

Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Learning

Are school-based interventions that promote a growth mindset effective in raising pupil attainment?

Summary

The idea of a 'growth mindset', or the belief that intelligence is something we can change, is becoming increasingly popular in education due to research by Carol Dweck. Dweck's findings suggest that having a growth mindset is associated with better academic achievement, and that it is possible for students to change their mindset to this more adaptive way-of-thinking. Research has begun to combine these two aspects and evaluate whether interventions to change a student's mindset from fixed to growth can have an impact on their academic attainment. A systematic review of the literature was carried out to establish whether these interventions are effective in schools. Five studies were included and evaluated in relation to their methodological quality, methodological relevance and topic relevance to the review question. The review found some evidence to suggest that interventions that promote a growth mindset are effective at raising pupil attainment; however, this was predominantly true for either pupils identified as at risk of educational disadvantage, or pupils who held a fixed mindset prior to the intervention. Further research areas are identified based on limitations of the design of studies and gaps in participant demographics that could increase the generalisability of findings.

Introduction

What is a 'growth mindset'?

The phrase 'growth mindset' developed from Carol Dweck's research into self-theories and motivation. Dweck (2000) argues that how people view intelligence can be classified into two main types. The first is referred to as a 'fixed mindset' or entity theory, where intelligence is viewed as something that is inherent and fixed, and little can be done to change it. The opposite is termed a 'growth mindset' or incremental theory, where a person believes that intelligence is changeable and something that can develop through effort and perseverance.

Growth Mindset Interventions

As the aim of the intervention is to promote a new way of thinking, there are many different ways in which this has been implemented. Previous research has involved a teaching session, followed by a short activity where a student writes to another student outlining the theory they have learnt (Paunesku et al., 2015). Interventions similar to this have ranged from 25 minutes (Yeager et al., 2014) to almost one academic year (Good, Aronson & Inzlicht, 2003).

There is at least one commercially available intervention based on implicit theories of intelligence, Brainology[®], which was developed by Dweck and her team. Brainology[®] is an online programme that teaches children to see intelligence as malleable using material based on neuroscience and what happens to the brain when we learn. The programme consists of approximately 2.5 hours of online sessions, which can be combined with up to 10 hours of classroom-based activities. There are also free resources available online, such as Mindset Kit (<https://www.mindsetkit.org/>), which provide structured sessions for teachers, parents or senior management teams to implement with children and staff.

Psychological Theory

The theory behind growth mindsets originally stemmed from research into how children respond to failure. Diener and Dweck (1978; 1980) identified two main types of responses: one termed 'helpless' and the other 'mastery-oriented'. Diener and Dweck (1978; 1980) found that children with a helpless response viewed the failure as out of their control and unchangeable, whereas those who were mastery-oriented persisted despite the failure and focussed on mastering a task. Dweck and Leggett (1988) extended this further by linking the helpless and mastery-oriented responses to the goals that a child pursued, and how their implicit beliefs can influence the type of goal they prefer. Dweck and Leggett (1988) suggested that a child who responds to failure in a helpless manner is likely to prefer performance goals (where the aim is to gain judgement about their current competence), and hold an entity theory of intelligence. On the other hand, a mastery-oriented child would be more likely to pursue learning goals (where the aim is to increase competence) and hold an incremental theory of intelligence. Research has supported this theory and found that implicit beliefs about intelligence predicted a pupil's response to failure and goals pursued (Robins & Pals, 2002), and, importantly, predicted their grades (Blackwell, Trzesniewski & Dweck, 2007), with incremental theorists scoring higher than entity theorists.

Considering educational psychology practice, if it is possible to change a child's implicit belief to that of a growth mindset, there are clear benefits that this could have for pupil motivation and academic achievement. Interventions that promote a growth mindset could be implemented universally across schools to encourage this way of thinking in all pupils. Considering the nature of the intervention, educational

psychologists would be in the ideal position to train and support schools with its implementation.

Rationale for Review

The increased focus on accountability and the need to apply evidence-based practice in schools (DfE & DH, 2015) means that school staff need to have an understanding of the research behind interventions and a clear rationale for their implementation with specific pupils. Hattie (2008) also suggests that it is no longer enough to rely on 'what works' but that educators need to be able to compare different interventions to know which is likely to be most effective and why. There is a need, therefore, for clear unbiased reviews of school-based interventions that enable comparisons to be made between them.

The effect of growth mindset interventions on academic attainment has been positively reviewed in literature focussing on wider topics, such as psychosocial interventions in general (Spitzer & Aronson, 2015; Yeager & Walton, 2011), but as yet there has been no systematic review into the effectiveness of these interventions on pupil attainment. A systematic review was considered necessary for the following reasons. Firstly, while Yeager and Walton (2011) approached their search systematically, only specific journals were searched and many studies may have been missed. Secondly, the authors of the reviews had also written at least one of the studies included in the review. A systematic review would remove any subjectivity that could occur due to a conflict of interest. Thirdly, past reviews have included studies completed with adult students from colleges and Universities (e.g. Aronson, Fried & Good, 2002) and the findings, therefore, are less relevant to school-age children.

The current review aims to establish whether interventions that stem from this psychological theory are having the positive impact on children's attainment as hypothesised in the literature. Due to the wide variation in how these interventions can be implemented, and a dearth of research into newly-produced manualised interventions, the review will include any school-based intervention that promotes a growth mindset, regardless of how and by whom it was implemented.

Review Question

Are school-based interventions that promote a growth mindset effective in raising pupil attainment?

Critical Review of the Evidence Base

A comprehensive literature search of the following electronic databases was carried out between 18th December 2015 and 2nd January 2016:

- PsycINFO
- PsycEXTRA
- PsycARTICLES
- ERIC (EBSCO)
- ERIC (ProQuest)
- Web of Science

Search terms were created to find studies that met the following conditions:

- The research design is experimental and consisted of an intervention condition.
- The intervention was based on Carol Dweck's theory of growth mindset.

- The intervention took place in a school and the participants were school-age children.

The exact search terms used are presented below in Table 1.

Table 1
Search Terms Entered into Databases.

1	2	3
implicit intelligen* OR	intervention* OR	school* OR
malleable intelligen* OR	experiment*	child* OR
growth mindset OR	AND	AND
fixed mindset OR		educat* OR
entity theory OR		pupil* OR
incremental theory		student* OR

Note. The asterix means that any word that contains all letters before the asterix will be included. For example, intelligen* would encompass intelligent and intelligence.

If available, a filter was applied so that only peer-reviewed studies were included in the search. The search yielded 732 studies in total, 318 of which were found to be duplicates. The research listed on the Brainology[®] website (<http://www.mindsetworks.com/>) was also screened on 2nd January 2016 for any additional studies that could be included; however, only duplicates were found. This meant 414 studies were title screened adhering to the inclusion and exclusion criteria outlined in Table 2. A further 334 studies were removed during title screening, leaving 80 requiring abstract screening applying the same criteria. 70 studies were removed during abstract screening, leaving 10 that would have been appropriate for full-text screening. Despite all abstracts being available in English, three of these studies, however, were not published in English and were, therefore, removed from the review as full-text screening was not possible.

