychologist Carol Dweck, this concept has led to many organisations and experts offering Growth Mindset interventions, talks, training, books and materials (Barnett & Macnamara, 2023). And the popularity of these programmes has created academic debate and evaluations led by many stakeholders.

Dweck (2000) outlines two theories concerning students' beliefs about their 'intelligence'. Some students adhere to an entity theory, wherein they perceive their intelligence as fixed, and they strive to prove their innate intelligence. Others embrace an incremental theory, viewing their intelligence as malleable and capable of improvement, motivating them to focus on enhancing their intelligence. Extensive literature indicates a significant association between the latter 'growth mindset' and academic achievement (Blackwell et al., 2007; Paunesku et al., 2015; Yeager et al., 2019). As a result, numerous researchers have explored the possibility of modifying students' mindsets through interventions to enhance their academic performance.

Dweck (2000) extends the concept of people beliefs about intelligence to encompass beliefs about other attributes, such as personality and social traits. However, for the purposes of this study, the focus will be solely on interventions aimed at altering students' beliefs regarding the malleability of intelligence, not other attributes.

Much of the literature available on the effect of growth mindset interventions in academic achievement is included in two meta-analyses published in 2022. The first one, by Macnamara and Burgoyne (2022), concluded that 'the apparent effects of growth mindset interventions on academic achievement are likely attributable to inadequate study designs, reporting flows and bias'. The other meta-analysis, by Burnette et al. (2022) reported a positive effect on academic outcomes. The contradiction between the results of these two studies, despite their use of similar data but different meta-analytic methodologies, raises some doubts about the suitability of meta-analytic approaches in addressing questions related to impact and causality, where they combine studies with varying levels of rigour.

The meta-analyses by Macnamara and Burgoyne (2022) incorporate the studies previously included in Sisk et al. (2018), and additionally all the available records up until August 2019. Macnamara and Burgoyne (2022) conducted three meta-analyses focusing exclusively on growth mindset treatments aimed at enhancing academic performance. However, the three differ in the requirement on the quality for the studies. The first meta-analysis included 63 studies ($N = 97{,}672$) and reveals a small effect size, $d = 0.05$. The second meta-analysis included only the studies demonstrating the influence of interventions on students' mindset (13 studies, $N = 18{,}355$). With this refined selection, they calculate an effect size of $d = 0.04$. Finally, they include only six studies (eight samples) with high-quality evidence ($N = 13{,}571$) and find an effect size of $d = 0.02$. The study concludes that the apparent effects of growth mindset on academic performance are not substantial and are likely attributable to issues with study designs, result reporting, or biases.

Macnamara and Burgoyne's (2022) third meta-analysis aimed to be highly selective in the quality of included studies, adhering to best practices in intervention design. However, because of the extensive number of best practice criteria, no study fully met all the requirements. As a result, the authors accepted studies that met at least 60% of these criteria, without prioritising which practices were most critical to the reliability of the study. Consequently, some of the six studies included in their most rigorous meta-analysis pose significant threats

to trustworthiness. For instance, Hoang (2018) and De Martino et al. (2019) exhibit attrition rates close to 60% without providing a clear analysis to account for the potential bias introduced by such high attrition rates.

The Burnette et al. (2022) meta-analysis includes articles reporting an impact of growth mindset intervention, using a randomised design, for studies published between 2002 and the end of 2020. The focus of the meta-analysis are two key moderators: the subsamples expected to benefit most and implementation fidelity. The authors included 32 studies (48 samples) that report an effect on academic performance. This includes 51,676 students. They find an effect of $d = 0.09$. When they isolate the effects for subgroups and of high implementation fidelity interventions, they find a bigger effect of $d = 0.14$.

Several criticisms of Macnamara and Burgoyne's (2022) meta-analysis are made by Tipton et al. (2022), who propose an alternative meta-analytic methodology. Some of the criticisms are associated with how the studies' quality was assessed, how the methodology incorporated the heterogeneity of effects among different groups, and how publication bias was addressed. Tipton et al. (2022) re-analysed the data from Macnamara and Burgoyne (2022) using Burnette et al.'s (2022) method. In their analysis, incorporating all Macnamara and Burgoyne's (2022) studies, they found a mean growth mindset effect of 0.09 SD. Among at-risk students, they observed an effect size of 0.15 SD. These results align better with the findings reported by Burnette et al. (2022).

Macnamara and Burgoyne (2023) respond to Tipton et al.'s (2022) critique by noting that Tipton et al.'s re-analysis not only adopts the methodology used by Burnette et al. (2022), but also introduces additional modifications to the dataset. These include altering effect sizes without accounting for baseline differences, changes in the coding of moderators, and the redefinition of 'at-risk' samples, all without clearly justifying these decisions. Macnamara and Burgoyne (2023) then re-analyse their data using Burnette et al.'s (2022) methodology, obtaining very similar results across all three of their meta-analyses from 2022.

The divergent results observed in these two meta-analyses raise an important question, especially considering the substantial resources allocated by school systems to such interventions: Is there a genuine causal relationship between mindset and academic performance?

Consequently, there is a reasonable basis for undertaking a new analysis to understand the potential impact of growth mindset interventions on academic achievement. The approach adopted in this new study involves a review methodology that emphasises the trustworthiness of findings across a spectrum of studies, with a specific focus on the primary issues that could potentially undermine their robustness. A conventional meta-analysis is not possible due to the problems that emerge from combining studies with varying outcomes, distinct student demographics and different (albeit similar) intervention strategies. Instead, the method detailed in the following section enables us to draw well-informed conclusions regarding the influence of growth mindset interventions while duly considering the quality of each individual study, without amalgamating them into a single effect size.

This paper presents a structured review of evaluation studies on growth mindset programmes aimed at influencing students' beliefs and perceptions about their academic potential, resulting in positive changes in both academic attitudes and performance. The paper explores the following research question:

> Is there a genuine causal relationship between mindset and academic performance?

The paper presents details of the review process, quality assessment criteria and a synthesis of findings. It also discusses the nature of biases in the existing studies and justification for not extending the review to a meta-analysis. The implications for research, policy and practice are also discussed.

METHODS

The methods described are for a structured review of existing evidence.

Search and inclusion criteria

The review included all studies that met the inclusion criteria (below) and were featured in the two most recent meta-analyses conducted by Macnamara and Burgoyne (2022) and Burnette et al. (2022). Additionally, an updated search was conducted to incorporate studies that became available after the cessation of the two meta-analyses' data collection, extending up to the end of 2023. The keywords used in the search are: ("growth mindset" OR "lay theory" OR "implicit theory") AND (intervention OR trial) AND (student*). The databases searched are ERIC, APA PsycINFO, Google Scholar and Science Direct.

To be considered for inclusion, studies had to meet the following criteria:

1. The study should evaluate a growth mindset intervention aimed at school-age children. The review search and selection focused exclusively on interventions related to incremental theories of intelligence, excluding those centred on incremental theories of personalities.
2. It must have employed a randomised controlled trial (RCT) design to assess the intervention's efficacy.
3. The study should have included an outcome from a measure of academic performance.
4. The research had to be published in the English language.
5. The study should provide the essential information required for calculating Cohen's *d* effect size for the entire randomised sample.

The number of studies found at each stage are shown in Figure 1.

