Case Study 1: An Evidence-Based Practice Review Report

Theme: School (setting) based interventions for children with special educational needs (SEN)

Is visualisation training effective in supporting school age children with comprehension difficulties?

Section 1: Summary

Training in visualisation, the ability to create mental images in an individual's mind, has been used for some time as an intervention to support reading and listening comprehension with a range of pupils. This review aims to evaluate the effectiveness of visualisation training interventions in school settings with a particular focus on supporting pupils with identified difficulties in the specific area of comprehension. A systematic review of the literature was undertaken using three online data bases. Six studies were selected for review and appraised using Gough's (2007) Weight of Evidence (WoE) Framework and an adapted Weight of Evidence A Coding Protocol from Gersten et al. (2005). Findings of the studies were examined and all but one of the studies found that visualisation had a positive effect on comprehension. The wide range of intervention delivery procedures and outcome measures was discussed as a limiting factor when drawing firm conclusions from this group of studies and areas for further investigation were identified. This review concluded that visualisation training interventions are an accessible and resource-friendly intervention option. Commonalities across the studies indicate that interventions involving an element of visualisation training could help to support pupils struggling with comprehension.

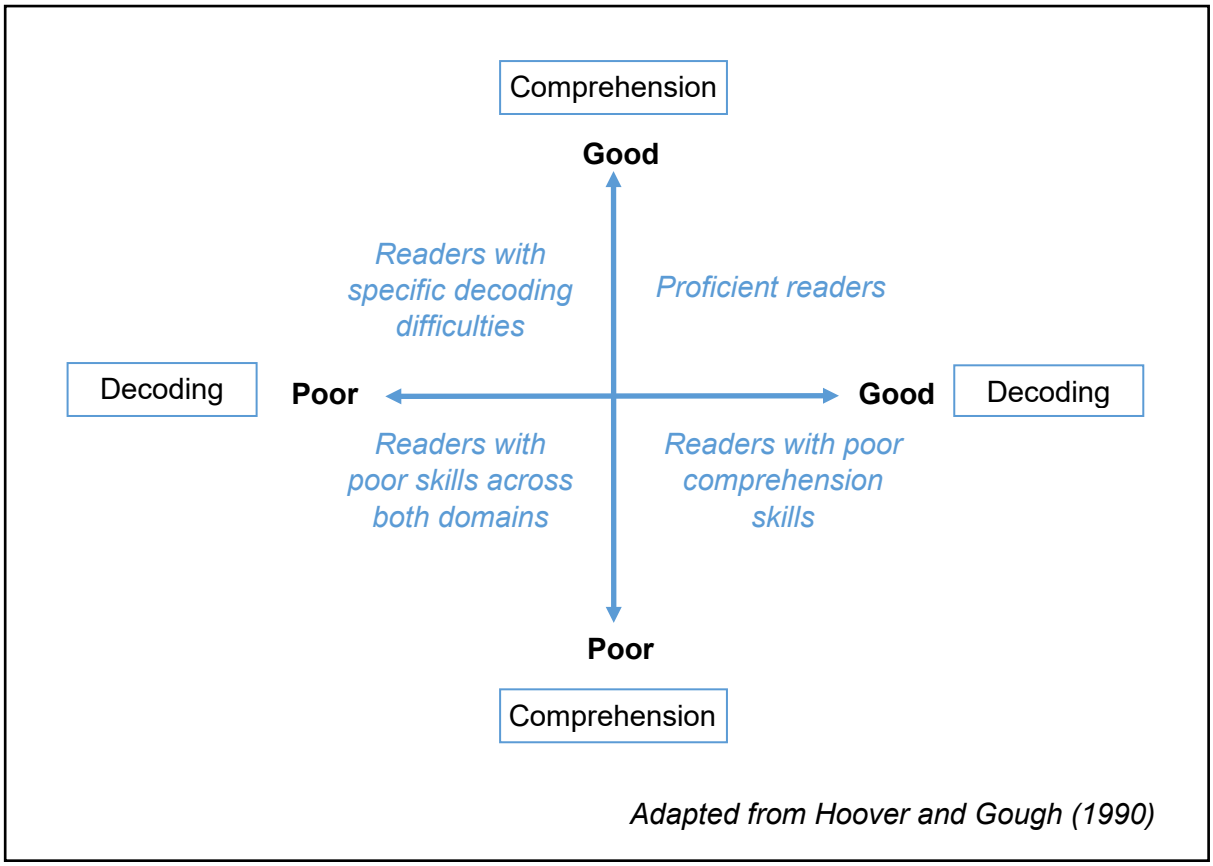
Section 2: Introduction

Reading Comprehension

The Simple View of Reading (Hoover & Gough, 1990) suggests that reading skills can be separated into two domains: decoding - concerning the reading of the words on the page and comprehension - the skill to understand the language of what is being read. Figure 1 illustrates the interaction of comprehension and decoding skills.

Figure 1

The Simple View of Reading and associated characteristics of each quadrant



Adapted from Hoover and Gough (1990)

Strong skills in both domains will lead to success in reading. Clearly the two domains depend to an extent on each other; a child will need to be able to decode a text before they can begin to understand it. However, it is not enough to assume that once decoding skills are mastered or any difficulties are addressed, comprehension will automatically flourish. Longitudinal studies of comprehension indicate that decoding is not necessarily a reliable predictor of comprehension skills (E.g. Oakhill & Cain, 2012).

Difficulties in each domain have different characteristics. Struggling to decode at an age appropriate level despite solid instruction and well targeted intervention could be an indicator of a specific literacy difficulty (Rose, 2009). There are a number of factors which may contribute to the development of such difficulties both environmental and at the level of individual differences. A genetic influence which predisposes some children to these difficulties, often described as Dyslexia, is also well documented (Pennington & Olson, 2005). However, arguably less well-established is the concept of a separate genetic basis to comprehension difficulties. Keenan et al. (2006) describe a genetic influence on comprehension skills independent of that on word decoding skills. Children with a profile such as this are often referred to as 'poor comprehenders'. As with all reading difficulties environmental influences such as exposure to language and educational experience play a large role in the development of comprehension skills (e.g. Perfetti et al., 2005).

As children develop reading comprehension they create a detailed mental representation of the text as a cohesive whole (Van Dijk & Kintsch, 1983). The term 'situation model' has been used to describe this representation (Van Dijk & Kintsch, 1983) and typically, poor comprehenders will struggle to achieve this. A possible

reason is that the development of a situation model depends on using inference to make links between the immediate information from the text and prior knowledge – either from an earlier piece of information from the same text or from general knowledge (Oakhill & Patel, 1991). Educators have often considered the ability to infer meaning as a desirable product of comprehension. However, it is increasingly recognised that the ability to infer information is one of the skills critical to underpinning comprehension itself and therefore difficulties inferring meaning will contribute to poor comprehension skills (Kintsch & Rawson, 2005). Similarly, integrative processing skills are stronger in good comprehenders than poor comprehenders (Oakhill et al., 1986) and therefore supporting children to develop comprehension skills should aim to address this area.