Table 2

Inclusion and Exclusion Criteria

	Inclusion Criteria	Exclusion Criteria	Rationale
Type of publication	Peer-reviewed research only	Non peer-reviewed research	Increased credibility and rigor of research process
Type of study	Study included an intervention	Study contained no intervention	Research question refers to intervention effectiveness
Psychological theory of intervention	Intervention refers to implicit theories and a growth mindset	Any type of intervention that does not directly refer to implicit theories or a growth mindset	Research question focussing on interventions stemming from this specific psychological theory only
Purpose of intervention	Intervention aims to promote a growth mindset in participants	Intervention aims to do anything other than promote a growth mindset in participants	Research question relates to intervention that promote a growth mindset
Outcome measures	At least one outcome measure of academic attainment	No outcome measure of academic attainment	Research question relates to raising pupil attainment
Language	Article published in English	Article published in any language other than English	Inability to accurately translate from original language
Setting	Intervention took place in a school	Intervention took place in a setting other than a school	Research question concerns effectiveness of school-based interventions
Location	Research took place in an Organisation for Economic Co-operation and Development	Research took place in a non-OECD country	Research more able to generalise to UK schools due to social, financial and educational

	Inclusion Criteria	Exclusion Criteria	Rationale
	(OECD) country		similarities.
Participants	Participants of school age (aged 4-17 at beginning of academic year in which intervention took place)	Participants not of school age (below 4 or above 18 at beginning of academic year in which intervention took place)	Research question concerns effectiveness of interventions for school-age children
Date	Published before 2 nd January 2016	Published after 2 nd January 2016	Final search date before analysis conducted

Seven studies were full-text screened using the inclusion and exclusion criteria. Three studies were removed at full-text screening and these are listed alongside the reason for exclusion in Appendix A. This left four articles suitable for inclusion in the review. It is important to note that only Study 2 from Blackwell et al. (2007) was included, and Yeager et al. (2014) included two studies that met the inclusion criteria (Study 2 and Study 3) and these were evaluated separately. This meant that five studies were included from four articles. A flow diagram of the systematic search process is shown in Figure 1. The final five studies are listed in Table 3 and summarised in Appendix B.

The studies were evaluated using the Weight of Evidence (WoE) Framework proposed by Harden and Gough (2012). Studies were first evaluated for WoE A, Methodological Quality. An adapted coding protocol for group-based designs from Kratochwill (2003) was used to evaluate the studies on a range of criteria and compare them using scores for seven main areas. The adaptations made to the coding

protocol are listed in Appendix C, and an example coding protocol can be found in Appendix D. The studies were then evaluated for WoE B, Methodological Relevance to the review question, and WoE C, Topic Relevance to the review question. Scores for each WoE were then averaged to provide an overall score, WoE D. Further details and the scores for each WoE are listed in Appendix E.

Figure 1

A Flow Diagram of the Literature Search

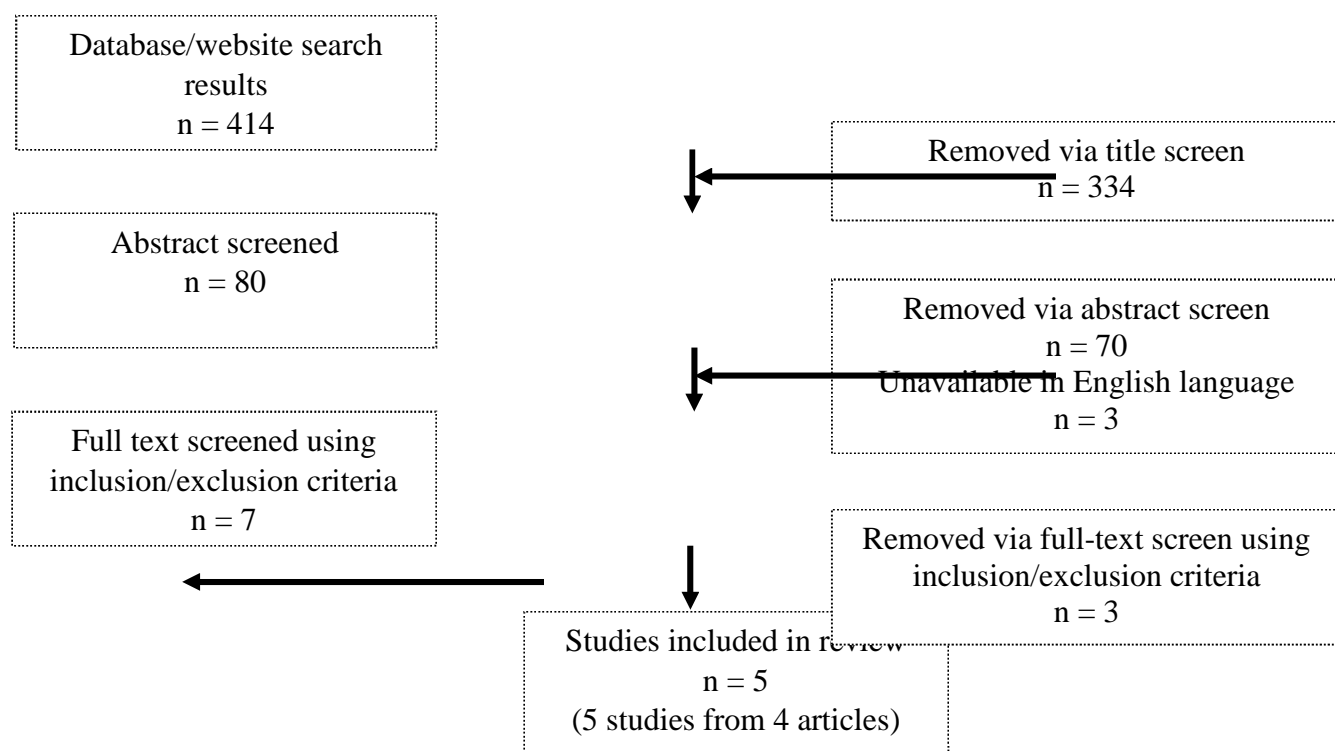


Table 3

List of Studies Included in the Review

Full Reference
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. <i>Child development, 78</i> (1), 246-263.

- Study 2

Full Reference

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.

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.

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

- Study 2
 - Study 3
-

Participants and Setting

The review included data from 2,051 students from the USA, ranging from 7th to 12th grade. Each of the studies drew samples from a single school, except for Paunesku et al. (2015), which sampled 13 schools. Demographics, such as gender, ethnicity and low socio-economic status (SES, as identified through free or reduced price school lunch) were reported for most schools and participants came from a range of backgrounds. The percentage of males and females was reported for all studies (for Yeager et al., 2014, see supplementary document) and was roughly equal. The percentage of pupils identified as low SES ranged from 0 to 90% (both in Paunesku et al., 2015). The ethnic majority of schools varied between studies: predominantly Hispanic/Latino (Good et al., 2003; Yeager et al., 2014, Study 3), African American (Blackwell, et al., 2007), Asian (school 1 in Paunesku et al., 2015) and White (multiple schools in Paunesku et al., 2015).

Clear information on demographics was incorporated into a study's score for WoE C. Pupils came from either public, charter or private schools and this is reflected in

scores for WoE A and WoE C, with studies that took place in a private school, or where information was unknown, given less weight than those that took place in state-funded schools. The majority of children who would require intervention attend state-funded schools (Kratochwill, 2003) and, therefore, private schools can be considered less representative of the target population.

Several of the studies targeted students identified as 'at-risk'. In both Yeager et al. (2014) Study 2 and 3, participants were taking a maths class that was considered lower ability than the class taken by most ninth grade pupils. In Blackwell et al. (2007), pupil's maths grades were at the 35th percentile nationally prior to the intervention, meaning they were relatively low-achieving. This is incorporated into scores for WoE C.

Attrition rates were low across studies (less than 20%, as described by Kratochwill, 2003), although a common feature was a lack of information on whether the mortality rate was equal across conditions. This was taken into account when scoring WoE A Comparison Group.

Design and Control Groups

All studies included a control group. For both Yeager et al. (2014) studies and Blackwell et al. (2007), the control group received a weaker form of the growth mindset intervention. In Blackwell et al. (2007), half the sessions received were the same for both the control and intervention group, with alternative sessions covering memory instead of a growth mindset. In the Yeager et al. (2014) studies, all participants received the same initial session. The experimental group then completed an activity aimed at promoting an incremental theory of personality, whereas the control group activity promoted an incremental theory of athletic ability.