Assessment of evidence

The studies meeting the inclusion criteria underwent an evaluation process to judge the quality of their findings. The framework used is the 'sieve' approach, as proposed by Gorard (2021), to assess the trustworthiness of each study's outcomes. The 'sieve' method has been used effectively by research review teams (e.g. Owen et al., 2022; Siddiqui & Ventista, 2018), and is the basis for the security ratings used by the EEF in England. The aim was to ensure that conclusions were influenced most by the most robust results concerning the potential impact of growth mindset interventions on student achievement. The 'sieve' method appropriately operationally defines quality and provides a consistent scale to rate the individual studies.

The 'sieve' approach outlines four key criteria for evidence assessment, as shown in Table 1. First, it examines whether the research design is suitable for the claims being made. This study is concerned with casual claims. Given that the review selection criteria required the use of RCTs, all the studies included in this review were awarded the highest rating (4) for this criterion (Table 1). This is a suitable design for a 'what works' question. The second criterion concerns the size of the randomised groups, the third addresses attrition levels, and the fourth criterion considers the data quality. Assessing studies based on these four criteria, their quality was rated on a scale from 0 to 4. A rating of 0 indicated the lowest level of quality, adding nothing to knowledge, whereas a rating of 4 signified the highest level,

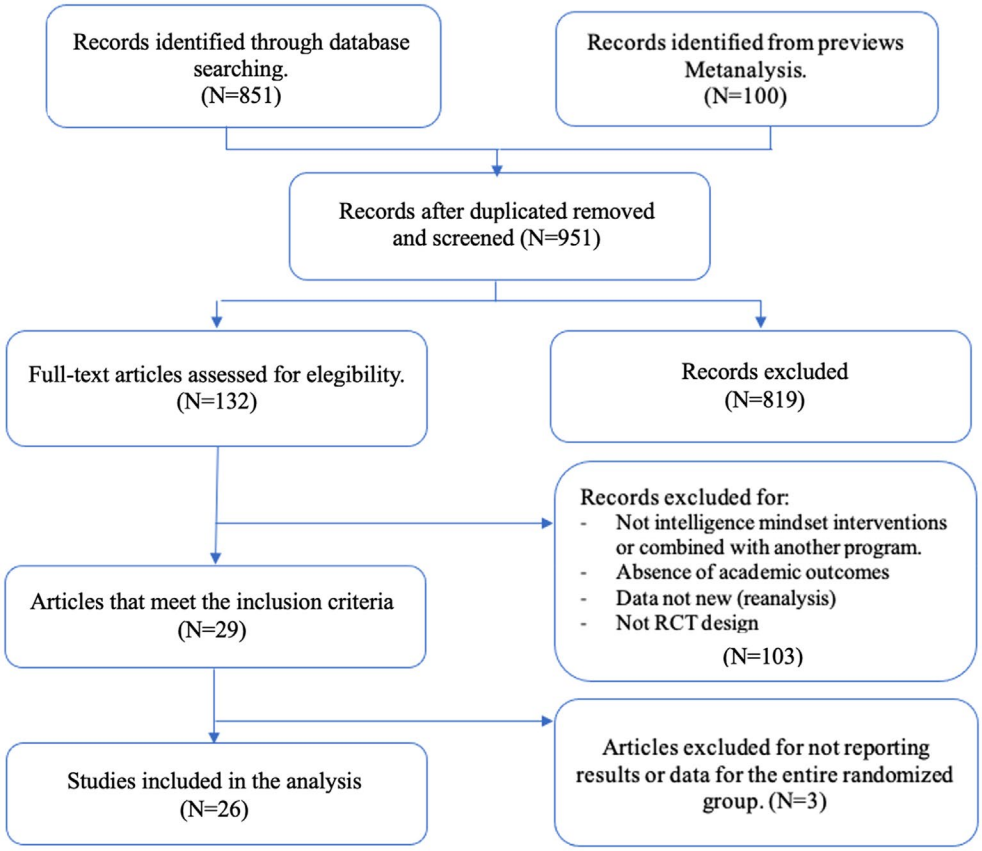

FIGURE 1 PRISMA flowchart of records discovered.

reserved for the most robust and reliable evidence. It is crucial to emphasise that ratings do not reflect the quality of the studies themselves, but rather the quality of the evidence they contribute toward addressing the research question concerning the impact of growth mindset interventions on student achievement.

A 4* rating was assigned to studies employing an RCT design with a sample size exceeding 400 students or 100 classrooms (depending on the unit of randomisation), exhibiting attrition rates of 10% or less, and demonstrating robust academic outcomes such as GPA, standardised tests or exam scores.

Studies receiving a 3* rating were characterised by slightly smaller sample sizes, though still surpassing 100 students, or they exhibited dropout rates between 10% and 20%. Studies rated as 0* to 2* faced more significant quality issues, such as smaller sample sizes or higher dropout rates. Given the substantial number of 3* and 4* studies found, which is unusual, this paper focuses on this high-quality evidence for addressing the research question.

The threshold of 400 students to achieve a 4* rating on the scale may appear somewhat arbitrary; however, the sieve methodology grants researchers the flexibility to determine the minimum threshold necessary, based on their professional judgement, to ensure the reliability of study outcomes. Nonetheless, it is recommended that each comparison group include at least several hundred cases to secure the highest rating for this item (Gorard, 2024). In this specific case, although the threshold was set at 400 students for randomisation at this level, lowering it to 350 or even 300 would have

TABLE 1 'Sieve' approach to estimate the trustworthiness of each study.

Design	Scale	Completeness of data	Data quality	Rating
Strong design for research question	Large number of cases per comparison group	Minimal missing data, no evidence of impact on findings	Standardised, independent, pre-specified, accurate	4
Good design for research question	Medium number of cases per comparison group	Some missing data, possible impact on findings	Standardised, independent, not pre-specified, some errors	3
Weak design for research question	Small number of cases per comparison group	Moderate missing data, likely impact on findings	Not standardised, independent, or pre-specified, some errors	2
Very weak design for research question	Very small number of cases per comparison group	High level of missing data, clear impact on findings	Weak measures, high level of error, too many outcomes	1
No consideration of design	A trivial scale of study, or number is unclear	Huge amount of missing data, or not reported	Very weak measures, or accuracy not addressed	0

Source: Gorard (2021).

minimal impact, as studies rated 3* for this item all report sample sizes below 250 at the unit of randomisation.

Regarding data completeness, studies were assigned a 3* rating when concerns about data competition were notable but not significant enough to cast doubt on the reliability of the results. Studies with data completeness below 90% were not awarded the highest rating for this item, and if attrition exceeded 30%, the rating was reduced to 2* or lower.

The fourth criterion in the sieve quality assessment is data quality. Most studies did not encounter issues in this category, as they employed standardised tests, GPA and end-of-year grades as their outcomes. A few studies exhibited lower data quality due to the use of less standardised tests as outcomes. However, most of these studies faced more significant concerns in other evaluation criteria, meaning that data quality was not the primary factor influencing their final rating.

To ensure accurate ratings of the studies, two additional experienced researchers independently assessed a subsample of the articles using the same framework at the outset. Furthermore, the same research team reviewed all studies with any uncertainty regarding their assigned ratings.

The assessments of the three reviewers were used to calculate Fleiss's Kappa, a statistic that measures interrater reliability, following Fleiss (1977). Fleiss's Kappa (κ) is an extension of Cohen's Kappa (κ), designed for scenarios involving more than two raters. It quantifies the level of agreement among raters who assign categorical ratings to a set of studies by comparing the observed agreement across all items with the agreement expected by chance. For the subsample of articles evaluated by the three researchers, Fleiss's Kappa (κ) was 0.714, indicating substantial agreement according to the benchmark proposed by Landis and Koch (1977).