The Role of Visualisation in Comprehension

One way of helping poor comprehenders integrate information is through encouraging the use of visualisation / mental imagery. Of note is the use of different terminology in the field with researchers using the terms ‘visualisation’, ‘mental imagery’ and ‘imagery. For the purposes of this review these terms will be used interchangeably. Oakhill and Patel (1991) suggested that by visualising the information in their minds, pupils’ ability to integrate information was automatically enhanced. The idea that generating a ‘picture in the mind’ can assist with understanding a text is not a new one. In 1976, Pressley found that mental imagery training did help children to record text after only a single twenty-minute session - an effect which has since been replicated. For example, Oakhill and Patel (1991) found an increase in inference generation and Gambrell and Bales (1986) concluded that inconsistencies in text were better spotted by pupils who had undergone

visualisation training than those who had not. Oakhill and Yuill (1996) grouped such factors together and proposed that there are three aspects central to successful comprehension. In addition to inference making, they argued that an understanding of text structure and the ability to monitor comprehension are all aspects which promote comprehension. The ability to generate visualisations supports each of these three areas.

Other researchers in the field have focussed on the role visualisation plays in combination with other skills. The dual coding concept of visualisation and verbalisation was proposed by Bell in 1986 and is based on earlier work by Paivio (1971) which proposed a dual coding system for reading. Paivio's Dual Coding Theory centres on the concept that information is processed in one of two ways: verbal and auditory processing of language and visual processing of nonverbal information such as images. Bell's Visualisation and Verbalisation programme for supporting reading comprehension is based largely on this concept and includes approaches from each of the two domains (Bell, 1991).

Visualisation Training

Some children spontaneously use visualisation when reading (Sadoski,1985). However, it is important to consider whether this is something that children can be trained to do, and if so, which methods might be most effective. Studies in this area employ a wide range of different methods for visualisation instruction from single one-off training sessions of short duration (e.g. Pressley, 1976) to longitudinal studies where regular interventions are established (e.g. Rader, 2009).

Reading Comprehension and Listening Comprehension

Once decoding skills have been mastered, reading comprehension shares skills with the comprehension of spoken language and the correlation between understanding written text and a verbally delivered equivalent is high (Perfetti et al., 2005). Methodological considerations regarding the assessment of comprehension skills in written and spoken language tasks cannot be ignored. The delivery of a spoken text will vary between assessors and between occasions, recordings can be unfamiliar and distracting; these are factors which need to be considered when assessing comprehension. For the purposes of this review, it was decided that evaluating the impact of visualisation training on both verbal and reading comprehension was relevant.

The ability to read and understand a wide range of texts underpins much of the current school curriculum. In addition, it could be argued that reading for pleasure should be seen as an end in itself. When faced with the challenge of supporting poor comprehenders, being able to answer the question of what works to improve reading comprehension is key and this is where educational psychologists as scientist practitioners can offer guidance. With limited human resources and time, educators need to employ strategies that have been shown to have a positive impact on the difficulties faced by struggling readers. Educational psychologists can offer a valuable insight into what strategies can be employed to promote children's reading comprehension skills. Therefore, examining whether visualisation training can be counted as one such strategy is relevant to educational psychology practice.

Review Question

This review aims to answer the question, 'Is visualisation training effective in supporting school age children with comprehension difficulties?'

Section 3: Critical Review of the Evidence

Literature Search

A systematic literature search was conducted on 16th January 2022 using three online databases: Educational Resources Information Center (ERIC Ebsco), Psycinfo and Web of Science. The search terms used are listed in Table 1. Searches were limited to peer-reviewed journal articles published after 1994 to ensure quality and relevance to the current education system. The year 1994 was chosen as a cut off to reflect the reforms to the national curriculum in the UK undertaken in 1994 following the Dearing Review of the curriculum and assessment system (Dearing, 1993) which resulted in changes to the national curriculum and statutory assessments, many of which are reflected in the current situation in schools. Search results were limited to articles published in English to aid understanding by the author. Ancestral searches were conducted on articles selected for inclusion in the review.

Table 1

Search Terms Used in the Searches of the Databases

Search Terms Used	Rationale

visualisation OR visualization OR imagery OR "mental imagery" OR "guided imagery"	This review seeks to evaluate the effectiveness of visualisation instruction. The terms 'mental imagery' or 'guided imagery' are related to visualisation and were noted during pilot searches.
intervention OR program* OR training OR strateg*	This review is concerned with evaluating the use of visualisation as a strategy taught through intervention.
"reading comprehension"	The effect of visualisation on reading comprehension skills in the focus of this review.

Note. Concepts were combined with 'AND'. Truncation (*) was used to include any ending of root words. Speech marks (") were used to include exact phrases.

Article Screening

Database searches yielded a total of 103 studies which were initially screened by title and then abstract against the inclusion and exclusion criteria detailed in Table 2 and during this process 5 additional studies were identified from ancestral searching. The remaining 14 studies were screened after reading the full article texts (See Appendix A). The flow chart in Figure 1 gives details of the search process and the numbers of studies excluded at each stage. Following this process, a total of 6 studies were identified for inclusion in this review. These studies are listed in Table 3.

Table 2

Inclusion and Exclusion Criteria with Rationale

Criterion	Inclusion	Exclusion	Rationale

1.	Intervention	The article includes the use of visualisation instruction as an intervention.	No use of visualisation instruction or visualisation is used only in combination.	This review is concerned with effectiveness of visualisation interventions.
2.	Design	Conclusions are drawn from primary data derived from quantitative studies.	The article is a review or meta-analysis or draws conclusions solely from qualitative data.	Original research conducted within an experimental design to allow the examination of effects.
3.	Outcome measures	At least one outcome measure relates to reading comprehension.	No outcome measures of reading comprehension are included.	Outcome measures are needed to evaluate the impact of the intervention on reading comprehension.
4.	Setting	The research is conducted in a school or other educational setting.	The research is not conducted in a school or other educational setting, for example at home or in a clinic.	This review is concerned with the impact of school or other education setting-based interventions.
5.	Age of participants	Participants are of compulsory school age in the UK (5-16)	Participants' ages fall outside of compulsory school age in the UK.	This review aims to evaluate visualisation instruction for children of

				compulsory school age.
6.	Reading ability level of participants	Participants have been identified as having difficulties with reading comprehension.	Participants have no difficulties with reading comprehension.	This review is concerned with effectiveness of intervention for poor comprehenders.

Figure 2

Flow diagram showing literature search and screening strategy.