In Good et al. (2003), the control groups received alternative interventions and one group received a combination of both an incremental and alternative intervention. In Paunesku et al. (2015), the control group received materials about a different aspect of neuroscience and other groups received either an alternative intervention or a combination of both the growth mindset and alternative intervention. The differences in control group between studies is factored into scores for WoE A Comparison Group and WoE B.

All studies randomly assigned participants to each group. Most studies included an intervention delivered at an individual level so this allowed for easy randomisation of participants. Blackwell et al. (2007) randomly assigned pre-existing classes to each condition, but pupils had been randomly allocated to these classes in the first place.

Intervention

The method and length of intervention varied between studies. The intervention in Good et al. (2003) consisted of two 90 minutes sessions, followed by weekly email communication with a mentor for two school semesters. Participants created a website for hypothetical students based on the information they were taught on the incremental theory of intelligence. Both Yeager et al. (2014) studies and Paunesku et al. (2015) adopted similar approaches. Participants first read material that promoted an incremental theory and then completed a written activity explaining this information to a hypothetical pupil. This activity lasted 25 minutes in both Yeager et al. (2014) studies and 45 minutes in Paunesku et al. (2015). The activity was completed on a computer in both Yeager et al. (2014) Study 3 and Paunesku et al. (2015). For the former of these, this was so that information could be translated into Spanish if necessary. Blackwell et al. (2007) delivered eight weekly 25 minute

sessions, four of which included reading, activities and discussions relating to an incremental theory of intelligence.

The studies that involved a weaker form of the intervention as a control group (Blackwell et al., 2007; Yeager et al., 2014, Study 2 and 3) allow the effectiveness of certain components of the intervention to be tested to some extent; however, all studies failed to adopt a design that allowed for all components of the intervention to be evaluated individually. For example, in Good et al. (2003), it is unknown whether the two initial sessions added any value to the intervention, or whether weekly correspondence for only one semester would have resulted in the same effect. This common flaw is reflected in WoE A Identifiable Components, for which all studies scored 0.

In terms of intervention fidelity, several of the interventions (Paunesku et al., 2015; Yeager et al., 2014, Study 2 and 3) were delivered through a computer programme or used predesigned materials, and thus can be considered manualised. Both Yeager et al. (2014) studies also used two coders to check whether participants had given a sufficient answer for the written activity. In Blackwell et al. (2007), the sessions were delivered by undergraduate assistants who received prereading and weekly training sessions prior to delivering each session and met for a debriefing afterwards. Participants were also tested on their knowledge of the intervention content after it had finished, and the intervention group received higher scores than the control. The mentors used in Good et al. (2003) received training on how to teach pupils the required message; however, little else is specified about how the intervention was monitored.

As a result of the manualisation and use of coders, both Yeager et al. (2014) studies

and Paunesku et al. (2015) scored higher on WoE A Implementation Fidelity than Blackwell et al. (2007) and Good et al. (2003). Manualised interventions provide a description of the procedure for implementing the intervention. This allows for more accurate replication of the intervention and, therefore, fidelity of implementation will be higher than an intervention not accompanied by descriptions of this type.

Measurement

Most studies used school grade as the attainment measure. Both Paunesku et al. (2015) and the Yeager et al. (2014) studies used a Grade Point Average (GPA) for all subjects considered 'core', which is factored into their WoE C score. Blackwell et al. (2007) only looked at mathematics grades, and Good et al. (2003) used end-of-year mathematics and reading scores assessed using the Texas Assessment of Academic Skills (TAAS), a standardised achievement test. Only Yeager et al. (2014) Study 2 reported a reliability coefficient, which resulted in the higher WoE A Measurement score.

All studies except Good et al. (2003) collected attainment scores at multiple time-points, enabling a comparison of pre and post intervention scores. Good et al. (2003) only collected grades post-intervention, and compared these to the control group. Most studies collected grades at one time-point post-intervention and, therefore, failed to include any follow-up. The exception to this is Yeager et al. (2014) Studies 2 and 3, which included grades for multiple time-points post-intervention; however, a separate follow-up was not included in the analysis. The lack of systematic follow-up in any study resulted in all receiving a score of 0 for WoE A Follow-up Assessment.

Outcomes and Effect Sizes

Blackwell et al. (2007) and Yeager et al. (2014) Study 2 were the only studies that found statistically significant improvements for their whole sample. It is important to note, however, that participants in both these studies were identified as relatively low-achieving for their age. The effect of intervention on grades for Blackwell et al. (2007) and Yeager et al. (2014) Study 2 was small to medium, and small respectively (Cohen, 1988). Good et al. (2003) found a medium effect on reading scores for all pupils who received the growth mindset intervention; for maths scores, however, a statistically significant effect was only found for females, which was large. Paunesku et al. (2015) found small effects only for those pupils identified as at risk of dropping out of school, which was measured by a low GPA or failure of at least one core subject. The statistical analysis, however, collapsed all interventions into one condition and this is reflected in the weighting criteria for WoE B and C. Yeager et al. (2014) Study 3 found a large effect on GPA but only for pupils who held an entity theory of intelligence prior to the intervention. Blackwell et al. (2007) tested whether pupils who originally held an entity theory benefitted more from the intervention; however, their results were not statistically significant ($p < .10$) and they suggest that this is due to a small sample size. Further information on effect sizes for statistically significant results can be found in Table 4.

Table 4

Effect Sizes and Descriptors for Statistically Significant Findings

Study	Measure	Comparison	Effect Size	Descriptor ^a	WoE D Descriptor
Blackwell et al. (2007) Study 2	Maths grade	Intervention vs control	$d = 0.45^b$	Small to medium	Medium

Study	Measure	Comparison	Effect Size	Descriptor ^a	WoE D Descriptor
Good et al. (2003)	TAAS Maths score	Incremental condition vs control for females only	$d = 1.13$	Large	Medium
		Combined condition vs control for females only	$d = 1.30$	Large	
	TAAS Reading score	Incremental condition vs control	$d = 0.52$	Medium	
Paunesku et al. (2015)	Grade Point Average	Any intervention vs control for students at risk of dropping-out of school	$d = 0.21^c$	Small	Medium
	Percentage of students who received satisfactory grades in all core subjects	Any intervention vs control for students at risk of dropping-out of school	$OR = 1.58$	Small	
Yeager et al. (2014) Study 2	Grade Point Average	Intervention vs control	$d = 0.34$	Small	Medium
	Percentage of students who received a D or below in Algebra I	Intervention vs control	$OR = 0.13$	Large	
Yeager et al. (2014) Study 3	Grade Point Average	Experimental vs control for baseline entity theorists	$d = 0.80$	Large	Medium
	Percentage of students who received a D or below in Algebra I	Experimental vs control for baseline entity theorists	$OR = 0.32$	Medium	

^a *d* descriptors from Cohen (1988) and *OR* descriptors from Rosenthal (1996).

^b The effect size was calculated by dividing the unstandardised coefficient by the standard deviation of grade score for the whole sample pre-intervention ($b = .53$, $SD = 1.19$).

^c The effect size was calculated by dividing 2 x the *t*-value by the square root of the degrees of freedom ($t = 2.30$, $df = 499$), as recommended in Kratochwill (2003).

Due to small sample sizes, all of the studies except for Paunesku et al. (2015) were underpowered for at least part of the statistical analysis. Many studies failed to report exact numbers for each intervention group, but the total number of participants was below that recommended by Cohen (1992) based on a medium effect size and alpha level of 0.05. Sample sizes became increasingly smaller as post-hoc analyses were carried out due to statistically significant interactions with mediating variables, such as gender (Good et al., 2003) and original theory of personality (Yeager et al., 2014, Study 3). In Good et al. (2003), the intervention did not have a statistically significant effect on maths scores for males ($p = .054$) but the effect size was medium ($d = .64$). A larger sample size may have led to statistically significant results for males as well as females. Whether a study had an adequate sample size is factored in to WoE A scores for Statistical Analysis and WoE B, with lower weight given to those deemed underpowered.