Transparency and openness

We adhered to the PRISMA 2020 guidelines for systematic reviews (Page et al., 2021). All data and research materials are available at https://osf.io/wv2eu/?view_only=c5db7e87d1af4f72bfb16faae6b49095. This review was not preregistered.

RESULTS

The analysis included 29 papers that met our selection criteria, namely, implementing RCTs to assess the impact of intelligence growth mindset interventions on the academic achievement of school-age students. Among these studies, three studies (Table 2) were excluded due to the inability to compute effect sizes for the randomised group, as they solely reported the intervention's impact within a subgroup (Paunesku et al., 2015; Yeager et al., 2016, 2019).

Rating of the quality of the evidence

In the 26 remaining papers, there were 62 unique academic outcomes. In cases where multiple academic outcomes were reported, the review prioritised the most trustworthy ones for analysis. For instance, if a study provides both an Intent to Treat (ITT) estimate with low attrition and Average Treatment Effect on the Treated (ATT) estimates with high attrition, the analysis uses the ITT estimates, as a better estimate of impact for the randomised group.

These outcomes were subsequently classified utilising the sieve analysis framework. Remarkably, 14 of the 62 outcomes received the highest security rating of 4*, signifying robust findings. Additionally, 16 outcomes were attributed a 3* rating, 25 were categorised as 2*, and 7 outcomes received ratings of 1* or 0* (Table 3).

Reasons for studies getting 2* or lower rating

Several studies in the analysis received ratings of 2* or lower, primarily due to specific factors affecting their research design and outcomes.

Nine papers, comprising a total of 25 outcomes, were categorised as 2* for two key reasons. Firstly, both Alan et al. (2019) and Wilkins (2014) employed relatively small-scale randomisation. Alan et al. (2019) randomised 16 schools, while Wilkins (2014) randomised 16 teachers. Despite involving a significant number of students, the limited number of units randomised raises concerns about potential imbalances between the treatment and control groups. Zhao et al. (2024) randomised a large cohort of students at the classroom level. However, the exact number of randomised classrooms was not specified in the study, though it is estimated that between 30 and 40 classrooms were included, with 20–25 classes in the treatment group and 10–15 in the control group. While this sample size is not particularly small, the groups were not balanced in terms of age and gender. A higher proportion of boys in the control group may have influenced the slower progress observed in language subjects compared to the relatively better progress in mathematics. Finally, the studies by Brougham and Kashubeck-West (2017), with a sample size of approximately 80 students, and Good et al. (2003), which randomised 138 participants into four groups, yielded findings that are less reliable compared to higher-rated studies.

TABLE 2 Reasons for excluding some studies.

Study	Reason for exclusion
Paunesku et al. (2015)	The study does not provide results for the entire randomised group; it reports results solely for at-risk students
Yeager et al. (2016)	The study only reports results for students who were either 1 standard deviation above or below the group's prior achievement measure, not for the entire randomised group
Yeager et al. (2019)	The study exclusively presents results for lower-achieving students and does not report results for the entire randomised group

TABLE 3 Number of outputs by security rating.

Security rating	Number of outputs
4*	14
3*	16
2*	25
1*	5
0*	2
Total	62

Rienzo et al. (2015), Bettinger et al. (2018) and Porter et al. (2020) received a 2* rating due to significant attrition rates, hovering around 30%. Lastly, Blau and Benolol (2016) received a 2* rating because the outcomes measured extended beyond strictly academic criteria, encompassing various aspects of a computer programming project.

Three studies with four outcomes ended up with a 1* rating. Schrodt (2015) had a small sample size of only 27 students. Blackwell et al. (2007) featured a sample size of 87 students, but they were organised into advisory groups, with just approximately 8 advisory groups being randomised. Dommett et al. (2013) randomised five schools into five different conditions, including two treatment conditions, two active control conditions, and one no-intervention control.

Two studies with two outcomes ended up in the lower category, with a 0* rating. Orosz et al. (2017) and Glerum et al. (2020) both had 55 students, and randomisation occurred at the classroom level, further reducing the effective sample size. The rating for each study appears in Table A1.

Effect sizes

Figures 2 and 3 present the effect sizes of the findings categorised by the security levels achieved by individual studies. In these graphs, findings on the left side are considered less robust, while those on the right are more secure. The effect sizes are on the y axis. Dots located below the horizontal zero line represent findings with negative effect sizes.

As shown in Figure 2, the weakest studies reported the biggest effect sizes. In fact, lower-rated findings exhibit a wider distribution compared to their higher-rated counterparts. Findings rated with a modest 0* to 2* security rating range from −0.7 to +1.9. In contrast, effect sizes for 3* and 4* rated outcomes span a narrower range of −0.15 to +0.4. Studies rated 4* are closely clustered around zero impact.

For convenience, Figure 3 provides a zoomed-in view of Figure 2, based only on 3* and 4* studies. As seen in Figure 2, 3* findings range from −0.15 to 0.38, while 4* ratings have an even narrower range, ranging from −0.01 to 0.064.

Table 4 shows in detail the effect size calculated for each study and for each of the outcomes included in the analysis.

Table 5 presents the results of each study included in this structured review in a concise manner, organised according to the strength of the evidence.

Results for studies with a 4* security rating

This section considers the five studies that carry a 5* security rating, encompassing a total of 14 outcomes. Three 4* studies reveal positive although small effect sizes. Outes-León et al. (2020) implemented a school-level RCT in Peru, enrolling 800 urban secondary schools.

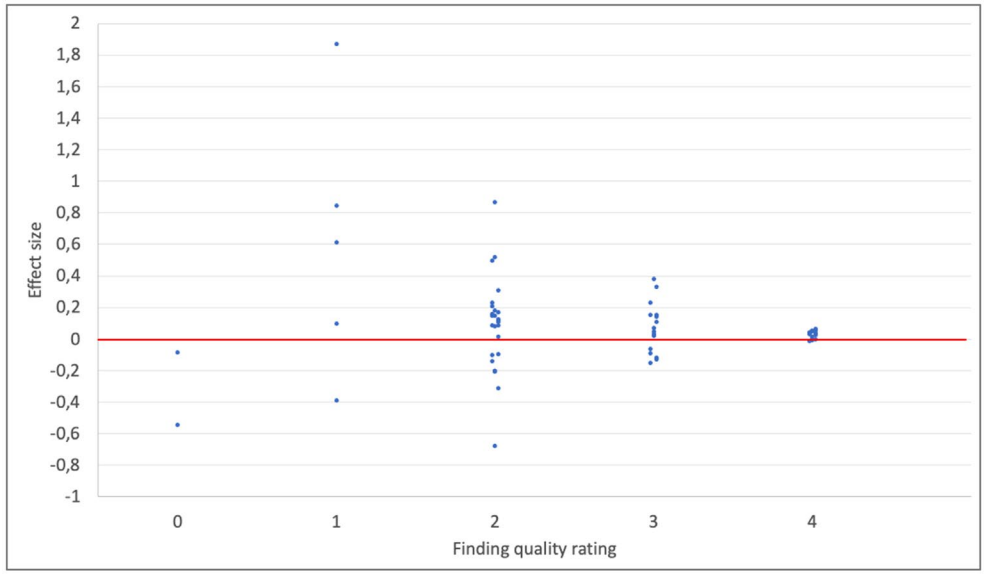

FIGURE 2 Effect sizes of growth mindset interventions by security level 0 to 4.

