

Case Study 1: An Evidence-Based Practice Review Report

***How effective is the Self-Regulated Strategy Development (SRSD)
instruction approach in improving reading comprehension of school-
age pupils with special educational needs (SEN)?***

Summary

The ability to construct meaning from the written word is an essential yet complex skill that is central to both accessing learning in school and navigating life in a text-rich society. However, many pupils with special educational needs (SEN) experience difficulties in reading comprehension, which can significantly impact their learning. Self-regulated strategy development (SRSD) is an intervention approach designed to improve learners' academic skills through the instruction of specific academic strategies and self-regulation skills. This systematic literature review evaluated the effectiveness of using the SRSD approach to deliver strategies targeting reading comprehension for this group of learners. The findings provide promising evidence for its efficacy in improving comprehension as measured by retell or summarisation, but the generalisation and maintenance of effects is less clear, which may limit the intervention's practical value. Implications for use and future research are discussed.

Introduction

Reading comprehension

While most children who develop decoding skills successfully will also develop their comprehension skills, it is recognised that the latter is not simply a by-product of the former (Snowling & Hulme, 2011). Predictors of reading comprehension difficulties include poor knowledge of language, vocabulary and grammar (Muter, Hulme, Snowling, & Stevenson, 2004) as well as deficits in correctly using reading strategies and self-regulation procedures (Hagaman & Casey, 2017). Affective factors may also influence reading skill, with lower intrinsic reading motivation shown to contribute to low ability readers' performance in reading assessments (Logan, Medford, & Hughes, 2011).

Many pupils who have special educational needs (SEN) experience difficulties in reading comprehension (Nation, Clarke, Wright, & Williams, 2006). This has significant educational and social implications, as difficulties are likely to compound when pupils progress through the education system, where there are increasing emphases on independent learning through reading (Mastropieri, Scruggs, Graetz, & Graet, 2003). If left unsupported, the increasing disparity between their reading abilities and the level of required reading materials can present a growing barrier to accessing the curriculum.

A key aspect of the role of an educational psychologist (EP) is to work collaboratively with teachers to improve the learning experiences of children

and young people with SEN. This can be done through supplementing school staff's professional skills in identifying and using evidence-based interventions, as recommended by the Special Educational Needs and Disability (SEND) Code of Practice (Department for Education [DfE] & Department of Health [DH], 2015). The purpose of this review is to contribute to the evidence base for reading comprehension interventions by evaluating the self-regulated strategy development (SRSD) model (Graham & Harris, 1993; Harris & Graham, 1996), which is an approach that shows promise in its effectiveness in supporting pupils with SEN.

Self-regulated strategy development

SRSD is a multi-component strategy instruction model designed to improve learners' strategic learning behaviour, self-regulation, and motivation (Harris, Graham, Mason, & Friedlander, 2008). Learners are taught task-specific strategies along with procedures for regulating their task behaviour and the use of strategies.

The SRSD model draws on social cognitive, constructivist and behavioural theoretical perspectives about teaching and learning (Harris & Graham, 2009). Partly based on the cognitive-behavioural model, the intervention is staged and emphasises interactive learning between the teacher and learners. The learners' awareness of the learning process and self-control is developed through a structured and explicit instruction of self-regulation and metacognitive skills. Teaching is reinforced through dialogue, modelling, individualised feedback, and gradually faded support for strategy use.

Informed by social-cognitive theories, another goal in SRSD is the development of self-efficacy through attributions of success to effort and strategy use, which in turn enhances intrinsic motivation in performing the learning task as well as the use of self-regulatory processes.

Instruction is given through six stages of strategy acquisition that scaffold the learner towards independently and effectively use of the strategy (see Table 1). In addition, explicit instruction of self-regulatory components are embedded into each instructional step. These may include goal setting, mnemonics, modelling, self-instruction, self-monitoring, and self-reinforcement (Harris et al., 2008; Mason, 2013).