It is also important to note that four of the five studies list Carol Dweck as one of the authors. As the psychological theory behind the intervention was initially proposed by Carol Dweck, it is inevitable that much of the research would have been completed by herself and those associated with her work at Stanford University. As it is impossible to know the exact links between each of the authors, this factor was not incorporated into any WoE score. Associated implications, however, are highlighted in the review conclusion.

Conclusions and Recommendations

The purpose of the review was to establish whether interventions that promote a growth mindset improve students' academic attainment. All studies found that some participants benefited from a growth mindset intervention for at least one measure of academic attainment; however, for some studies, the benefits were isolated to a specific group of students. Paunesku et al. (2015) found that interventions were only effective for pupils with a low GPA or who had failed at least one core subject. In relation to this, all participants in both Blackwell et al. (2007) and the Yeager et al. (2014) studies were described as low achieving students. It can, therefore, be concluded that the majority of evidence into growth mindset interventions suggests that they are effective at raising attainment for pupils at risk of educational disadvantage.

The only study that found the intervention effective for generic students was Good et al. (2003), which suggested that the intervention was not as effective for males. This result was potentially due to the small sample size and should, therefore, be treated with caution. Yeager et al. (2014) Study 3 found that the intervention was only successful in improving attainment for those pupils who held an entity theory or fixed mindset before the intervention. This result is in line with psychological theory (Dweck & Leggett, 1988) as these pupils would have more scope to change than pupils who already hold an incremental theory. Blackwell et al. (2007) also lent some weight towards the idea that interventions may be more effective for these pupils; however, their findings were not statistically significant. A small sample was used so further research is needed to establish whether interventions should predominantly target this group of students.

Effect sizes were largest for Good et al. (2003), which implemented the longest intervention; however, a limitation of all studies was the lack of separate analysis for individual intervention components. No conclusion can be drawn, therefore, as to whether the length or duration of an intervention increases its effectiveness.

Relating this to practice, the evidence suggests that interventions of this type can improve attainment for pupils at risk of educational disadvantage, and potentially more so for students who hold a fixed mindset. This means that it is worthwhile for practitioners to establish students' current implicit beliefs prior to implementing this intervention. As the evidence is inconclusive on which intervention components are deemed necessary or more effective than others, it can be left up to the judgement of educators as to which method of intervention they choose to adopt. It is important, however, that any intervention is carefully monitored against intended outcomes, as the evidence-base behind it contains some methodological weaknesses.

Areas for Further Research

Firstly, there are issues around the generalisability of findings. No studies were found that included pupils younger than seventh grade; therefore, implementing a growth mindset intervention with pupils younger than 12 years old could not be considered evidence-based. Further research is needed with this younger age group to establish whether this type of intervention would be effective for these pupils.

It is also important to highlight that all studies took place within the USA and generalising findings to the UK education system should be done with caution.

Methods for teacher training will differ between countries and it is possible students within the UK may not hold exactly the same beliefs about intelligence as students in

the USA. Again, further research is needed to establish whether the growth mindset interventions are equally as effective outside of the USA.

Secondly, there were limitations in the design of many studies that would need to be addressed in future research to make the evidence-base more solid. A common theme within studies was insufficient sample size; therefore, higher powered studies with more participants are necessary before any confident conclusions can be drawn. Studies also failed to evaluate individual components of interventions and future research should consider whether the length or method of delivery of an intervention can impact its effectiveness. The lack of follow-up in any study was also highlighted in this review. Many studies targeted at-risk pupils and it is vital for educators to know whether any short-term advantage for these pupils is maintained as they progress through education.

Finally, further studies conducted by researchers not associated with Carol Dweck would add more weight to the current evidence-base. As Carol Dweck has developed commercially available fixed mindset interventions, and written many books on the subject, she can be seen to make financial gains from promoting the effectiveness of this type of intervention. It is possible that the researchers may have been subject to experimenter bias as they would have had a vested interest in finding favourable results. In relation to this, establishing whether any publication bias is present in the research literature would be useful, as it is possible that only research that yielded positive results was submitted or accepted for publication.

References

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38(2), 113-125.

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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Department for Education and Department for Health (DfE & DH). (2015). Special educational needs and disability code of practice: 0 to 25 years. London: The Stationery Office. (revised January 2015). See www.education.gov.uk

Diener, C. I., & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy and achievement cognitions following failure. *Journal of Personality and Social Psychology, 36*(5), 451-462.

Diener, C. I., & Dweck, C. S. (1980). An analysis of learned helplessness: II. The processing of success. *Journal of Personality and Social Psychology, 39*(5), 940-952.

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

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273.

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.

Harden, A., & Gough, D. (2012). Quality and relevance appraisal. In Gough, D., Oliver, S., & Thomas, J. (Eds). *An Introduction to Systematic Reviews* (pp. 153-179). London: Sage Publications Ltd.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Kratochwill, T. R. (2003). Task Force on Evidence-Based Interventions in School Psychology. Retrieved January 20, 2016, from http://www.indiana.edu/~ebi/documents/_workingfiles/EBImanual1.pdf

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

Robins, R. W., & Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity, 1*(4), 313-336.

Rosenthal, J. (1996). Qualitative Descriptors of Strength of Association and Effect Size, *Journal of Social Service Research*, 21(4), 37-59.

Spitzer, B., & Aronson, J. (2015). Minding and mending the gap: Social psychological interventions to reduce educational disparities. *British Journal of Educational Psychology*, 85(1), 1-18.

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

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

Appendix A: List of Studies Excluded at Full Text Screening

Study	Reason for Exclusion
Moreno, J., González-Cutre, D., Martín-Albo,	Outcome measure is performance

<p>J., & Cervelló, E. (2010). Motivation and performance in physical education: An experimental test. <i>Journal of Sports Science & Medicine</i>, 9(1), 79-85.</p>	<p>in a lateral movement test in physical education, not a measure of pupil attainment.</p>
<p>Cutts, Q., Cutts, E., Draper, S., O'Donnell, P., & Saffrey, P. (2010, March). Manipulating mindset to positively influence introductory programming performance. In <i>Proceedings of the 41st ACM technical symposium on Computer science education</i> (pp. 431-435). ACM.</p>	<p>Participants are University students and not of school age.</p>
<p>Pepi, A., Alesi, M., & Geraci, M. (2004). Theories of intelligence in children with reading disabilities: a training proposal 1. <i>Psychological reports</i>, 95(3), 949-952.</p>	<p>Intervention was metacognitive and not related to promoting a growth mindset.</p>

Appendix B: Summary of Included Studies

Study	Location	Sample	Design	Intervention	Relevant Measures	Primary Outcomes and Findings
Good, Aronson and Inzlicht (2003)	USA (Texas)	138 seventh grade pupils 70% free or reduced-price school lunch. 67% Hispanic, 20% White, 13% Black.	Experimental with post measure only. Random assignment to one of four conditions (incremental teaching, attribution teaching, incremental/attribution combined, control) at individual student level.	Two 90 minute face-to-face sessions with mentor and weekly email communication throughout school year to assist pupil in developing a website conveying intervention message.	Texas Assessment of Academic Skills (TAAS) in reading and maths taken at the end of the school year.	Maths: statistically significant large effects for intervention (incremental $d = 1.13$, combined $d = 1.30$) compared to control for females. Results were not statistically significant for males (incremental $p = .054$, $d = .64$). Reading: statistically significant effects for incremental ($d = .52$) but not for combined.
Blackwell et al. (2007) Study 2	USA (New York City)	91 seventh grade pupils Intervention: 48 Control: 43 79% free school lunch.	Experimental with two pre measures and one post measure. Preassigned groups randomly allocated to control or	Eight weekly 25 minute sessions held during pupil advisory session already scheduled. Control group had session 1, 2, 5 and 6 same as	Mathematics grades at three time points over a year. Intervention delivered between 2 nd and 3 rd time	Statistically significant effect of intervention on change in grades between time 2 and time 3 ($b = .53$). Maths grades declined between points 2 and 3 for control but not for intervention