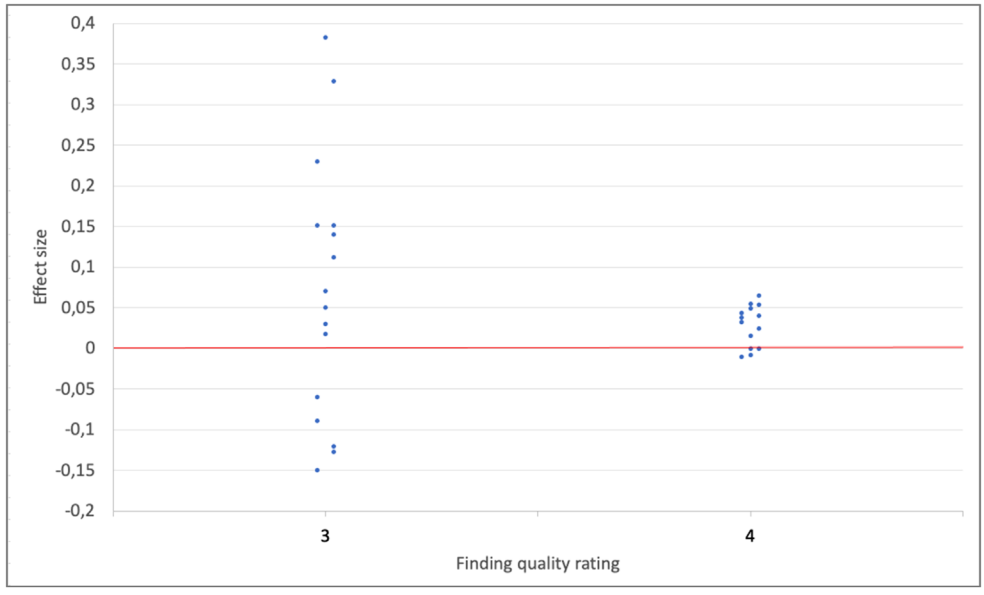

FIGURE 3 Effect sizes of growth mindset interventions by security levels 3 and 4.

Within this framework, 400 schools in the treatment group received materials and information to implement three 30-min sessions about growth mindset. It is worth noting that it was voluntary for teachers to implement these sessions, leading to approximately half of the schools not participating. Due to the low compliance, an ITT estimate and ATT estimates were calculated. The review took into consideration the ITT estimates, because it has minimal dropout (1%), resulting in a 4* security rating. Two months post-intervention, the effect on maths standardised tests yielded $d = 0.054$, while reading standardised tests showed $d = 0.04$. Follow-up assessments conducted 14 months later exhibited maths $d = 0.038$ and reading $d = 0.044$.

TABLE 4 Effect size for individual studies according to the security level 0 to 4.

TABLE 5 Strength of evidence and impact for included studies.

Strength of evidence	Positive	Unclear/mixed/very small	Negative
4🔒		5	
3🔒	3	3	1
2🔒	5	3	1
1🔒	2		1
0🔒		1	1

With similar results, Rege et al. (2021) carried out an individual-level RCT in Norway encompassing 6541 secondary school students. The intervention sought to convey the concept that engaging in challenging tasks can enhance one's abilities over time, particularly during adolescence when the brain experiences significant growth. The study found an ITT impact ($d = 0.025$) on advance maths course enrolment and passing rates. The authors explain that in some schools the choice of maths course was made before the intervention, and even though the students could change their choices later, this situation might attenuate the effect size of the impact.

Zhou et al. (2023) conducted an individual-level RCT in rural China, involving 1680 Year 5 students. A significant proportion of these students were classified as low-income and had absent parents. The study evaluated the effect of an intensive mindset intervention, which consisted of weekly 2-h sessions over 5 weeks, on official centrally graded exams in mathematics and Chinese. The outcomes were measured both shortly after the intervention and 6 months later. With very low attrition rates (2% at midline and 4.1% at endline), the study found little evidence of an impact on test scores, reporting effect sizes between 0.03 and 0.06 standard deviations in Chinese and mathematics, respectively, at both time points.

The remaining two studies, also categorised with a 4* finding security rating, exhibited effect sizes around 0, encompassing both marginal negative and marginal positive results. Foliano et al. (2019) conducted a school-level RCT involving 101 primary schools (encompassing 5018 Year 6 students) in England. The treatment group's headteachers and Year 6 teachers received training to deliver an eight-week programme (up to 2.5 h a week) aimed at developing a growth mindset in their students. Despite the intervention, no improvement was noted in Key Stage 2 tests in reading, grammar, punctuation, spelling or maths in comparison to the control group (KS2 maths $d = -0.01$, KS2 reading $d = 0.00$, KS2 grammar punctuation and spelling $d = 0.00$).

Similarly, Ganimian (2020) randomly assigned 202 public secondary schools in Argentina to either a treatment or control group, including 9805 students. The treatment involved Grade 12 students participating in a single session focused on the malleability of intelligence. Despite implementing the intervention as intended, no evidence emerged of an impact on maths or reading national assessments (with effect sizes of $d = 0.015$ and $d = -0.008$, respectively).

Results for studies with a 3* security rating

This section presents discussion of seven studies characterised by a 3* security rating, with a total of 16 outcomes. These studies present robust RCT designs, involving a substantial number of students with relatively low attrition rates. Nevertheless, they do not achieve the 4* rating because they are not as strong as the studies described in the previous section. The primary reasons for these 3* ratings include that there are relatively

smaller sample sizes, typically ranging between 100 and 200 students, in contrast to the 4* studies, which encompassed sample sizes exceeding 400 students (Burnette et al., 2018; Wanzek et al., 2021; Xu et al., 2021). Additionally, Chao et al. (2017) randomised 107 classes but distributed them across six different groups, resulting in 17 to 20 classes per group. Furthermore, three studies received 3* ratings due to dropout rates ranging between 13% and 19% (Binning et al., 2019; Polley, 2018; Porter et al., 2022).

Among the 3* studies, four demonstrated positive effects. The experiment conducted by Xu et al. (2021) yielded the most substantial effect size among both 3* and 4* studies. This study was conducted in a laboratory setting and involved 140 16-year-old students from two public high schools. In the experiment, students in the growth mindset condition engaged in a 10-min reading and writing task on brain function and the malleability of intelligence, whereas the control group undertook a similar-length task concerning general brain functioning (without addressing intelligence malleability). Subsequently, both groups participated in a 12-min learning task related to the topic of sound travel and the Doppler effect, an area in which participants were novices. Post-assessment revealed that students in the treatment group outperformed their peers in the control group in both comprehension and transfer tests (Cohen's $d = 0.33$ for comprehension and 0.38 for transfer). This notably larger effect, in comparison to the other 3* rated findings, could potentially be attributed to the evaluation of a more specialised learning task within a controlled laboratory environment.

Porter et al. (2022) explored a teacher-delivered growth mindset intervention, involving 50 middle school teachers and 1996 students. Teachers in the treatment group underwent training receiving curriculum guides, lesson plans and research on growth mindset. They subsequently crafted personalised intervention plans for their students and administered lessons lasting between 30 min and 1 h during regular classes. Students exposed to these lessons outperformed their counterparts in the control group, both in classes where growth mindset was taught (Cohen's $d = 0.23$) and in other classes (Cohen's $d = 0.14$).