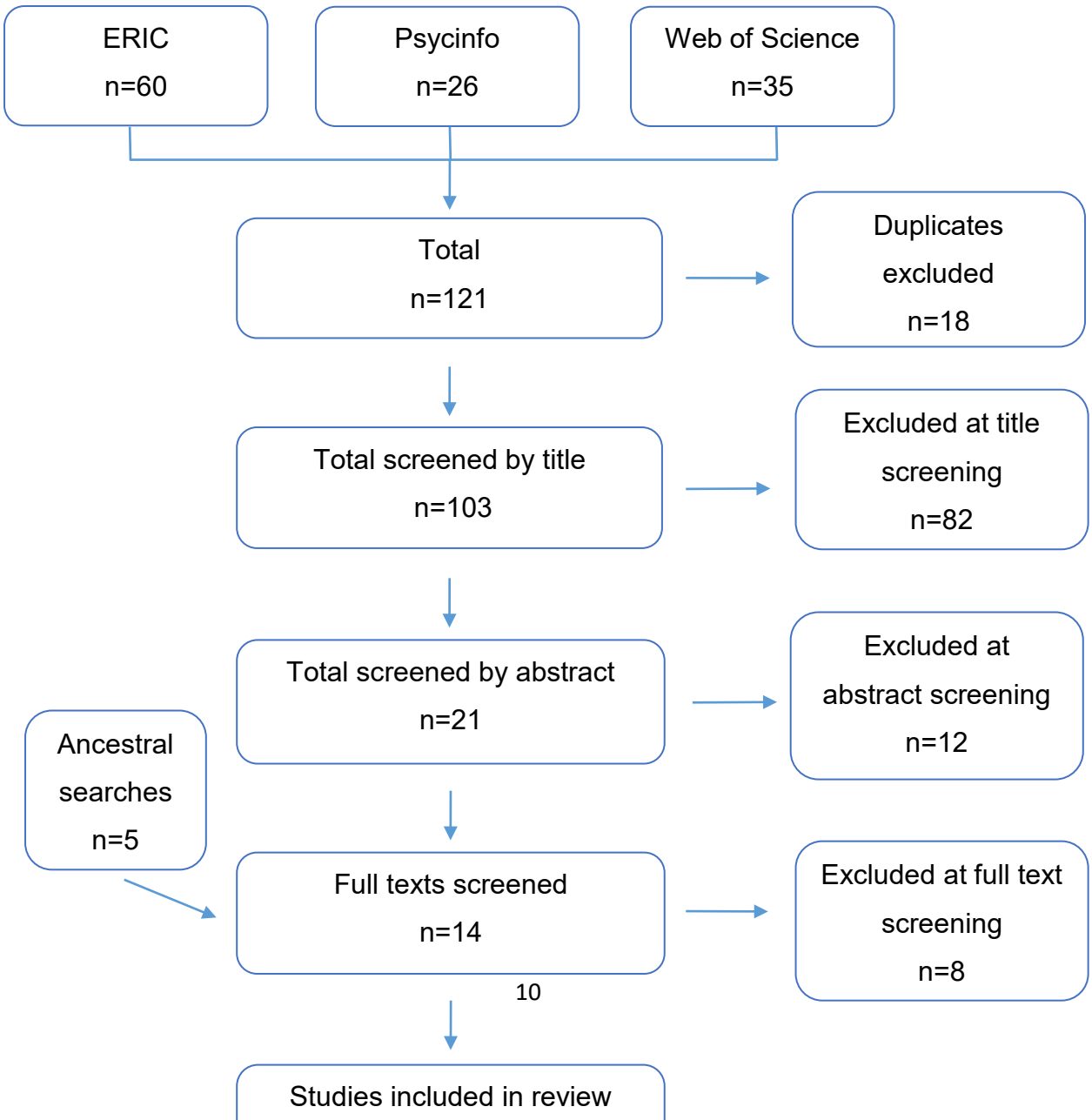


Table 3

References of Studies Included in the Review

1	Cariglia-Bull, T., & Pressley, M. (1990). Short-term memory differences between children predict imagery effects when sentences are read. *Journal of Experimental Child Psychology*, *49*(3), 384-398.
2	Center, Y., Freeman, L., Robertson, G., & Outhred, L. (1999). The effect of visual imagery training on the reading and listening comprehension of low listening comprehenders in year 2. *Journal of Research in Reading*, *22*(3), 241-256.
3	Francey, G., & Cain, K. (2015). Effect of imagery training on children's comprehension of pronouns. *The Journal of Educational Research*, *108*(1), 1-9.
4	Joffe, V. L., Cain, K., & Marić, N. (2007). Comprehension problems in children with specific language impairment: does mental imagery training help? *International Journal of Language & Communication Disorders*, *42*(6), 648-664.
5	Johnson-Glenberg, M. C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: Verbal versus visual strategies. *Journal of Educational Psychology*, *92*(4), 772.

6	Rader, L. A. (2009). Teaching students to visualize: Nine key questions for success. *Preventing School Failure: Alternative Education for Children and Youth*, *54*(2), 126-132.

Weight of Evidence

A Weight of Evidence (WoE) framework (Gough, 2007) was used to critically appraise the 6 studies selected for this review. Each study was individually scored against criteria for methodological quality (WoE A), methodological relevance (WoE B), and topic relevance for this review (WoE C). Scores across each of the categories were then combined and an average taken to arrive at an overall WoE rating, WoE D. A summary of the WoE ratings for the studies can be seen in Table 4.

Table 4

A summary of the WoE ratings for the studies in the review

Study	WoE A	WoE B	WoE C	WoE D
Cariglia-Bull & Pressley (1990)	2 Medium	2.3 Medium	1.7 Medium	2 Medium
Center et al. (1999)	3 High	2.3 Medium	2.6 High	2.6 High
Francey & Cain (2015)	3 High	2 Medium	1.7 Medium	2.2 Medium
Joffe et al. (2007)	3 High	2 Medium	1.7 Medium	2.2 Medium
Johnson-Glenberg (2000)	3 High	2.3 Medium	1.7 Medium	2.3 Medium
Rader (2009)	3 High	2 Medium	2.3 Medium	2.4 Medium

Note. ≥ 2.5 = High, 1.5 -2.5 = Medium, < 1.5 = Low

In order to appraise studies against Wo E A criteria a coding protocol (Gersten et al., 2005) was used which allowed a detailed evaluation of design aspects including quality indicators for describing participants; implementation of the intervention; description of comparison conditions; outcome measures and data analysis. Minor adaptations for clarity and relevance were made by the author, see appendix B.

A summary of the WoE A judgements for each study can be seen in Table 5.

Table 5

Summary of the WoE A judgements for the 6 studies

Study	Essential Criteria Category				Total Essential & score	Total Desirable & score	Overall WoE A & rating
	Participant Description	Implementation / Comparison	Outcome measures	Data Analysis	>10 = 2 9 - 10 = 1 <9 = 0	>4 = 2 2 - 4 = 1 <2 = 0	≥3 = High 2 = Medium <2 = Low
Cariglia-Bull & Pressley (1990)	4	2	1	1	8 Score = 0	5 Score = 2	2 Medium
Center et al. (1999)	5	3	2	2	12 Score = 1	5 Score = 2	3 High
Francey & Cain (2015)	4	2	2	2	10 Score = 1	5 Score = 2	3 High
Joffe et al. (2007)	3	3	2	2	10 Score = 1	6 Score = 2	3 High
Johnson-Glenberg (2000)	4	3	2	2	11 Score = 2	6 Score = 2	4 High
Rader (2009)	4	3	2	1	10 Score = 1	5 Score = 2	3 High

A Coding Protocol used to arrive at judgements for WoE B was developed by the author of this review and concerned the relevance and quality of the methodology used in the studies examined. Three categories: 'Study Design', 'Comparison Cohort' and 'Outcome Measures' were chosen to assess each study. The rating procedure, rationale and details of the protocol can be found in appendix B.

A WoE C coding protocol was also developed by the author and concerned 'Comprehension Measure, Visualisation Training and SEN / Poor Comprehender Status. These categories were selected to evaluate the level of relevance to the review question. Details of this protocol are also included in appendix B.

Appendix C contains completed protocols for WoE A, B and C for each of the categories against which the studies were appraised.

Mapping the Field

The six studies included in this review used quantitative designs which aimed to assess the effect of visualisation training on the comprehension skills of intervention versus comparison groups. Details of study designs, participants and measures are provided in Table 6.