Study	Location	Sample	Design	Intervention	Relevant Measures	Primary Outcomes and Findings
		52% African American, 45% Latino, 3% White and Asian	intervention condition.	intervention.	points.	group.
Paunesku et al. (2015)	USA (east, west and southwest)	1,594 students from 13 high schools	Experimental with pre and post measures of GPA, but only post measure used for analysis. Random assignment to one of four conditions (growth mindset intervention, sense-of-purpose intervention, combined interventions, control) at individual student level.	One 45 minute online session involving reading an article and two writing exercises (summarising the online session and applying content to hypothetical pupil).	GPA in core academic subjects for fall (pre-intervention) and spring (post). Percentage of satisfactory grades (A, B, C or Pass).	Statistically significant interaction between risk (students at risk of dropping out of high school) and intervention for growth mindset intervention and nonsignificant ($p = .071$) for combined intervention. Statistically significant effect of all interventions collapsed together for students at risk of dropping out of high school but not for other students. 'At-risk' students in the intervention conditions (all combined) had higher percentage of

Study	Location	Sample	Design	Intervention	Relevant Measures	Primary Outcomes and Findings
						satisfactory grades than at-risk students in the control group (49% vs 41%).
Yeager et al. (2014) Study 2	USA (Nothern Carolina)	78 lower performing ninth grade students from one high school. 1% free/reduced price school lunch 44% White, 44% Asian American, 12% Latino	Experimental. Condition randomly assigned at individual student level.	Overview of how the brain changes given to both groups. 25 minute activity completed 2 weeks later supporting people’s potential to change – pupils read an article and quotes, then wrote their own narrative to share with future students.	End and intermediate semester GPA (6 scores, 3 for each semester) for English, math and science. Percentage of students who received a D or below in Algebra I	Statistically significant effect of intervention on GPA ($d = .34$). Grades declined more for control group than intervention group. Less students received a D or below in the intervention group compared to the control group (14% vs 2%, $OR = .13$).
Yeager et al. (2014) Study 3	USA (California)	150 ninth grade students from one low-performing high school. 73% free or	Experimental – close replication of Study 2. Condition randomly assigned at	Overview of how the brain changes given to both groups. 25 minute computer-based activity completed 2 weeks later (with	End of semester GPA (2 scores, 1 for each semester) for English, math	Statistically significant effect on grades for students who held an entity theory prior to the intervention ($d = .80$). Nonsignificant effect for those that held an

Study	Location	Sample	Design	Intervention	Relevant Measures	Primary Outcomes and Findings
		reduced price school lunch. 77% Hispanic/Latino, 10% White, 5% Black/African American, 8% Other	individual student level.	Spanish option) supporting people’s potential to change – pupils read an article and quotes, then wrote their own narrative to share with future students.	and science. Percentage of students who received a D or below in Algebra I	incremental theory prior to intervention. For students who held an entity theory, less students received a D or below in the intervention group compared to the control group (42% vs 19%, $OR = .32$).

Appendix C: Items Removed from Coding Protocol

The following outlines the sections removed from the Kratochwill (2003) Coding Protocol for Group Based Designs, along with the rationale for removal.

Section Removed	Rationale
B7 Coding and B8 Interactive process	No study included qualitative data.
C Primary/Secondary Outcomes Are Statistically Significant	Only outcomes that measure pupil attainment are relevant to the review question. Effect sizes for these outcomes were calculated if statistically significant and reported in the review.
D Educational/Clinical Significance	Participants were not from a clinical sample and one study only included post-intervention data.
E1 Evidence for primary outcomes	Score taken from section 'C Primary/Secondary Outcomes Are Statistically Significant', which was removed.
G Replication	Any replication of a study would have been treated as an additional study and analysed separately relating to the review question.
H2 Site of Implementation: Non School Site	All interventions took place in a school setting, as per inclusion criteria.
G Intervention Style or Orientation	All interventions focussed on promoting an incremental theory or growth mindset, as per the inclusion criteria.

Appendix D: Example Coding Protocol

Study 1: Blackwell et al. (2007)

Coding Protocol: Group-Based Design

Domain:

- School- and community-based intervention programs for social and behavioral problems
- Academic intervention programs
- Family and parent intervention programs
- School-wide and classroom-based programs
- Comprehensive and coordinated school health services

Name of Coder(s): _____ Date: 30-01-2016

Full Study Reference in APA format:

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.

Intervention Name (description from study): Incremental theory training

Study ID Number (Unique Identifier): 1

Type of Publication: (Check one)

- Book/Monograph
- Journal article
- Book chapter
- Other (specify):

I. General Characteristics

A. General Design Characteristics

- A1. Random assignment designs (if random assignment design, select one of the following)
 - A1.1 Completely randomized design
 - A1.2 Randomized block design (between-subjects variation)
 - A1.3 Randomized block design (within-subjects variation)
 - A1.4 Randomized hierarchical design
- A2. Nonrandomized designs (if nonrandom assignment design, select one of the following)
 - A2.1 Nonrandomized design
 - A2.2 Nonrandomized block design (between-participants variation)
 - A2.3 Nonrandomized block design (within-participants variation)
 - A2.4 Nonrandomized hierarchical design
 - A2.5 Optional coding of Quasi-experimental designs (see Appendix C)

A3. Overall confidence of judgment on how participants were assigned (select one of the following)

- A3.1 Very low (little basis)
- A3.2 Low (guess)
- A3.3 Moderate (weak inference)
- A3.4 High (strong inference)
- A3.5 Very high (explicitly stated)
- A3.6 N/A
- A3.7 Unknown/unable to code

B. Statistical Treatment/Data Analysis (answer B1 through B6)

- B1. Appropriate unit of analysis yes no
- B2. Familywise error rate controlled yes no N/A
- B3. Sufficiently large *N* yes no
 - Statistical Test: Regression growth curve with knot point and t-test comparisons.
 - Alpha level: 0.05
 - ES: Medium
 - N* required: 67
- B4. Total size of sample (start of the study): 99 completed questionnaire, 95 participated in intervention
- B5. Intervention group sample size: 48
- B6. Control group sample size: 43

C. Type of Program (select one)

- C1. Universal prevention program
- C2. Selective prevention program
- C3. Targeted prevention program
- C4. Intervention/Treatment
- C5. Unknown

D. Stage of the Program (select one)

- D1. Model/demonstration programs
- D2. Early stage programs
- D3. Established/institutionalized programs
- D4. Unknown

E. Concurrent or Historical Intervention Exposure (select one)

- E1. Current exposure
- E2. Prior exposure
- E3. Unknown

II. Key Features for Coding Studies and Rating Level of Evidence/ Support
(3=Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence)

A. Measurement (answer A1 through A4)

A1. Use of outcome measures that produce reliable scores for the majority of primary outcomes. The table for Primary/Secondary Outcomes Statistically Significant allows for listing separate outcomes and will facilitate decision making regarding measurement (select one of the following)

- A1.1 Yes
- A1.2 No
- A1.3 Unknown/unable to code

A2. Multi-method (select one of the following)

- A2.1 Yes
- A2.2 No
- A2.3 N/A
- A2.4 Unknown/unable to code

A3. Multi-source (select one of the following)

- A3.1 Yes
- A3.2 No
- A3.3 N/A
- A3.4 Unknown/unable to code

A4. Validity of measures reported (select one of the following)

- A5.1 Yes validated with specific target group
- A5.2 In part, validated for general population only
- A5.3 No
- A5.4 Unknown/unable to code

Rating for Measurement (select 0, 1, 2, or 3): 3 2 1 0

B. Comparison Group

B1. Type of Comparison Group (select one of the following)