Table 1
Six stages of instruction for strategy acquisition in SRSD (adapted from Mason, 2013)

SRSD stages	Activities
Develop pre-skills	Learners' prior knowledge about the task and strategy is assessed and remediation is provided when needed.
Discuss the strategy	The strategy to be learned is described, a purpose for using the strategy is established, and the benefits of using the strategy are presented.
Model the strategy	The teacher cognitively models how to use and apply the strategy for the task by thinking out loud.
Memorise the strategy	Learners memorise the strategy steps until they are fluent in understanding any mnemonics and meanings.
Guided practice	Instruction is scaffolded from teacher-learner collaborative practice towards independence.
Independent practice	The teacher encourages independent practice across tasks and settings to foster generalisation and maintenance

SRSD was initially developed to address pupils' difficulties in writing (Harris & Graham, 1996). It has evolved over time and has been used for other academic content areas, including math and reading (e.g. Jitendra et al., 2015; Mason, 2004), and within individual, small group, or whole classroom settings (What Works Clearinghouse, 2017). Different strategies may be taught, depending on the content area as well as the needs of the learner. Mnemonic devices are often used to help learners remember to include all required steps or elements when carrying out the strategy.

Lessons are typically delivered at least three times a week, and may last between 20 and 60 minutes (Harris & Graham, 2009). Over the course of the intervention, the control over the use of strategies is gradually transferred to the learner. Instruction is recursive, and continues until the learner reaches the criterion for mastery. As such, there is no set numbers of sessions, as lessons may be repeated and revisited based on the needs of individual pupils.

SRSD has primarily been researched in the area of school-based writing instruction (Harris & Graham, 2009). Within this domain, it was demonstrated to be effective with a variety of groups of pupils with SEN, such as autism spectrum disorders (ASD; Asaro-Saddler, 2016), attention deficit hyperactivity disorder (ADHD; Reid, Hagaman, & Graham, 2014), and emotional or behavioural difficulties (EBD; Sreckovic, Knowles, & Lane, 2014). Most studies were carried out in the United States, although one

large-scale UK-based efficacy trial (Torgerson et al., 2014) has also found a strong positive effect on the writing outcomes of low-attaining Year 6 pupils across 23 schools in West Yorkshire.

This review will focus on evaluating the use of SRSD to teach strategies for reading, which has received relatively little research attention. At the time of writing, no systematic review of literature has been published focusing on SRSD's use in improving reading comprehension for pupils with SEN.

Review Question

In light of the aforementioned rationale, this review will aim to answer the review question:

How effective is the Self-Regulated Strategy Development (SRSD) instruction approach in improving reading comprehension of school-age pupils with special educational needs (SEN)?

Critical Review of the Evidence Base

Literature search

A systematic literature search was carried out on 25th January 2018 using three online databases (PsycINFO, ERIC, and the Web of Science) to locate articles across the fields of psychology, education, and social sciences. The searches included the terms listed in Table 2, and were limited to journal articles as per the inclusion criteria listed in Table 3.

Table 2
Search terms used in the systematic literature search

Search terms	Rationale
"Self-regulated Strategy Development" or SRSD or TWA or RAP	<ul style="list-style-type: none"> The review seeks to evaluate studies that have implemented interventions using the SRSD approach. "TWA" and "RAP" were included as additional search terms as pilot searches have revealed these to be commonly used reading comprehension strategies taught using the SRSD model.
reading comprehen*	<ul style="list-style-type: none"> The review seeks to evaluate the impact of the interventions on reading comprehension skills.

Note. RAP = "Read a paragraph, Ask myself "What was the main idea and two details?" and Put it into my own words" (a three-step paraphrasing strategy); TWA = "Think before reading, think *while* reading, think *after* reading" (a nine-step reading comprehension strategy)

Screening of articles

The searches yielded 68 articles across the three databases. Following the removal of duplicates (n = 26) and articles that did not meet the inclusion criteria at title and abstract screening (n = 18), the 24 remaining articles were screened at full text to determine their eligibility for inclusion in the review. Nineteen studies were further excluded with reasons listed in Appendix A.

The five studies that were eligible for the final review are listed in Table 4, and the characteristics and key findings of each study are summarised Appendix B. This selection process is represented graphically in Figure 1.