Table 6

Overview of Information about the Design, Participants and Measures of the Studies for Review

Study	Study Type and Design	Location	Participants	Intervention Investigated	Context of Intervention	Outcome Variables
Cariglia-Bull & Pressley (1990)	Randomised Control Trial (balanced for gender) Dependent pre-test and post-test with untreated control group	Middle to high socio-economic area of medium sized city in Canada	165 participants Age: 9 to 12 years All groups: range of abilities, not necessarily SEN or poor comprehenders	One session only: instruction to use mental imagery	Students asked to recall sentences after being assessed for reading comprehension and short-term memory ability.	Pre-test and post test scores collected. Elements of sentences recalled, scores per minute Correlation with short term memory and reading comprehension
Center et al. (1999)	Quasi Experimental Design: random	Suburbs of Sydney Australia	66 participants Age: 7 years	Visualisation training by trained instructor	Listening comprehension lessons of 20	Pre-test and post test scores. Correlation of

	allocation of matched pairs from a selected group Dependent pre and post-test with untreated control group		Intervention group: 33 poor comprehenders (participants scored lowest third on test of comprehension) Control: 33 adequate comprehenders		minutes 3 x week for four weeks for both groups. Intervention group lessons included visualisation training	both reading comprehension scores and listening comprehension scores between matched pairs.
Francey & Cain (2015)	Quasi Experimental Design: Dependent pre and post-test with contrasting same intervention group	North West England Mix of rural and urban schools	34 participants Age: 9 to 10 years Intervention: 17 poor listening comprehenders Comparison: 17 good listening comprehenders	Mental imagery training involving coloured images at first and then a requirement to generate own images	3 x 30 min imagery training on consecutive days. Participants exposed to sentences and stories	Identification of pronoun antecedents and pronouns in cloze text converted to scores. Assessed on understanding of pronouns after imagery intervention

Joffe et al. (2007)	Quasi Experimental Design: Dependent pre and post-test with untreated control group	Multicultural, multilingual London borough school	25 participants Age: 7 to 11 years Intervention group: 9 SLI[a] RB pupils Control group: 16 TD[b] peers	Mental imagery training	Before and after 5 x 30 min sessions of mental imagery training, participants listened to short stories.	Responses to verbal questions concerning literal and inferential aspects of the stories were scored.
Johnson-Glenberg (2000)	Quasi Experimental Design: Dependent pre and post-test with comparison group and untreated control group	95% white American students from Colorado. Very few students received free school meals	45 participants Age: 8 to 11 years Intervention group: 23 Contrast training group: 22 Control: 14 All: poor comprehenders / adequate decoders	Visualisation / verbal training	Pre – tests of 11 different measures of comprehension were followed by 28 x 30 min sessions over 10 weeks of training in visualising / verbalizing. Post test scores of the	Pre and post test scores of 11 different measures of comprehension were analysed and compared to a reciprocal teaching training and a without instruction control group.

			(teacher identification)		same measures were taken.	
Rader (2009)	Quasi Experimental Design: Dependent pre and post-test with untreated control group	Urban elementary school setting with mostly Hispanic and White participants of average socioeconomic backgrounds. Groups in separate schools	69 participants Age: 6 to 8 years Intervention 2 x groups of 33 and Control 2x groups of 36 All: speech and language needs and / or risk of reading failure	Visualisation and verbal training centred around 9 questions	Weekly intervention lessons over 2 years delivered by specifically trained teachers using a script.	Pre and post test scores concerning features of comprehension and recall were analysed and compared to scores from control groups

Note. [a] Speech and Language Impairment, [b] Typically Developing Peers

Participants

Across the 6 studies, a total of 404 participants took part in the studies in this review with an age range from 6 to 12 years. Participants were all in mainstream primary schools with the intervention group in one study (Joffe et al., 2007) recruiting participants to the intervention group from a resource base for pupils with speech and language impairment. Sample sizes varied from 25 to 165 with the smallest number in an intervention group of 9 (Joffe et al., 2007). Attrition was low across all the studies and where there was attrition it was mentioned in the studies.

All but one of the studies (Cariglia-Bull & Pressley, 1990) recruited participants with identified comprehension difficulties, a factor which gave this study a lower WoE C rating. Methods for identifying participants varied between the studies as two of the studies, Center et al. (1999) and Francey and Cain (2015), used measures to assess pupils for their comprehension ability prior to selection / intervention whereas Johnson-Glenberg (2000) relied on teacher identification. This was not considered a reason to reduce the WoE ratings for this study as it was judged that teachers who know the children well will be able to use their professional knowledge to identify pupils against these criteria. Two of the studies recruited participants with speech and language impairment (Joffe et al., 2007 and Rader, 2009) The study from Rader (2009) did not achieve a high score in WoE C ratings as participants were identified as either having speech and language needs and / or being at risk of reading failure. It was decided that this represented a wide and diverse group which could affect measures of outcomes in unpredictable ways. The inclusion of pupils with speech and language impairment was not heavily penalised in the WoE ratings as this type of difficulty is typified by difficulties in comprehension; however, it is possible that pupils in the speech and language resource base in one study (Joffe et al., 2007)

may have had a primary need involving difficulties with speech production rather than a receptive language difficulty. Therefore, the WoE C rating for this study was limited to 'medium'.

Research Design

One of the studies (Cariglia-Bull & Pressley, 1990) was a randomised controlled trial (balanced for gender). The remaining studies reviewed were quasi experimental designs with either untreated control groups (Center et al.,1990; Joffe et al., 2007; Rader, 2009) or comparison control groups (Francey & Cain, 2015; Johnson-Glenberg, 2000) which received an alternative intervention of equivalent length and intensity. The nature of the control or comparison group was highly relevant to this review. It was judged that comparing with a non-intervention group (Cariglia-Bull & Pressley, 1990; Center et al., 1990, Joffe et al., 2007; Rader, 2009) would not allow for isolation of the effects of the visualisation intervention. Positive outcomes could be associated with extra attention or small group work as described by the 'Hawthorne Effect' (Sedgwick & Greenwood, 2015). Therefore, studies with a comparison cohort experiencing an alternative intervention or attention (Francey & Cain, 2015) were scored more highly than those who experienced no intervention for this measure on WoE B. An additional consideration is whether the comparison groups had a similar literacy or general ability level to the intervention group. This was considered important as individual differences in reading comprehension levels could account for variations in impact. Studies in which all the participants had a similar level of comprehension skill (Cariglia-Bull & Pressley, 1990); Center et al., 1999; Johnson-Glenberg, 2000; Rader, 2009) were scored more highly than those in

which there was an identified difference in reading comprehension skill level between groups (Francey & Cain, 2015; Joffe et al., 2007).

Intervention

There was considerable heterogeneity between the interventions used across the 6 studies. Only one study (Rader, 2009) used the participants' own teachers to deliver the interventions. In order to maintain fidelity, the author of this study provided the teachers with training and scripts in order to deliver the sessions. The control groups were parallel classes who received no intervention, however one of the classes did show improvement due to a teacher running their own unrelated intervention.

The length of the studies varied between one session (Cariglia-Bull & Pressley, 1990) and regular sessions over two years (Rader, 2009). The remaining studies all involved several sessions delivered over a period of 3 days (Francey & Cain, 2015) to 10 weeks (Johnson-Glenberg, 2000). This makes it difficult to draw comparisons between the studies and is a limitation of this review.

The procedures and content of the visualisation training interventions differed across studies. Center et al., (1999) used physical objects as well as pictures to train participants to generate mental images. One study (Johnson-Glenberg, 2000) used an established programme of reading intervention, Bell's 1986 Visualising and verbalising for language comprehension and thinking intervention (Bell, 1991). In this programme, visualisation skills are built up progressively starting with one word and leading, eventually to a longer paragraph text over sequential lessons. In this programme, visualisation is combined with a focus on verbalisation. A low rating for

the Visualisation Training category in WoE C reflects this combined approach and the lack of clarity regarding which element of the training produced any benefits seen. Joffe et al., 2007 used a mental imagery training programme devised by Gambrell and Bales (Gambrell & Bales, 1986) to structure their intervention. The advantage of this is that the training programme has been used previously with some prior analysis of its effects, enabling the study to engage in a degree of replication of previous findings. (Joffe et al., 2007). The remaining studies devised their own visualisation interventions but all involved visual stimuli such as pictures to encourage pupils to generate mental images in response to text.