- B1.1 Typical contact
- B1.2 Typical contact (other) specify:
- B1.3 Attention placebo
- B1.4 Intervention elements placebo
- B1.5 Alternative intervention
- B1.6 Pharmacotherapy
- B1.7 No intervention
- B1.8 Wait list/delayed intervention
- B1.9 Minimal contact
- B1.10 Unable to identify comparison group

Rating for Comparison Group (select 0, 1, 2, or 3): 3 2 1 0

B2. Overall confidence rating in judgment of type of comparison group (select one of the following)

- B2.1 Very low (little basis)
- B2.2 Low (guess)
- B2.3 Moderate (weak inference)
- B2.4 High (strong inference)
- B2.5 Very high (explicitly stated)
- B2.6 Unknown/Unable to code

B3. Counterbalancing of Change Agents (answer B3.1 to B3.3)

- B3.1 By change agent
- B3.2 Statistical: for maths class not intervention tutor
- B3.3. Other:

B4. Group Equivalence Established (select one of the following)

- B4.1 Random assignment
- B4.2 Posthoc matched set
- B4.3 Statistical matching
- B4.4 Post hoc test for group equivalence

B5. Equivalent Mortality (answer B5.1 through B5.3)

- B5.1 Low Attrition (less than 20% for Post)
- B5.2 Low Attrition (less than 30% for follow-up)
- B5.3 Intent to intervene analysis carried out
Findings _____

~~C. Primary/Secondary Outcomes Are Statistically Significant~~

~~C1. Evidence of appropriate statistical analysis for **primary outcomes** (answer C1.1 through C1.3)~~

- ~~C1.1 Appropriate unit of analysis (rate from previous code)~~
- ~~C1.2 Familywise/experimenterwise error rate controlled when applicable (rate from previous code)~~
- ~~C1.3 Sufficiently large *N* (rate from previous code)~~

~~C2. Percentage of **primary outcomes** that are significant (select one of the following)~~

- ~~C2.1 Significant primary outcomes for at least 75% of the total primary outcome measures for each key construct~~
- ~~C2.2 Significant primary outcomes for between 50% and 74% of the total primary outcome measures for each key construct~~
- ~~C2.3 Significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct~~

Rating for Primary Outcomes Statistically Significant (select 0, 1, 2, or 3): 3 2 1 0

~~C3. Evidence of appropriate statistical analysis for **secondary outcomes** (answer C3.1 through C3.3)~~

- ~~C3.1 Appropriate unit of analysis~~
- ~~C3.2 Familywise/experimenterwise error rate controlled when applicable (rate from previous code)~~
- ~~C3.3 Sufficiently large *N* (rate from previous code)~~

~~C4. Percentage of **secondary outcomes** that are significant (select one of the following)~~

- ~~C4.1 Significant secondary outcomes for at least 75% of the total secondary outcome measures for each key construct~~
- ~~C4.2 Significant secondary outcomes for between 50% and 74% of the total secondary outcome measures for each key construct~~
- ~~C4.3 Significant secondary outcomes for between 25% and 49% of the total secondary outcome measures for any key construct~~

~~**Rating for Secondary Outcomes Statistically Significant** (select 0, 1, 2, or 3): 3
2 1 0~~

~~C5. Overall Summary of Questions Investigated~~

~~C5.1 Main effect analyses conducted (select one) yes no~~

~~C5.2 Moderator effect analyses conducted (select one) yes no~~

~~Specify results:~~

~~C5.3. Mediator analyses conducted (select one) yes no~~

~~Specify results:~~

C. Primary/Secondary Outcomes Statistically Significant (only list $p \leq .05$)
 (list primary outcomes first in alphabetical order, followed by secondary outcomes in alphabetical order)

Outcomes	Primary vs. Secondary	Who Changed	What Changed	Source	Treatment Information	Outcome Measure Used	Reliability	ES	(1-)
Outcome #1:	<input type="checkbox"/> Primary <input type="checkbox"/> Secondary <input type="checkbox"/> Unknown	<input type="checkbox"/> Child <input type="checkbox"/> Teacher <input type="checkbox"/> Parent/sign. adult <input type="checkbox"/> Ecology <input type="checkbox"/> Other <input type="checkbox"/> Unknown	<input type="checkbox"/> Behaviour <input type="checkbox"/> Attitude <input type="checkbox"/> Knowledge <input type="checkbox"/> Other <input type="checkbox"/> Unknown	<input type="checkbox"/> Self Report <input type="checkbox"/> Parent Report <input type="checkbox"/> Teacher Report <input type="checkbox"/> Observation <input type="checkbox"/> Test <input type="checkbox"/> Other <input type="checkbox"/> Unknown					

Null Findings/Negative Outcomes Associated with the Intervention (listed alphabetically by outcome)

Outcomes	Primary vs. Secondary	Who Was Targeted for Change	What Was Targeted for Change	Source	Note null/negative outcomes	Outcome Measure Used	Reliability	ES
Outcome #1:	<input type="checkbox"/> Primary <input type="checkbox"/> Secondary <input type="checkbox"/> Unknown	<input type="checkbox"/> Child <input type="checkbox"/> Teacher <input type="checkbox"/> Parent/sign. adult <input type="checkbox"/> Ecology <input type="checkbox"/> Other <input type="checkbox"/> Unknown	<input type="checkbox"/> Behaviour <input type="checkbox"/> Attitude <input type="checkbox"/> Knowledge <input type="checkbox"/> Other <input type="checkbox"/> Unknown	<input type="checkbox"/> Self Report <input type="checkbox"/> Parent Report <input type="checkbox"/> Teacher Report <input type="checkbox"/> Observation <input type="checkbox"/> Test <input type="checkbox"/> Other <input type="checkbox"/> Unknown				

Type of Data Effect Size is Based On

- (check all that apply)
- Means and SDs
 - t*-value or *F*-value
 - Chi-square (*df* = 1)
 - Frequencies or proportions (dichotomous)
 - Frequencies or proportions (polytomous)
 - Other (specify):
 - Unknown

Confidence Rating in ES Computation

- (select one of the following)
- Highly estimated (e.g., only have *N* *p*-value)
 - Moderate estimation (e.g., have complex but complete statistics)
 - Some estimation (e.g., unconventional statistics that require conversion)
 - Slight estimation (e.g., use significance testing statistics rather than descriptives)
 - No estimation (e.g., all descriptive data is present)

D. Educational/Clinical Significance

Outcome Variables:	Pretest	Posttest	Follow-Up
D1. Categorical Diagnosis Data	Diagnostic information regarding inclusion into the study presented: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Positive change in diagnostic criteria from pre to posttest: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Positive change in diagnostic criteria from posttest to follow-up: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
D2. Outcome Assessed via continuous Variables		Positive change in percentage of participants showing clinical improvement from pre to posttest: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Positive change in percentage of participants showing clinical improvement from posttest to follow-up: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
D3. Subjective Evaluation: The importance of behaviour change is evaluated by individuals in direct contact with the participant.	Importance of behavior change is evaluated: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Importance of behavior change from pre to posttest is evaluated positively by individuals in direct contact with the participant: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Importance of behavior change from posttest to follow-up is evaluated positively by individuals in direct contact with the participant: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
D4. Social Comparison: Behavior of participant at pre, post, and follow-up is compared to normative data (e.g., a typical peer).	Participant's behavior is compared to normative data <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Participant's behavior has improved from pre to posttest when compared to normative data: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Participant's behavior has improved from posttest to follow-up when compared to normative data: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown

Rating for Educational/Clinical Significance (select 0, 1, 2, or 3): 3 2 1 0

E. Identifiable Components (answer E1 through E7)

- ~~E1. Evidence for primary outcomes (rate from previous code): 3 2 1 0~~
- E2. Design allows for analysis of identifiable components (select one) yes no
- E3. Total number of components: Unknown
- E4. Number of components linked to primary outcomes: Unknown
- Additional criteria to code descriptively:
- E5. Clear documentation of essential components (select one) yes no
- E6. Procedures for adapting the intervention are described in detail (select one) yes no
- E7. Contextual features of the intervention are documented (select one) yes no