Polley (2018) conducted an RCT in Bangladesh involving 600 secondary school students and a 45-min computer-based growth mindset intervention. The control group received an identical intervention, with the sole distinction being the absence of commentary on whether the brain changes during the learning process. The intervention resulted in an average increase in test scores of 0.12σ (with Cohen's d values of 0.112 for maths, 0.151 for science, 0.018 for English, and 0.151 for world studies exams).

Burnette et al. (2018) conducted an RCT to assess a 45-min online growth mindset intervention with 222 10th-grade adolescent girls from four rural, low-income high schools in the United States. They found an effect on GPA with a Cohen's d of 0.07.

Three studies rated 3* reported negative results. Chao et al. (2017) focused on primary students in low-income areas of India, assessing the impact of a growth mindset intervention alongside an incentive system (small rewards given to students). In this study, 107 classes (949 students) were randomised into six groups, comprising three treatments that combined a growth mindset intervention with different incentive interventions (treatment 1: only growth mindset, treatment 2: growth mindset + incentive 1, treatment 3: growth mindset + incentive 2) and three control interventions with no growth mindset but with different types of incentives (control, control + incentive 1, control + incentive 2). The growth mindset intervention exhibited a negative effect on standardised test scores ($d = -0.13$). Binning et al. (2019) conducted a study involving 598 students from two high schools, randomising them into either a growth mindset or control (study skills) intervention, resulting in a negative effect on GPA ($d = -0.09$). Lastly, Wanzek et al. (2021) compared the effect of a reading intervention versus a reading plus mindset intervention in 240 fourth-grade students with reading difficulties. This study reported negative or near-null effect sizes across various reading components (oral reading fluency $d = -0.15$, non-word reading $d = 0.05$, phonological processing $d = -0.12$, word reading $d = -0.06$, reading comprehension $d = 0.03$).

Bias due to conflict of interest?

It is clear that smaller and weaker studies have tended to produce the largest apparent effect sizes, whereas the 4* studies have shown little or no promise from the intervention. Another factor to consider is the role of conflicts of interest for the researchers, given the proliferation of enterprises and consultancy groups that have developed around the growth mindset theory. Moreover, since the theory's widespread popularity, a substantial amount of funding has been allocated by schools, school districts, governments and non-governmental organisations for mindset training, interventions, books and related initiatives.

In assessing whether researchers' conflicts of interest are related to experimental results, the review examined all studies where at least one of the researchers had a conflict of interest. This was defined as any direct involvement such as being a founder, co-founder, owner, part of the board, or receiving paid compensation from organisations that sell products related to mindset training, including books, seminars or consulting services, or if the researcher charged substantial fees for talks on this topic.

As shown in Table 6, seven papers had authors with an apparent conflict of interest, encompassing nine outcomes: one 4* outcome, three 3* outcomes, three 2* outcome and two 1* outcomes.

Figure 4 provides a visual representation of the findings. Studies found to have conflicts of interest are marked with crosses, while those without conflicts of interest are indicated by circles. Notably, most studies with conflicts of interest show positive effect sizes, apart from Orosz et al. (2017), a 1* rated study that reports a small negative effect. Conversely, studies without conflicts of interest display a range of impacts, including both positive and bigger negative results.

Figure 5 provides a closer look at only the 3* and 4* findings. Even though the four outcomes reported in studies with conflict of interest have positive effects, the size of the effects does not differ much from the size of effects reported by authors without a conflict of interest.

Considering that 3* and 4* studies conducted by authors with conflicts of interest do not have bigger effect sizes than the larger effect size from studies without such conflicts, the review does not find much cause for concern that the size of the results we are currently observing are exaggerated due to these conflicts. Nevertheless, it is noteworthy that studies involving conflicts of interest mostly report positive effect sizes. This could mean that some studies with negative results might not be getting published, partly due to the conflicts of interest held by their authors. This suspicion is further supported by the fact that the three studies excluded from our analysis because they failed to report results for the entire randomised group and did not provide sufficient information to calculate results for the entire sample, were authored by individuals with conflicts of interest (Paunesku et al., 2015; Yeager et al., 2016, 2019). Although these papers explicitly focused on assessing the impact on specific groups (such as at-risk students or low achievers), and two of them had even pre-registered this focus, it is surprising that they did not report results for the entire randomised group alongside the outcomes for their targeted group of interest.

The conflict-of-interest bias could affect findings of any meta-analysis due to the aggregation of effect sizes. Conversely, in our analysis, because we display each effect size independently, the absence of certain negative results does not substantially alter the overall conclusion. This can explain why the conflict-of-interest was so concerning in Macnamara and Burgoyne's (2022) meta-analysis. They reveal a significantly lower average effect size for published studies lacking financial incentives, in contrast to a notably higher effect size for studies that do have financial incentives. This disparity is primarily driven by the scarcity of negative results within studies with financial incentives.

TABLE 6 Studies with an apparent conflict of interest.

Study	Authors	Reasons of conflict of interest	Source
Bettinger et al. (2018)	E. Bettinger, S. Ludvigsen, M. Rege, I.F. Solli, D. Yeager	D.S. Yeager offers mindset talks for a fee	https://thelavinagency.com/speakers/david-yeager/
Blackwell et al. (2007)	L.S. Blackwell, K.H. Trzesniewski, C.S. Dweck	L.S. Blackwell is the president and co-founder of Mindset Works, C.S. Dweck co-founded MindsetWorks, only in 2015, she separated from the organisation to avoid conflicts of interest	https://www.mindsetworks.com/about-us/default https://www.speakerbookingagency.com/talent/carol-dweck https://www.allamericanspeakers.com/celebritytalentbios/Lisa+Blackwell,+PhD/399102
Burnette et al. (2018)	J.L. Burnette, M.V. Russell, C. L. Hoyt, K. Orvidas, L. Widman	K. Orvidas facilitates Self-Help Workshops on mindsets for a fee	https://www.kaseyorvidas.com/programs
Good et al. (2003)	C. Good, J. Aronson, M. Inzlicht	J. Aronson serves as an executive advisor to Casa Laxmi, providing expertise in the field of mindset, and has authored books on this subject	https://www.casalaxmi.com/our-team/executive-advisors/
Porter et al. (2022)	T. Porter, D. Catalán Molina, A. Cimpian, S. Roberts, A. Fredericks, L.S. Blackwell, K. Trzesniewski	Lisa S. Blackwell is the president and co-founder of Mindset Works	https://www.mindsetworks.com/about-us/default.
Rege et al. (2021)	M. Rege, I.F. Solli, S. Ludvigsen, R. Crosnoe, C. Muller, A. Duckworth, P. Hanselman, C.S. Dweck, E. Bettinger, G. Walton, D.S. Yeager	David S. Yeager offers mindset talks for a fee. Angela Duckworth is a member of Amazon's Insight to Entrepreneurs team, where she actively promotes the growth mindset. She is also an author of books on the topic and a co-founder of Character Lab, an organisation dedicated to fostering a growth mindset among other initiatives	https://www.amazon.com/Angela-Duckworth-Insights-for-Entrepreneurs/b?ie=UTF8&node=17395092011 https://angeladuckworth.com/about-angela/about-character-lab/ https://thelavinagency.com/speakers/david-yeager/
Orosz et al. (2017)	G. Orosz, S. Péter-Szarka, B. Bothe, I. Tóth-Király, R. Berger	Gábor Orosz is a co-founder of the Heroes Square Initiative, an organisation that offers mindset programmes for schools and teachers Rony Berger is an Advisory Board Member of the Center for Compassion and Altruism Research and Education, an organisation offering training programmes, on growth mindset among other topics	https://hosoktere.org/kik-vagyunk/#tortenetunk https://ccare.stanford.edu/people/rony-berger-psyd/