Outcome Measures

While five of the studies measure the effects of intervention on general reading comprehension, Francey and Cain (2015) were concerned with the effect of visualisation on pronoun use, due to the reliance on integration of information in a reading or listening task as well as inference skills that this measure requires. In this study a further level of analysis relating to the distance between antecedent information and pronoun use was also measured, with the hypothesis that the greater the distance, the more demanding the task in relation to comprehension.

When selecting outcome measures for pre and post-tests, the authors overcame the potential challenges of reliability and validity in different ways and to different degrees. Cariglia-Bull and Pressley (1990) used a standardised reading comprehension test, namely the Gates-MacGinitie test. They chose the test as none of the pupils in the district had previous exposure therefore the test was likely to be more reliable. A review of this test by Cooter (Cooter, 1989) found that although this

test had been normed on a large sample with representation from pupils with a range of backgrounds in order to develop reliability, the subject of validity was not well addressed in the development of this test. Francey and Cain (2015) attempted to improve reliability and validity by conducting two standardised measures when grouping pupils into good and poor comprehender categories before measuring impact of the intervention with their own devised measure of pronoun comprehension which was tailored to the task to ensure a higher degree of validity.

Rader (2009) also used a measure created specifically for the study by the author. As Rader was concerned with measuring retell ability, a passage was created and skills practised in isolation before administering and scoring via a rubric. This attention to detail aimed to increase the validity and reliability of this non-standardised measure.

Center et al. (1999) used a range of standardised tests in combination with a story structure retelling test. To ensure validity for the poorer comprehenders the researchers adapted the test so that the stories were read to them. This ensured that the test was measuring comprehension rather than any difficulties with decoding skills as it was noted that these students had lower skills in this area.

Joffe et al. (2007) used a story comprehension measure devised by Bishop and Adams (1992) which the authors selected due to this measure having previously been used successfully with pupils with language difficulties.

Johnson-Glenberg (2000) used a total of eleven different standardised measures to create a detailed picture of pre and post test performance. These included: WRAT (Wide Range Achievement Test) Word Recognition; Gates-MacGinitie comprehension; DTLA (Detroit Test of Learning Aptitude); WISC (Wechsler

Intelligence Scale for Children) including digit Span; visual measures, visual imagery-paired word (imagery phase only) and visual open-ended questions.

Findings

Table 7

Findings across the studies in the review. Effect sizes are included as reported in each study where available. Where there were multiple outcome measures, results from the most relevant to the review question were selected.

Study	Sample Size	Outcome Measure	Significance	Effect size	WoE D
Cariglia-Bull & Pressley (1990)	165	Sentences recalled per minute of reading. Between group comparison.	t(163) =1.30,p>0.10 (No significant effect)	Not given	2 Medium
Center et al. (1999)	66	Neale Reading Comprehension Test. Between group comparison	F=4.66,p<0.035 (Significant effect)	η_p^2 0.44	2.6 High
Francey & Cain (2015)	34	Mean effects on cloze test pre and post test with near and far pronoun antecedents. Within group comparison	Near antecedent: P<0.0125 (Significant effect) Far antecedent:	19% increase in scores	2.2 Medium

			No information about significance given	11% increase in scores	
Joffe et al. (2007)	25	Pre and post test of literal and inferential comprehension questions. Within group comparison.	F(1,23) = 35.69, p<0.001 (Significant effect)	η_p^2 0.61	2.2 Medium
Johnson-Glenberg (2000)	45	Gates-MacGinitie comprehension test. Recall of main ideas. Between group comparison	2.03 gain, p < 0.10 (No significant effect) 2.94 gain, p<0.06 (Significant effect)	*Cohen's d* = 0.35 *Cohen's d* = 0.71	2.3 Medium
Rader (2009)	69	State Development Reading Assessment. Between group comparison with benchmark scores.	8.5% mean increase across two year groups (not enough information to calculate significance)	Not given	2.4 Medium

Table 7 provides a summary of the key findings across the studies. In 4 of the studies reviewed, significant benefits of visualisation training were seen. (Center et al., 1999; Francey & Cain, 2015; Joffe et al., 2007; Johnson-Glenberg, 2000). One of the studies found no significant benefits of the intervention (Cariglia-Bull & Pressley, 1990). This study had the lowest WoE scores of the 6 studies reviewed and scored lower in WoE B. In addition, this study was rated 'low' for both 'Outcome Measures and 'Data Analysis' categories on WoE A.

It should be noted that within the studies which found significant benefits, not all findings were consistent. In one study (Johnson-Glenberg, 2000), a recognised reading comprehension assessment showed no significant effect after the mental imagery intervention but a strong positive effect was seen against an author - designed measure of the recall of main ideas from the text.

A general theme of the studies was that the benefits of visualisation training were greater for poor comprehenders than for adequate comprehenders. This supports previous findings from Oakhill and Yuill (1996) who suggest that good comprehenders could already be using mental imagery or, if not, other successful comprehension strategies. Another theme was that pupils with difficulties in either working memory (Francey & Cain, 2015) or auditory short-term memory (Cariglia-Bull & Pressley, 1990) may find creating mental images more challenging.

Section 4 Conclusions and Recommendations

This review aims to answer the question of whether visualisation training is effective in supporting school age children with comprehension difficulties and the findings of most of the studies in the review indicate that there is a direct benefit of mental imagery training on comprehension and it is therefore a potentially effective intervention for poor comprehenders. The incorporation of visual training into school practice is relatively easy; there are a range of existing approaches and programmes (E.g. Bell, 1991) but even without these, the principles of mental imagery training are well within the abilities of most education practitioners - a significant advantage of this approach.

A range of outcome measures were generated across the 6 studies and although this resulted in lower WoE ratings for some, it also demonstrates that visualisation can have a positive effect across a relatively wide range of measures. Similarly, there were differences seen in the training programmes employed by the studies – particularly in terms of the duration and intensity of intervention delivery, yet benefits were seen in most studies. All studies included visual representations to support pupils to self-generate mental images. Findings did not indicate that this resulted in any confusion. Further research could focus on how important the incorporation of visual material into visualisation training is and whether pupils are better able to generate their own images with or without a pictorial example.

The study by Johnson-Glenberg (2008) followed an established programme: Bell's visualisation, verbalisation training (Bell, 1991). Positive effects were seen in pupils' ability to recall the main ideas of a story but not in a formal reading comprehension task applied at post-test. The question arising from these particular findings is to

what extent the inclusion of verbalisation training alongside the visualisation element was beneficial. A recommendation of this review is that further research into the relative benefits of each approach is undertaken.

The findings of the studies largely support a dual coding theoretical approach, (e.g. Paivio,1971) with importance given to the visual aspects of information processing in order to support comprehension. This poses important considerations for educational practice. In UK primary schools, young children spend a good deal of time engaging with picture books which contain a multitude of rich visual and contextual information. An important consideration is whether moving to chapter books, some pupils find the loss of this visual information particularly challenging. One recommendation of this review is that the role of mental imagery training could be examined at this period, with the question of whether visualisation training has a role to play as pupils 'progress' to texts without pictures.