Rating for Identifiable Components (select 0, 1, 2, or 3): 3 2 1 0

F. Implementation Fidelity

- F1. Evidence of Acceptable Adherence (answer F1.1 through F1.3)
 - F1.1 Ongoing supervision/consultation
 - F1.2 Coding intervention sessions/lessons or procedures
 - F1.3 Audio/video tape implementation (select F1.3.1 or F1.3.2):
 - F1.3.1 Entire intervention
 - F1.3.2 Part of intervention
- F2. Manualization (select all that apply)
 - F2.1 Written material involving a detailed account of the exact procedures and the sequence in which they are to be used
 - F2.2 Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used
 - F2.3 Written material involving an overview of broad principles and a description of the intervention phases
 - F2.4 Formal or informal training session involving an overview of broad principles and a description of the intervention phases
- F3. Adaptation procedures are specified (select one) yes no unknown

Rating for Implementation Fidelity (select 0, 1, 2, or 3): 3 2 1 0

~~**G. Replication** (answer G1, G2, G3, and G4)~~

- ~~G1. Same Intervention~~
- ~~G2. Same Target Problem~~
- ~~G3. Independent evaluation~~

~~**Rating for Replication** (select 0, 1, 2, or 3): 3 2 1 0~~

H. Site of Implementation

- H1. School (if school is the site, select one of the following options)

- H1.1 Public
- H1.2 Private
- H1.3 Charter
- H1.4 University Affiliated
- H1.5 Alternative
- H1.6 Not specified/unknown

~~H2. Non School Site (if it is a non school site, select one of the following options)~~

- H2.1 Home
- H2.2 University Clinic
- H2.3 Summer Program
- H2.4 Outpatient Hospital
- H2.5 Partial inpatient/day Intervention Program
- H2.6 Inpatient Hospital
- H2.7 Private Practice
- H2.8 Mental Health Center
- H2.9 Residential Treatment Facility
- H2.10 Other (specify): _____
- H2.11 Unknown/insufficient information provided

Rating for Site of Implementation (select 0, 1, 2, or 3): 3 2 1 0

I. Follow Up Assessment

- Timing of follow up assessment: N/A
- Number of participants included in the follow up assessment: N/A
- Consistency of assessment method used: N/A

Rating for Follow Up Assessment (select 0, 1, 2, or 3): 3 2 1 0

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

- A1. Sampling procedures described in detail yes no
 Specify rationale for selection: Opportunity sample
 Specify rationale for sample size: None
- A1.1 Inclusion/exclusion criteria specified yes no
- A1.2 Inclusion/exclusion criteria similar to school practice yes no
- A1.3 Specified criteria related to concern yes no
- A2. Participant Characteristics Specified for all groups

Particip ants from all Groups	Gra de/ age	Gend er	Eth nicit y or Mult ieth nic	Ethn ic Iden tity	Race(s)	Accult uration	Pri - mary Lang uage	SE S	Fa mily Str uct ure	Loc ale	Disabil ity	Func tional Descri ptors
Child/St	Gra	49	Unk	Unk	52%	Unkno	Unkn	79	Unk	Urb	Unkno	Low

Participants from all Groups	Grade/age	Gender	Ethnicity or Multiethnic	Ethnicity	Race(s)	Acculturation	Primary Language	SES	Family Structure	Locale	Disability	Functional Descriptors
udent	de 7	femal e 50 male	now n	now n	African American, 45% Latino, 3% White and Asian	wn	own	% free lunch	no wn	an	wn	achieving – maths scores at 35 th percentile

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes yes no

Specify: Gender, Low SES

A3.2 Have relevance to inclusion criteria yes no

A4. Receptivity/acceptance by target participant population (treatment group)

Teacher rated classroom-motivation change as positive or negative. 27% positive in intervention compared to 9% control

A5. Generalization of Effects:

A5.1 Generalization over time

A5.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated yes no

Specify: _____

A5.1.2 Procedures for maintaining outcomes are specified yes no

Specify: _____

A5.2 Generalization across settings

A5.2.1 Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context yes no

Specify: _____

A5.2.2 Documentation of efforts to ensure application of intervention to other settings yes no

Specify: _____

A5.2.3 Impact on implementers or context is sustained yes no

Specify: _____

A5.3 Generalization across persons

Evidence is provided regarding the degree to which outcomes are manifested with participants who are different than the original group of participants for with the intervention was evaluated yes no

Specify: _____

B. Length of Intervention (select B1 or B2)

B1. Unknown/insufficient information provided

- B2. Information provided (if information is provided, specify one of the following:)
 - B2.1 weeks: 8
 - B2.2 months
 - B2.3 years
 - B2.4 other

C. Intensity/dosage of Intervention (select C1 or C2)

- C1. Unknown/insufficient information provided
- C2. Information provided (if information is provided, specify both of the following:)
 - C2.1 length of intervention session: 25 minutes
 - C2.2 frequency of intervention session: Weekly

D. Dosage Response (select D1 or D2)

- D1. Unknown/insufficient information provided
- D2. Information provided (if information is provided, answer D2.1)
 - D2.1 Describe positive outcomes associated with higher dosage:

E. Program Implementer (select all that apply)

- E1. Research Staff
- E2. School Specialty Staff
- E3. Teachers
- E4. Educational Assistants
- E5. Parents
- E6. College Students
- E7. Peers
- E8. Other: Undergraduate assistants
- E9. Unknown/insufficient information provided

F. Characteristics of the Intervener

- F1. Highly similar to target participants on key variables (e.g., race, gender, SES)
- F2. Somewhat similar to target participants on key variables: One male, one female, one African American, one Hispanic
- F3. Different from target participants on key variables

G. Intervention Style or Orientation (select all that apply)

- G1. Behavioral
- G2. Cognitive-behavioral
- G3. Experiential
- G4. Humanistic/interpersonal
- G5. Psychodynamic/insight-oriented
- G6. other (specify):
- G7. Unknown/insufficient information provided

H. Cost Analysis Data (select G1 or G2)

- H1. Unknown/insufficient information provided

H2. Information provided (if information is provided, answer H2.1)

H2.1 Estimated Cost of Implementation: _____

I. Training and Support Resources (select all that apply)

I1. Simple orientation given to change agents

I2. Training workshops conducted

of Workshops provided: Unclear (states weekly so likely 8)

Average length of training: Unknown

Who conducted training (select all that apply)

I2.1 Project Director

I2.2 Graduate/project assistants

I2.3 Other (please specify):

I2.3 Unknown

I3. Ongoing technical support

I4. Program materials obtained

I5. Special Facilities

I6. Other (specify):

J. Feasibility

J1. Level of difficulty in training intervention agents (select one of the following)

J1.1 High

J1.2 Moderate

J1.3 Low

J1.4 Unknown

J2. Cost to train intervention agents (specify if known): _____

J3. Rating of cost to train intervention agents (select one of the following)

J3.1 High

J3.2 Moderate

J3.3 Low

J3.4 Unknown

Summary of Evidence for Group-Based Design Studies

Indicator	Overall Evidence Rating NNR = No numerical rating or 0 - 3	Description of Evidence Strong Promising Weak No/limited evidence or Descriptive ratings
General Characteristics		
General Design Characteristics	NNR	
Statistical Treatment	NNR	
Type of Program	NNR	
Stage of Program	NNR	
Concurrent/Historical Intervention Exposure	NNR	
Key Features		
Measurement	0	No/limited evidence
Comparison Group	2	Promising
Primary/Secondary Outcomes are Statistically Significant		
Educational/Clinical Significance		
Identifiable Components	0	No/limited evidence
Implementation Fidelity	1	Weak
Replication		
Site of Implementation	3	Strong
Follow Up Assessment Conducted	0	No/limited evidence
Descriptive or Supplemental Criteria		
External validity indicators	NNR	
Length of Intervention	NNR	
Intensity/dosage	NNR	
Dosage Response	NNR	
Program Implementer	NNR	
Characteristics of the Intervener	NNR	
Intervention		

Style/Orientation		
Cost Analysis Data Provided	NNR	
Training and Support Resources	NNR	
Feasibility	NNR	Low

Appendix E: Weight of Evidence

Weight of Evidence A: Methodological Quality

WoE A refers to the quality of the study compared to other studies using the same design. For this review, items were assessed using the coding manual and amended coding protocol for group-based designs from Kratochwill (2003).