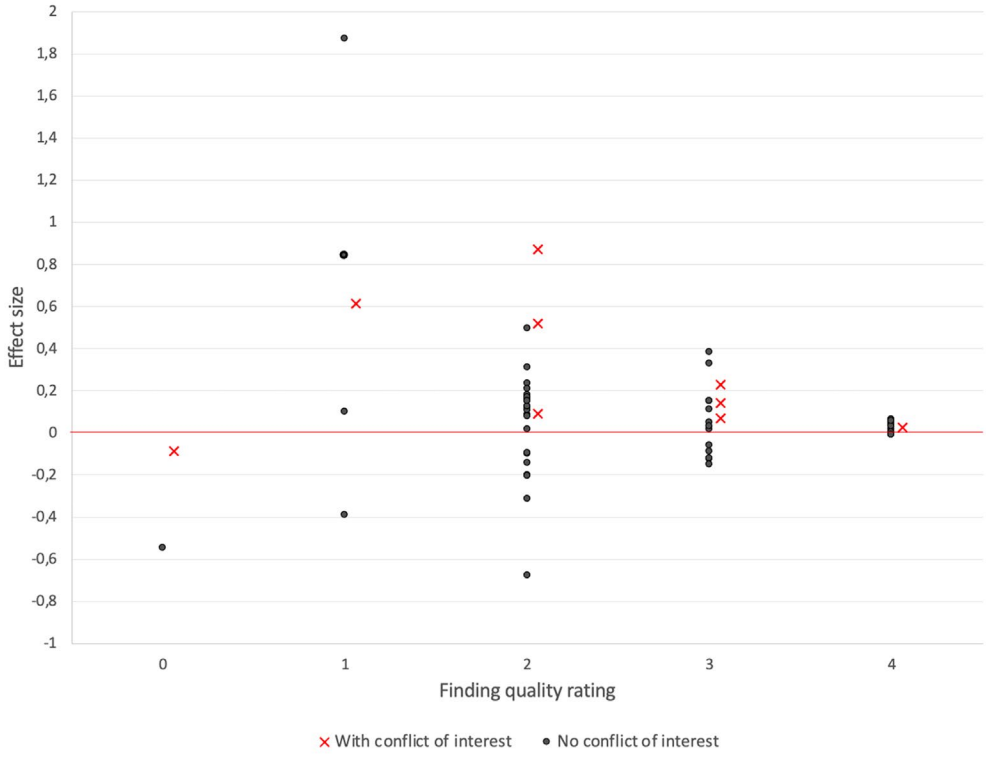

FIGURE 4 Effect size by conflict of interest by security levels 0 to 4.

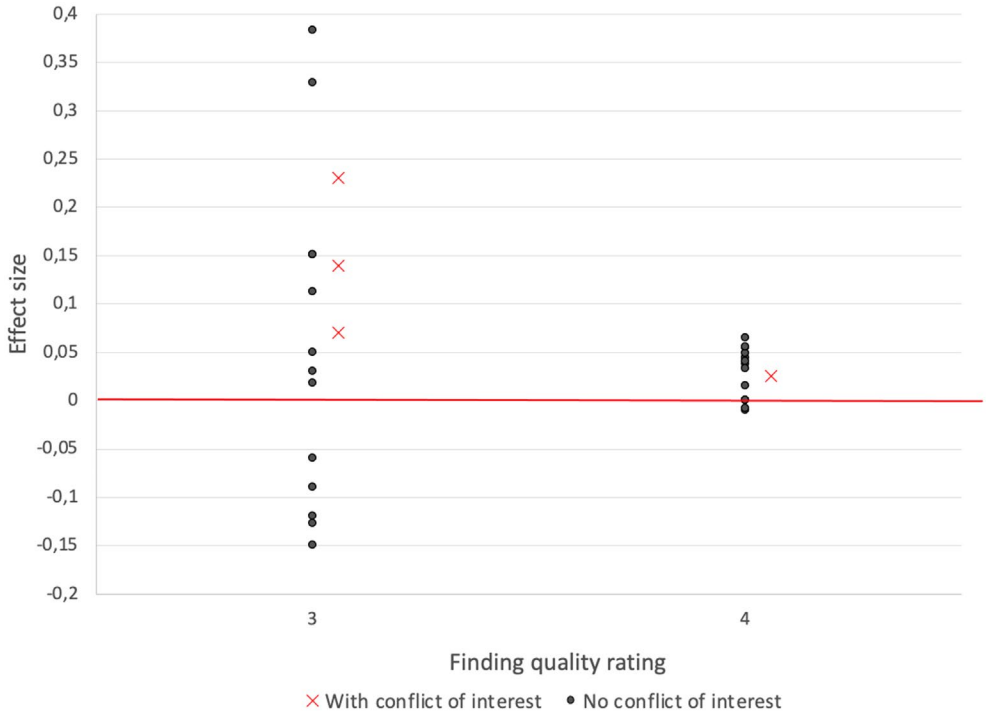

FIGURE 5 Effect size by conflict of interest, 3* and 4* findings.

DISCUSSION

The review reveals a large number of experiments that have tested the effectiveness of growth mindset interventions in enhancing students' academic achievement. While some of these experiments had major problems, which make it difficult to trust their findings, many RCTs produced robust findings, due to the rigorous research design, independent evaluations, substantial sample sizes, minimal data gaps and data quality.

At the same time, the analysis illustrates that the 'sieve' analysis, as proposed by Gorard (2021) for assessing evidence, is an effective and valuable tool to evaluate the trustworthiness of evidence. Papers evaluated more favourably using this framework tend to exhibit greater coherence and consistency in their results.

Due to the abundance of RCTs on this topic, and the availability of some very high-quality evidence, it is most reasonable to answer our research question taking into consideration the most robust research. Among outcomes rated 3* and 4*, the effect sizes span from −0.15 to +0.38. It is worth noting that the two notably higher effect sizes, exceeding 0.3, emanate from a single laboratory setting experiment. Recognising the substantial divergence between such controlled conditions and the real-life educational environments where the intervention would be implemented, it is reasonable to consider excluding these outcomes in a review of pragmatic trials. This would result in a more constrained range of effect sizes, ranging from −0.15 to +0.23.

When we give greater significance to the 4* studies, because of their heightened methodological trustworthiness, the range of effect sizes narrows further, encompassing −0.01 to 0.065. Considering these factors, we can cautiously infer that implementing a growth mindset intervention might, in the most favourable circumstances, produce a very modest impact, potentially reaching a Cohen's *d* effect size of 0.05, on the academic performance of school-age students. However, given the null impact observed in two of the four most rigorous evaluations (Foliano et al., 2019; Ganimian, 2020), we do not recommend the allocation of substantial resources for implementing public policies aimed at promoting a growth mindset among school-age students. And this is the headline finding of the review.

Nevertheless, due to the cost-effectiveness of growth mindset interventions and the modest yet positive outcomes reported in the remaining three highly rigorous evaluations (Outes-León et al., 2020; Rege et al., 2021; Zhou et al., 2023), we suggest that teachers be allowed to implement these interventions in their classrooms, as long as it is at minimal cost and with limited impact on students' time.