Further research is also suggested to examine whether visualisation supports or hinders working memory. Although one study (Cariglia-Bull & Pressley, 1990) concluded that mental imagery training can potentially overload auditory short-term memory and Francey & Cain (2015) found a smaller benefit when working memory was stretched over a greater distance between concepts to remember, other studies in the review found that visualisation training supported poor comprehenders. Further research into the relationship, if any, between auditory short-term memory, working memory and supportive comprehension interventions would be useful to enable practitioners to best support this subset of struggling readers.

The studies reviewed found that visualisation training can be useful across a range of populations. For example, Joffe et al. (2007) concluded that pupils with specific

language impairment benefitted from a mental imagery training intervention. However, these findings should be treated cautiously, as the typically developing control group received no intervention rather than comparison measures and so it is difficult to establish that visualisation training was the causal factor.

None of the studies provided outcomes measures after a significant time interval post intervention. This would have given valuable information about the sustainability of any positive effects seen. It could be argued that once participants have been trained in the generation of mental images, this ability persists and is generalised to new texts indefinitely, but this has not been assessed by any of the studies in this review and is a limiting factor when drawing conclusions about the beneficial effects of the interventions used. Therefore, a further recommendation of this review is that research incorporating follow up post-test measures is conducted to evaluate the longer-term impact of visualisation interventions on reading and listening comprehension.

References

Andrews, R., Sadoski, M., & Paivio, A. (2002). Imagery and text - A dual coding theory of reading and writing/Response. *British Journal of Educational Psychology*, *72*, 147.

Bell, N., & Lindamood, P. (1991). Visualizing and verbalizing: For language comprehension and thinking (p. 440). *Academy of Reading Publications.*

Bishop, D. V., & Adams, C. (1992). Comprehension problems in children with specific language impairment: Literal and inferential meaning. *Journal of Speech, Language, and Hearing Research*, 35(1), 119-129.

Cooter, R. B. (1989). Test review: Gates-MacGinitie reading tests, levels 5/6, 7/9, and 10/12. *Journal of Reading*, 32(7), 656-658.

Dearing, R. (1993). *The National Curriculum and its Assessment* (pp. 31-32). London: School Curriculum and Assessment Authority.

Gambrell, L. B., & Bales, R. J. (1986). Mental imagery and the comprehension-monitoring performance of fourth-and fifth-grade poor readers. *Reading Research Quarterly*, 454-464.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C. & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71, 149–164.

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22 (2), 213–228.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127-160.

Keenan, J. M., Betjemann, R. S., Wadsworth, S. J., DeFries, J. C., & Olson, R. K. (2006). Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading*, *29*(1), 75-91.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling and C. Hulme (eds), The Science of Reading: A Handbook (Oxford: Blackwells) pp. 209–226

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*(2), 91-121.

Oakhill, J., & Patel, S. (1991). Can imagery training help children who have comprehension problems? *Journal of Research in Reading*, *14*(2), 106-115.

Oakhill, J., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In J. C. Cornoldi and J. Oakhill (Eds.), Reading Comprehension Difficulties. Processes and Intervention (pp. 69-92).

Paivio, A. (1971). Imagery and language. In *Imagery* (pp. 7-32). Academic Press.

Pennington, B. F., & Olson, R. K. (2005). Genetics of dyslexia. In M. J. Snowling and C. Hulme (eds), The Science of Reading: A Handbook (Oxford: Blackwells), pp. 453-472

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling and C. Hulme (eds), The Science of Reading: A Handbook (Oxford: Blackwells) pp. 227 – 247

Pressley, G. M. (1976). Mental imagery helps eight-year-olds remember what they read. *Journal of Educational Psychology*, *68*(3), 355.

Rose, J. (2009). *Identifying and teaching children and young people with dyslexia and literacy difficulties* (DCSF-00659-2009). DCSF Publications.

Sadoski, M. (1985). Commentary: The natural use of imagery in story comprehension and recall: Replication and extension. *Reading Research Quarterly*, *20*(5), 658-667.

Sedgwick, P., & Greenwood, N. (2015). Understanding the Hawthorne effect. *British Medical Journal*, *351*.

Snowling, M. J., & Hulme, C. E. (2005). The Science of Reading: A handbook. Blackwell Publishing.

Van Dijk T. A., Kintsch W. (1983). Strategies of Discourse Comprehension. New York, NY: Academic Press.

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*, *80*(3), 351-361.

Appendix A: Full-Text Screened Articles

1. Borduin, B. J., Borduin, C. M., & Manley Charles M.; ORCID: https://orcid.org/0000-0002-1002-5020, C. M. A. I.-B. (1994). The use of imagery training to improve reading comprehension of second graders. *The Journal of Genetic Psychology: Research and Theory on Human Development*, *155*(1), 115–118.

   (Excluded under criteria 6 of exclusion criteria)

2. Cariglia-Bull, T., & Pressley, M. (1990). Short-term memory differences between children predict imagery effects when sentences are read. Journal of Experimental Child Psychology, 49(3), 384-398.

   (Included in this review)

3. Center, Y., Freeman, L., Robertson, G., & Outhred, L. (1999). The Effect of Visual Imagery Training on the Reading and Listening Comprehension of Low Listening Comprehenders in Year 2. *Journal of Research in Reading*, *22*(3), 241–256.

   (Included in this review)

4. Chang, Y.-M., Wang, I.-C., & Ma, M.-Y. (2016). Efficacy of supplementary image schemes on reading motivation and comprehension. *Special Issue: Exploring Computing's Impact on Design and Creativity in MST Education.*, *12*(5), 1153–1162.

   (Excluded under criteria 6 of exclusion criteria)

5. Compagnone Dunn, L. A. (2021). Using elements of a screenplay to promote visualization and increase reading comprehension in students with disabilities and striving readers. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *82*(10-A), No-Specified.

   (Excluded under criteria 1 of exclusion criteria)

6. Francey, G., & Cain, K. (2015). Effect of imagery training on children's comprehension of pronouns. The Journal of Educational Research, 108(1), 1-9.

   (Included in this review)

7. Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children’s reading comprehension. *JOURNAL OF EDUCATIONAL PSYCHOLOGY*, *96*(3), 424–436.

   (Excluded under criteria 6 of exclusion criteria)

8. Grünke, M., & Leidig, T. (2017). The Effects of an Intervention Combining Peer Tutoring with Story Mapping on the Text Comprehension of Struggling Readers: A Case Report. *Educational Research Quarterly*, *41*(1), 43–62.

   (Excluded under criteria 1 of exclusion criteria)

9. Joffe, V. L., Cain, K., & Marić, N. (2007). Comprehension Problems in Children with Specific Language Impairment: Does Mental Imagery Training Help? *International Journal of Language & Communication Disorders*, *42*(6), 648–664.

   (Included in this review)

10. Johnson-Glenberg, M. C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: Verbal versus visual strategies. *JOURNAL OF EDUCATIONAL PSYCHOLOGY*, *92*(4), 772–782.

(Included in this review)

11. McTigue, E. M. (2010). Teaching young readers imagery in storytelling: what color is the monkey?. *The Reading Teacher*, *64*(1), 53-56.

(Excluded under criteria 6 of exclusion criteria)

12. Rader, L. A. (2009). Teaching Students to Visualize: Nine Key Questions for Success. *Preventing School Failure*, *54*(2), 126–132.