The scores for Measurement, Comparison Group, Identifiable Components, Implementation Fidelity, Site of Implementation and Follow-up Assessment were taken directly from the Kratochwill (2003) coding protocol. Despite all studies scoring 0 for Identifiable Components and Follow-up Assessment, these sections were still included in WoE A so that any newly published study or future review can be evaluated using the same criteria. As the section 'Primary/Secondary Outcomes Are Statistically Significant' was removed from the coding protocol, the criteria used for Statistical Analysis score is outlined below.

Weighting	Criteria ^a
High (3)	Statistical analysis includes all of the following: <ul style="list-style-type: none"> • Appropriate unit of analysis • Familywise error rate controlled • Sufficiently large sample size
Medium (2)	Statistical analysis includes two of the following: <ul style="list-style-type: none"> • Appropriate unit of analysis • Familywise error rate controlled • Sufficiently large sample size
Low (1)	Statistical analysis includes one of the following: <ul style="list-style-type: none"> • Appropriate unit of analysis • Familywise error rate controlled • Sufficiently large sample size
Zero (0)	Statistical analysis includes none of the following: <ul style="list-style-type: none"> • Appropriate unit of analysis • Familywise error rate controlled • Sufficiently large sample size

^a Sufficient sample size was calculated from Cohen (1992) based on a medium effect size and alpha level of 0.05. A medium effect size was chosen as no other effect size has been suggested for this intervention and no prior meta-analysis has been conducted relating to this topic.

The average WoE A score is converted to a qualitative descriptor based on the follow ranges:

- High = above 2.5
- Medium = 1.5 – 2.49
- Low = 0.5 – 1.49
- Zero = less than 0.5

Study	Statistical Analysis	Measurement	Comparison Group	Identifiable Components	Implementation Fidelity	Site of Implementation	Follow-up Assessment	Average	Descriptor
Blackwell et al. (2007) Study 2	2	0	2	0	1	3	0	1.14	Low
Good et al. (2003)	1	0	1	0	1	2	0	0.71	Low
Paunesku et al. (2015)	2	0	2	0	2	3	0	1.29	Low
Yeager et al. (2014) Study 2	2	1	2	0	2	2	0	1.29	Low
Yeager et al. (2014) Study 3	1	0	2	0	2	2	0	1	Low

Weight of Evidence B: Methodological Relevance

WoE B refers to the appropriateness of the study design for answering the review question. In this case, whether an intervention is effective at improving academic attainment. For a study to receive a rating, all criteria for that weighting must be met.

Weighting	Criteria
High (3)	<ul style="list-style-type: none"> • Study includes an ‘active’ control group that receives a different type of intervention. • Participants are randomly allocated to condition group. • Study includes a comparison of pre and post scores for relevant attainment measure. • Statistical analysis compares effect of growth mindset intervention to other intervention. • Sample size is adequate for all statistical analyses.
Medium (2)	<ul style="list-style-type: none"> • Study includes a ‘no intervention’ control group, e.g. wait list. • Participants are randomly allocated to condition group or group equivalence established through posthoc analysis. • Study includes a comparison of post scores for relevant attainment measure between intervention and control. • Statistical analysis compares effect of intervention to control condition (but may collapse interventions to one condition if no difference found between them). • Sample size may be lower than required for statistical analyses.
Low (1)	<ul style="list-style-type: none"> • Study does not include a control group. • Study includes a comparison of pre and post scores for relevant attainment measure. • Sample size may be lower than required for statistical analyses.
Zero (0)	<ul style="list-style-type: none"> • Study meets none of the above criteria.

Rationale for WoE B Criteria:

- An ‘active’ control group allows for an analysis of whether the intervention is more effective than other similar interventions.
- Random allocation of participants to group removes any bias that could arise from differences in participants between groups.
- The inclusion of pre and post scores allows for a) statistical analysis of pre-scores to establish any difference between group prior to the intervention, and b) an analysis of the change in scores as a result of the intervention.
- In order to establish whether the growth mindset intervention was effective, the statistical analysis must compare the effect of this intervention to alternative interventions.

- The sample size chosen for the study based on intended statistical analysis must be considered sufficient using Cohen (1992) based on a medium effect size and alpha level of 0.05.

Weight of Evidence C: Topic Relevance

WoE C refers to the relevance of the focus of the study to the specific review question. For a study to receive a rating, all criteria for that weighting must be met.

Weighting	Criteria
High (3)	<ul style="list-style-type: none"> • Participants are sampled from more than one school. • Most participants attend a state-funded school. • At least two of the following demographics (age, gender, measure of SES and ethnicity) are provided for both participants and the local area and the sample is deemed representative. • Attainment measure covers all 'core' subjects (mathematics, English and science). • There is evidence that the attainment measure used has high reliability and validity ($r = .70$ or higher) (Kratochwill, 2003). • Intervention is universal and targets all students. • Results from growth mindset intervention analysed separately from any other intervention.
Medium (2)	<ul style="list-style-type: none"> • Participants are sampled from only one school. • Most participants attend a privately funded school or type of school is unknown. • At least two of the following demographics (age, gender, measure of SES and ethnicity) are provided for the sample only. • Attainment measure covers at least two different subjects. • No statistical information on reliability of validity of attainment measure is provided. • Intervention targets a subset of students identified as 'at-risk'.
Low (1)	<ul style="list-style-type: none"> • Demographics (age, gender, measure of SES and ethnicity) of participants is unknown. • Attainment measure covers only one subject.
Zero (0)	<ul style="list-style-type: none"> • Study meets none of the above criteria.

Rationale for WoE C Criteria:

- Findings can be generalised more widely if the sample covers multiple schools and schools that are state-funded as opposed to privately funded.

- Detailed information on demographics also allows further generalisation if a wide range of demographics are covered and the sample is deemed representative of the population.
- The research question refers to pupil attainment; therefore, a measure that covers all core subjects would be considered a more relevant, and results more generalizable, than a measure for just one subject.
- The use of reliable and valid measures allows for greater confidence that any change would be reliable and the appropriate construct has been targeted.
- The research question refers to school-based interventions for raising attainment generally and does not relate to a specific population of students.
- In order to establish whether the growth mindset intervention was effective, the statistical analysis must compare the effect of this intervention to alternative interventions.

Weight of Evidence D: Overall

The WoE D score is an average score for WoE A, WoE B and WoE C. This score is converted to a qualitative descriptor based on the follow ranges:

- High = above 2.5
- Medium = 1.5 – 2.49
- Low = 0.5 – 1.49
- Zero = less than 0.5

Study	WoE A Methodological Quality	WoE B Methodological Relevance	WoE C Topic Relevance	WoE D Overall
Blackwell et al. (2007) Study 2	Low (1.14)	High (3)	Low (1)	Medium (1.71)
Good et al. (2003)	Low (0.71)	Medium (2)	Medium (2)	Medium (1.57)
Paunesku et al. (2015)	Low (1.29)	Medium (2)	Medium (2)	Medium (1.76)
Yeager et al. (2014) Study 2	Low (1.29)	High (3)	Medium (2)	Medium (2.10)
Yeager et al. (2014) Study 3	Low (1)	Medium (2)	Medium (2)	Medium (1.67)