Limitations

The conclusions drawn from this structured review are based on various studies employing RCT designs. Therefore, the reliability of this study's conclusions depends on the extent to which we can trust the findings of the papers included in the review. To safeguard against this, we assessed the quality of the studies using the 'sieve' approach, as proposed by Gorard (2021). However, this evaluation framework focuses on the most relevant factors affecting study quality, particularly those that pose the greatest threat to the reliability of their conclusions. Nevertheless, other aspects that could potentially jeopardise the quality of a study have not been considered in this review. For example, if the RCTs are unblinded, this could introduce teacher bias or research expectancy effects. We have also not addressed the potential issue of studies attempting to measure too many outcomes. Finally, preregistration of studies was not a requirement for achieving the highest quality rating.

Another limitation of this study is that only studies published in English were included. As a result, studies published in other languages, as well as unpublished studies, were excluded, potentially leading to publication bias. Due to this bias, where studies with positive results are more likely to be published, it is possible that studies with negative or neutral outcomes have not been published and, therefore, have not been included in this review.

Future research direction

While the literature consistently demonstrates a correlation between academic outcomes and growth mindset, this study reveals that the most rigorous causal studies show only weak, if any, effects of mindset interventions on academic performance. This is despite many interventions successfully shifting students from a fixed mindset to a growth mindset. This finding prompts the critical question of whether these interventions are genuinely altering students' underlying beliefs or merely influencing their responses to mindset questionnaires.

Future research should investigate the true malleability of the growth mindset and assess whether the widely adopted, low-cost, low-intensity interventions are genuinely capable of fostering meaningful changes in students' beliefs, or if they primarily affect how students respond to questionnaires. It is vital to distinguish whether the observed shifts in mindset represent deeper, enduring changes in students' attitudes and behaviours, or whether they merely reflect superficial adjustments in self-reporting.

Moreover, future studies should focus on evaluating more intensive interventions that engage students at a deeper, experiential level. These interventions should prioritise how students actively experience and apply growth mindset principles, as opposed to simply being introduced to them in an abstract and with stories about other people. Such comprehensive engagement could result in more profound shifts in students' core beliefs and produce long-term improvements in their learning and development.

AUTHOR CONTRIBUTIONS

Carolina Gazmuri: Conceptualization; investigation; writing – original draft; methodology; validation; visualization; writing – review and editing; software; formal analysis; project administration; data curation.

ACKNOWLEDGEMENTS

The author would like to acknowledge the valuable guidance and feedback from Nadia Siddiqui and Stephen Gorard.

CONFLICT OF INTEREST STATEMENT

The author declares that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OSF at https://osf.io/wv2eu/?view_only=c5db7e87d1af4f72bfb16faae6b49095, reference number doi 10.17605/osf.io/wv2eu.

ETHICS STATEMENT

I adhered to the PRISMA 2020 guidelines for systematic reviews (Page et al., 2021). This review was not preregistered.

ORCID

Carolina Gazmuri https://orcid.org/0009-0002-8014-3257

REFERENCES

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

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education* (pp. 1–181). Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00001-8

Barnett, M. K., & Macnamara, B. N. (2023). Individual responses versus aggregate group-level results: Examining the strength of evidence for growth mindset interventions on academic performance. *Journal of Intelligence*, *11*(6), 104. https://doi.org/10.3390/jintelligence11060104

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

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

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

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

Brougham, L., & Kashubeck-West, S. (2017). Impact of a growth mindset intervention on academic performance of students at two urban high schools. *Professional School Counseling*, *21*(1), 2156759X1876493. 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. (2022). A systematic review and meta-analysis of growth mindset interventions: For whom, how, and why might such interventions work? *Psychological Bulletin*, *149*, 174–205. https://doi.org/10.1037/bul0000368

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. *British Journal of Educational Psychology*, *88*(3), 428–445. https://doi.org/10.1111/bjep.12192

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

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

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

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, *45*(3–4), 294–309. https://doi.org/10.1007/s10464-010-9300-6

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

Elias, M. J., & Haynes, N. M. (2008). Social competence, social support, and academic achievement in minority, low-income, urban elementary school children. *School Psychology Quarterly*, *23*(4), 474–495. https://doi.org/10.1037/1045-3830.23.4.474

Fleiss, J. L. (1977). The measurement of interrater agreement. In *Statistical methods for rates and proportions* (2nd ed., pp. 212–236). Wiley.

Foliano, F., Rolfe, H., Buzzeo, J., Runge, J., & Wilkinson, D. (2019). *Changing Mindsets: Effectiveness trial Changing Mindsets Evaluation Report*. www.educationendowmentfoundation.org.uk

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

Glerum, J., Loyens, S. M. M., & Rikers, R. M. J. P. (2020). Is an online mindset intervention effective in vocational education? *Interactive Learning Environments*, *28*(7), 821–830. https://doi.org/10.1080/10494820.2018.1552877

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

Gorard, S. (2021). *How to make sense of statistics: Everything you need to know about using numbers in social science*. SAGE.
Gorard, S. (2024). Judging the relative trustworthiness of research results: How to do it and why it matters. *Review of Education*, *12*(1), e3448. https://doi.org/10.1002/rev3.3448
Heckman, J. J., & Rubinstein, Y. (2001). The benefits of skill: The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, *91*(2), 145–154. https://doi.org/10.1257/aer.91.2.145
Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482. https://doi.org/10.1086/504455
Hoang, T. V. (2018). *Growth mindset and task value interventions in college algebra* [Doctoral dissertation, Texas State University]. Texas State University. https://digital.library.txstate.edu/handle/10877/7423
Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., & Borghans, L. (2014). Improving Cognitive and Non-cognitive skills to Promote Lifetime Success. http://www.nber.org/papers/w20749
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.
Macnamara, B. N., & Burgoyne, A. P. (2022). 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/bul00003522022
Macnamara, B. N., & Burgoyne, A. P. (2023). 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). *Psychological Bulletin*, *149*(3–4), 242–258. https://doi.org/10.1037/bul0000394
Orosz, G., Péter-Szarka, S., Bothe, 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*(MAR), 311. https://doi.org/10.3389/fpsyg.2017.00311
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. http://www.worldbank.org/prwp
Owen, K., Watkins, R., & Hughes, C. (2022). From evidence-informed to evidence-based: An evidence building framework for education. *Review of Education*, *10*(1), 1. https://doi.org/10.1002/rev3.3342
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., & Chou, R. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71.
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
Polley, T. (2018). Believing is Achieving: Targeting Beliefs about Intelligence to Increase Learning in Bangladeshi Secondary Schools. https://sites.duke.edu/thomaspolley/
Porter, T., Catalán Molina, D., Cimpian, A., Roberts, S., Fredericks, A., Blackwell, L. S., & Trzesniewski, K. (2022). Growth-mindset intervention delivered by teachers boosts achievement in early adolescence. *Psychological Science*, *33*(7), 1086–1096. https://doi.org/10.1177/09567976211061109
Porter, T., Martinus, A., Ross, R., Cyster, C. F., & Trzesniewski, K. (2020). Changing learner beliefs in south African townships: An evaluation of a growth mindset intervention. *Social Psychological and Personality Science*, *11*(7), 991–998. https://doi.org/10.1177/1948550620909738
Rege, M., Hanselman, P., Solli, I. F., Dweck, C. S., Ludvigsen, S., Bettinger, E., Crosnoe, R., Muller, C., Walton, G., Duckworth, A., Yeager, D. S., Levi, M., Shankar, M., Brock, T., Romero, C., Macrander, C., Wilson, T., Konar, E., Horng, E., & Bloom, H. (2021). How can we inspire nations of learners? An investigation of growth mindset and challenge-seeking in two countries or to, *76*(5), 755–767. https://doi.org/10.3886/ICPSR37353.v1
Rienzo, C., Rolfe, H., & Wilkinson, D. (2015). *Education Endowment Changing Mindsets Evaluation report and Executive summary Independent evaluators*. www.educationendowmentfoundation.org.uk
Schrodt, K. M. (2015). *The Relationship Among Mindset Instruction, Kindergarteners' Performance, and Motivation in Writer's Workshop*. Middle Tennessee State University.
Siddiqui, N., & Ventista, O. (2018). A review of school-based interventions for the improvement of social emotional skills and wider outcomes of education. *International Journal of Educational Research*, *90*, 117–132.
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
Tipton, E., Bryan, C. J., Murray, J. S., & Lynn Schneider, B. (2022). Why meta-analyses of growth mindset and other interventions should follow best practices for examining heterogeneity. *Psychological Bulletin*, *149*(3–4), 229–241. https://doi.org/10.13140/RG.2.2.34070.01605
Wanzek, J., Al Otaiba, S., Petscher, Y., Lemons, C. J., Gesel, S. A., Fluhler, S., Donegan, R. E., & Rivas, B. K. (2021). Comparing the effects of Reading intervention versus Reading and mindset intervention for upper