(Included in this review)

13. Schwamborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *COMPUTERS IN HUMAN BEHAVIOR*, *27*(1), 89–93.

(Excluded under criteria 1 and 6 of exclusion criteria)

14. Vidal-Abarca, E., & Gilabert, R. (1995). Teaching strategies to create visual representations of key ideas in content area text materials: A long-term intervention inserted in school curriculum. *Special Issue: Process-Oriented Instruction: Improving Student Learning.*, *10*(4), 433–447.

(Excluded under criteria 6 of exclusion criteria)

Appendix B

Weight of Evidence (Gough, 2007) Protocols Used

WoE A Protocol Adaptations

The Gersten et al., (2005) protocol for evaluating Weight of Evidence A (Gough, 2007) was adapted for clarity and relevance.

Adaptations are noted below:

Strikethrough of text (~~example~~) denotes omissions

{…} denotes additions

[…] details the rationale for the adaptation

Adapted Essential Quality Indicators

Quality indicators for describing participants

Was sufficient information provided to determine/confirm whether the participants

demonstrated the ~~disability(ies) or~~ difficulties presented?

[Rationale: to better reflect the individual differences in relation to poor comprehension skills]

Quality indicators for Implementation of the intervention and Description of Comparison Conditions

Was the intervention clearly described and ~~specified~~?

[Rationale: Not all visualisation / mental imagery training would be classed as specific, therefore this is not applicable; a description would be considered adequate]

Adapted Desirable Quality Indicators

~~Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?~~

{Is attrition comparable across samples?}

{Was severe overall attrition documented? If so, is overall attrition less than 30%?}

[Rationale: it was considered that this question would be more likely to reflect the strengths and limitations of the studies if split into two separate questions. The first part of the question was removed as the response of 'unknown' would give the same information and improve clarity]

~~Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures? Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?~~

{Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures?}

{Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?}

[Rationale: it was considered that this question should be split into two 2 separate questions. In studies in this area, data collectors could be educational professionals known to the students. which would result in a score of zero even if other aspects of reliability had been considered.]

Did the research report include actual audio, videotape {or written} excerpts that capture the nature of the intervention?

[Rationale: some studies used written and other visual material and therefore records of these should be considered valuable as excerpts.]

WoE A Final Protocol Used

Study: …….

Essential Quality Indicators

Quality indicators for describing participants

Was sufficient information provided to determine/confirm whether the participants demonstrated the difficulties presented?

☐Yes
☐No
☐N/A
☐Unknown/Unable to Code

Was sufficient information provided on participants (Age, Ethnicity, Social Economic Status, information on any disabilities)?

☐Yes
☐No
☐N/A
☐Unknown/Unable to Code

Did the study use appropriate participants for the research question?

☐Yes
☐No
☐N/A
☐Unknown/Unable to Code

Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?

☐Yes
☐No
☐N/A
☐Unknown/Unable to Code

Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

☐Yes
☐No
☐N/A
☐Unknown/Unable to Code

Quality indicators for Implementation of the Intervention and Description of Comparison Conditions

Was the intervention clearly described?

☐Yes
☐No
☐N/A
☐Unknown/Unable to Code

Was the fidelity of implementation described and assessed?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?
☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Quality Indicators for Outcome Measures

Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?

☐Yes
☐ No

☐ N/A
☐ Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured at the appropriate times?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Quality Indicators for Data Analysis

Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the limit of analysis in the study?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Did the research report include not only inferential statistics but also effect size calculations?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Essential Quality Indicators Total Score:

Desirable Quality Indicators

Is attrition comparable across samples?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Was severe overall attrition documented? If so, is overall attrition less than 30%?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Was evidence of the criterion-related validity and construct validity of the measures provided?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Did the research team assess not only surface features of fidelity implementation (e.g. number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Was any documentation of the nature of instruction or series provided in comparison conditions?

☐Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Did the research report include actual audio, videotape or written excerpts that capture the nature of the intervention?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Were results presented in a clear, coherent fashion?

☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Desirable Quality Indicators Total Score:

Total Score

Essential Quality Indicators:	>10 = 2, 9 - 10 = 1, <9 = 0	Study score:
Desirable Quality Indicators:	>4 = 2, 2 - 4 = 1, <2 = 0	Study score:
Total Score and rating:	≥3 = High 2 = Medium <2 = Low	Study score: Study rating: **……. Quality**

<u>WoE B Coding Protocol developed by the author</u>

Weight of Evidence B concerns the relevance and quality of the methodology used in the studies examined. The author of this review developed the coding protocol for Weight of Evidence B in order to evaluate the studies against three categories: 'Study Design', 'Comparison Cohort' and 'Outcome Measures'.

'Study Design' was chosen because this review is concerned with the efficacy of visualisation training as an intervention. The evaluation of this aspect was informed by evidence typologies, as detailed in Petticrew and Roberts, 2003, which show that randomised control trials are the most appropriate design to answer an efficacy question. Pilot or feasibility studies, which may be less powered than a definitive randomised control trial also score highly. At the lower end of the hierarchy, qualitative studies and case reports would be less suitable in providing a reliable answer to the question of efficacy. Cohort studies, such as the majority of the studies considered in this review achieve a score in the middle range.

‘Comparison Cohort’ was included in order to evaluate how the comparison groups were selected in the studies. It is possible that an impact could be found but it not be due to the visualisation aspect of the intervention, but to a different feature of the intervention.

‘Outcome measures’ considered the quality of pre and post tests in terms of the number and range of measures used, how comparable they were between pre and post test and whether the issue of reliability between assessors was addressed (if more than one assessor was used). Reading or listening comprehension is a complex skill which is best captured using a range of measures rather than a single measure. Studies which scored highly in this category included several measures which were comparable pre and post intervention with controls for reliability .

Weight of Evidence B Criteria, Ratings and Rationale

Criteria	Ratings		Rationale
Study Design	3.	Study design includes randomised control group	Randomisation of allocation to intervention / control group reduces bias. A control group allows comparison of outcome measures.
	2.	Study design includes a control group but is not randomised	
	1.	Cohort study without control	
		Single case experimental study	
Comparison Cohort	3	Groups matched for skill level in reading comprehension AND exposed to an equivalent intervention	Comparison of groups with similar skill levels allows for control of the impact of

	2	Matched for skill level but no equivalent intervention OR Equivalent intervention but not matched for skill level	individual differences on the outcome measures. Use of an equivalent intervention ensures both groups had similar exposure to input and attention from trainers.
	1	Comparison group has a different skill level and receives no intervention	
Outcome Measures	3	Comparable pre and post-test measures AND a range of measures used AND the issue of reliability addressed	If outcome measures cannot be compared or are not reliable this will result in unreliable outcome measures.
	2	Pre and post-test measures are not comparable OR a range of measures was not used	If a range of measures is not used the outcome measures would be a weak indication of the true impact of the intervention
	1	Pre and post-test measures are not comparable AND a range of measures was not used AND the issue of reliability was not addressed	

Note. WoE B ratings are calculated as an average across all three categories and considered ‘High’ for scores ≥ 2.5, ‘Medium’ for scores between 1.5 and 2.5, and ‘Low’ for scores < 1.5.

WoE C Coding Protocol developed by the author

Weight of Evidence C concerns the relevance of the studies examined to the topic of the review. The author of this review developed the coding protocol for Weight of Evidence C in order to evaluate the studies against three categories: 'Comprehension Measure', 'Visualisation Training Method and 'SEN/ Poor Comprehender Status'.

'Comprehension Measure' was chosen because this review is concerned with the effect of visualisation training on reading comprehension. Some studies decided to use listening rather than reading comprehension in order to measure the impact of mental imagery training. As stated in the introduction section, listening and reading comprehension share a common skill set. By using listening comprehension with subjects who are unable to decode at the level of complexity required for assessment, the researchers were able to eliminate any effect of decoding skills.

'Visualisation Training Method' was considered important to consider as it is important that an intervention is thorough and explicit, with time taken for feedback from pupils to ensure that they understand and are able to employ the principles of visualisation. Studies which demonstrated an awareness of the need for training to be clear and explicit scored more highly than those which used generic measures with little opportunity to assess whether participants realised what was required.

'SEN and Poor Comprehender Status' varied between the studies. The current review is concerned with the impact on poor comprehenders. A higher score would indicate that the participants were adequate decoders and poor comprehenders as this is the target participant status for this review. Lower scores indicate that the participant cohort are typically developing readers or pupils with additional needs.

Weight of Evidence C Criteria, Ratings and Rationale

Criteria	Ratings		Rationale
Comprehension Measure	3.	Study includes a measure of both reading comprehension and verbal comprehension	This study is concerned with evaluating the impact of intervention on reading comprehension. Studies which assessed only verbal comprehension are still considered a valid comparison as requisite skills of language comprehension and memory underpin both. Including both measures scores most highly as the decoding ability of some pupils may prevent them from accessing the language of a text.
	2.	Study includes a measure of either reading comprehension or verbal comprehension	
	1.	Study does not include a measure of either reading or verbal comprehension	
Visualisation training method	3	Participants are given explicit training on how to make mental images in their mind in more than one session	For participants, especially those identified as having comprehension difficulties it is important that an intervention is thorough and explicit, with time taken for feedback from pupils to ensure that they understand and are able to employ the principles of visualisation. Giving participants an image of what they could have imagined did not increase the score as each mental representation is has a unique quality.
	2	Participants are given explicit training on how to make mental images on only one occasion	
	1	Participants are given only vague, brief instructions.	

SEN and poor comprehender status	3	Participants are identified against agreed criteria as having poor comprehension skills but adequate decoding skills.	By focusing on poor comprehenders as opposed to poor readers in general, this review seeks to look at whether visualisation can help pupils understand a text better by forming mental images to aid memory. If pupils are struggling to decode text it may indicate that they are unable to access the language and concepts that they could understand if it were read to them therefore their main reading difficulty would not relate to comprehension and they would not be good candidates for a class based visualisation intervention. Pupils with identified language difficulties are also included under the umbrella of 'poor comprehenders' as their language needs are likely to impact on reading comprehension.
	2	Participants are identified as poor comprehenders and poor decoders but this is countered by verbal delivery of texts. Pupils are identified as having a general language difficulty	
	1	Pupils are not identified as having comprehension difficulties and are typically developing for their age.	

Note. WoE C ratings are calculated as an average across all three categories and considered 'High' for scores ≥ 2.5, 'Medium' for scores between 1.5 and 2.5, and 'Low' for scores < 1.5.

Appendix C

Completed Example of WoE A Rating for Study 1: Cariglia-Bull, T., & Pressley, M. (1990). Short-term memory differences between children predict imagery effects when sentences are read. *Journal of Experimental Child Psychology*, *49*(3), 384-398.

Essential Quality Indicators

Quality indicators for describing participants

Was sufficient information provided to determine / confirm whether the participants demonstrated the difficulties presented?

☒Yes
☐No
☐N/A
☐Unknown/Unable to Code

Was sufficient information provided on participants (Age, Ethnicity, Social Economic Status, information on any disabilities)?

☒Yes
☐No
☐N/A
☐Unknown/Unable to Code

Did the study use appropriate participants for the research question?

☒Yes
☐No
☐N/A
☐Unknown/Unable to Code

Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?
☒Yes
☐No

☐N/A
☐Unknown/Unable to Code

Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

☐Yes
☐No
☐N/A
☒Unknown/Unable to Code

Quality indicators for Implementation of the Intervention and Description of Comparison Conditions

Was the intervention clearly described?

☒Yes
☐No
☐N/A
☐Unknown/Unable to Code

Was the fidelity of implementation described and assessed?

☐ Yes
☐ No
☐ N/A
☒ Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?
☒ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Quality Indicators for Outcome Measures
Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?

☐Yes
☒ No

☐ N/A
☐ Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured at the appropriate times?

☒ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Quality Indicators for Data Analysis

Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the limit of analysis in the study?

☒ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Did the research report include not only inferential statistics but also effect size calculations?

☐ Yes
☒ No
☐ N/A
☐ Unknown/Unable to Code

Essential Quality Indicators Total Score: 8

Desirable Quality Indicators

Is attrition comparable across samples?

☒ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Was severe overall attrition documented? If so, is overall attrition less than 30%?

☒ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures?

☐ Yes
☒ No
☐ N/A
☐ Unknown/Unable to Code

Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?

☐ Yes
☒ No
☐ N/A
☐ Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?

☐ Yes
☒ No
☐ N/A
☐ Unknown/Unable to Code

Was evidence of the criterion-related validity and construct validity of the measures provided?

☒ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Did the research team assess not only surface features of fidelity implementation (e.g. number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?

☐ Yes
☐ No
☐ N/A
☒ Unknown/Unable to Code

Was any documentation of the nature of instruction or series provided in comparison conditions?

☐Yes
☐ No
☒ N/A
☐ Unknown/Unable to Code

Did the research report include actual audio, videotape or written excerpts that capture the nature of the intervention?

☒ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Were results presented in a clear, coherent fashion?

☒ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Desirable Quality Indicators Total Score: 5

Total Score

Essential Quality Indicators:	>10 = 2, 9 - 10 = 1, <9 = 0	Study score: 1
Desirable Quality Indicators:	>4 = 2, 2 - 4 = 1, <2 = 0	Study score: 2
Total Score and rating:	≥3 = High Quality 2 = Medium Quality <2 = Low Quality	Study score: 2 Study rating: **Medium Quality**

WoE B Ratings for Each Study

Study	WoE B Criteria			Average WoE B & rating
	Study Design	Comparison Cohort	Outcome Measures	
Cariglia-Bull & Pressley (1990)	3 High	2 Medium	2 Medium	2.3 Medium
Center et al. (1999)	2 Medium	3 High	2 Medium	2.3 Medium
Francey & Cain (2015)	2 Medium	2 Medium	2 Medium	2 Medium
Joffe et al.(2007)	2 Medium	1 Low	3 High	2 Medium
Johnson-Glenberg (2000)	2 Medium	2 Medium	3 High	2.3 Medium
Rader (2009)	2 Medium	2 Medium	2 Medium	2 Medium

Note. ≥ 2.5 = High, 1.5 -2.5 = Medium, < 1.5 = Low

WoE C Ratings for Each Study

Study	WoE C Criteria			Average WoE C & rating
	Comprehension Measure	Visualisation Training	SEN / Comprehender status	
Cariglia-Bull & Pressley (1990)	2 Medium	2 Medium	1 Low	1.7 Medium
Center et al. (1999)	2 Medium	3 High	3 High	2.6 High
Francey & Cain (2015)	1 Low	2 Medium	2 Medium	1.7 Medium
Joffe et al.(2007)	3 High	3 High	1 Low	1.7 Medium
Johnson-Glenberg (2000)	1 Low	1 Low	3 High	1.7 Medium
Rader (2009)	2 Medium	3 High	2 Medium	2.3 Medium

Note. ≥ 2.5 = High, 1.5 -2.5 = Medium, < 1.5 = Low