elementary students with Reading difficulties. *Journal of Learning Disabilities*, *54*(3), 203–220. https://doi.org/10.1177/0022219420949281

Wilkins, P. B. B. (2014). Efficacy of a growth mindset intervention to increase student achievement. http://digitalcommons.gardner-webb.edu/education_etd

Xu, Q., Wormington, S., Guzman-Alvarez, A., & Wang, M. T. (2021). Why does a growth mindset intervention impact achievement differently across secondary schools? Unpacking the causal mediation mechanism from a National Multisite Randomised Experiment. *Journal of Research on Educational Effectiveness*, *14*(3), 617–644. https://doi.org/10.1080/19345747.2021.1894520

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., Hulleman, C. S., Hinojosa, C., Lee, H. Y., O'Brien, J., Romero, C., Paunesku, D., Schneider, B., 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

Zhao, S., Yu, C., Jin, L., & Lin, D. (2024). Helping students to face academic failures: Evaluation of a growth mindset intervention among primary school students in China. *Applied Psychology. Health and Well-Being*, *16*(2), 397–420.

Zhou, L., Fagernas, S., Hussain, I., Li, L., Pelkonen, P., & Huang, Z. (2023). Changing beliefs and attitudes of primary school children: Growth mindset in rural China.

How to cite this article: Gazmuri, C. (2025). Can growth mindset interventions improve academic achievement? A structured review of the existing evidence. *Review of Education*, *13*, e70066. https://doi.org/10.1002/rev3.70066

APPENDIX

TABLE A1 Main reasons for the rating of each study.

Paper	Rating	Main concern influencing the rating.
Alan et al. (2019)-S2	2	Despite a large number of students, the randomisation units are only 16 schools, leading to the likelihood of unbalanced groups. Specifically, the two groups exhibit imbalances in verbal test scores. Additionally, only short-term scores were considered due to a 48% dropout rate in the long term
Bettinger et al. (2018)	2	The study experienced a 28% dropout rate because of students not taking the test. Only the overall test effect were reported and not some portions of the test reported in the paper
Binning et al. (2019)	3	The study does not provide clarity regarding the number of students included in the regression analysis used to calculate the impact. However, based on the available information, a dropout rate of 13% has been calculated
Blackwell et al. (2007)	1	Although 95 students participated, they were grouped into advisory groups consisting of 12–14 students, and these advisory groups were randomised. Consequently, we determined that approximately 7 or 8 advisory groups were randomised, which is a small randomisation size
Blau and Benolol (2016)	2	The outputs are not robust for assessing academic achievement. They evaluated a programming project based on criteria such as the clarity of the idea, computing quality, creativity, and the aesthetics of the design. Additionally, the sample size of the study encompassed a little over 100 students
Brougham and Kashubeck-West (2017)	2	The randomisation encompasses only 89 students, and the dropout rate exceeds 20%
Burnette et al. (2018)	3	With a sample size of 222 students, it is not sufficiently large to qualify as a robust 4-star study
Chao et al. (2017)	3	The randomised units were classes, with 107 classes assigned to 6 different groups, resulting in groups consisting of 17 to 20 classes each. This size is not big enough to achieve a 4-star rating
Dommett et al. (2013)	1	The study had a small randomisation size, with only 5 schools assigned to 5 different intervention conditions
Foliano et al. (2019)	4	The study includes a significant sample size, with 101 schools assigned to either the control or treatment group. Despite an attrition rate close to 10%, a 4-star rating has still been awarded
Ganimian (2020)	4	A substantial sample size is featured in the study, with 202 schools allocated to either the control or treatment group. Despite an attrition rate nearing 10%, the study has still been awarded a 4-star rating
Glerum et al. (2020)	0	The study had a limited sample size, consisting of only 55 students, with randomisation conducted at the classroom level, further reducing the effective sample size
Good et al. (2003)	2	138 students were randomised into four groups, and the size of each group was insufficient to warrant a 3 or 4-star rating
Orosz et al. (2017)	0	The study had a limited sample size, consisting of only 55 students, with randomisation conducted at the teacher level, further reducing the effective sample size

TABLE A1 (Continued)

Paper	Rating	Main concern influencing the rating.
Outes-León et al. (2020)	4	The study benefits from a large sample size of 800 randomised schools, with minimal dropout rates for the ITT results
Polley (2018)	2–3	Clarity regarding the dropout rate is lacking in the paper; however, based on calculations, it has been estimated to range between 14% and 33% for various outcomes. Consequently, the study is rated between 2 and 3, depending on the specific outcome
Porter et al. (2020)	1–2	The study exhibits a substantial dropout rate, ranging from 36% to 57% for various outcomes. Consequently, the study is rated between 1 and 2 stars, depending on the specific outcome
Porter et al. (2022)	3	With a randomised sample of only 52 teachers, the study's sample size limits its rating to a maximum of 3 stars
Rege et al. (2021)	4	Given the substantial sample size of 6541 students and very low attrition rates for the ITT results, the study merits a 4-star rating
Rienzo et al. (2015)	2	The study has a significant dropout rate of 38%
Schrodt (2015)	1	The study has a small sample size of only 27 students
Wanzek et al. (2021)	3	A sample size of 240 students qualifies the study for a 3-star rating
Wilkins (2014)	2	The study features a small sample size, with only 16 teachers randomised
Xu et al. (2021)	3	A sample size of 140 students qualifies the study for a 3-star rating
Zhao et al. (2024)	2	Randomisation occurred at the classroom level; however, the exact number of classrooms in each group was not specified. I estimated a total of around 30–40 classrooms, with 20–25 in the treatment group and 10–15 in the control group. The groups were not balanced in terms of age or gender. There were more boys in the control group, which may have influenced the slower progress observed in the two language subjects, in contrast to the relatively better progress in mathematics
Zhou et al. (2023)	4	The sample size consisted of 1680 students, with an attrition rate between 2% and 4%. The data quality is high, as it includes results from officially centrally graded exams in both Chinese and mathematics