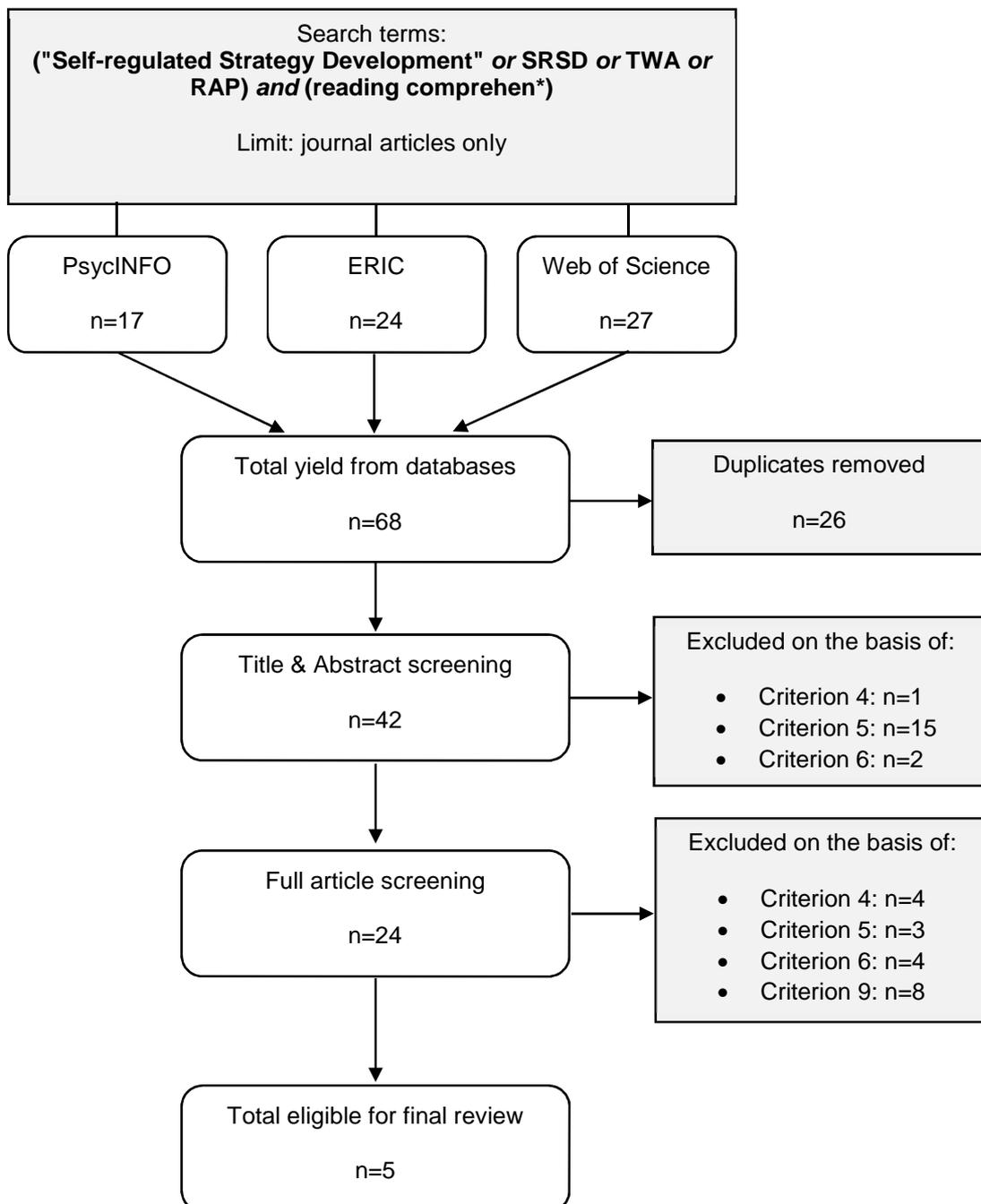


Figure 1. Flow Diagram of the Article Selection Process

Table 3
Inclusion and exclusion criteria for the literature search

Criteria	Inclusion	Exclusion	Rationale
1 Type of publication	The article is published in a peer reviewed journal	The article is not published in a peer reviewed journal	This ensures the quality of the article has been checked
2 Language	The article is written up in English	The article is not written up in English	This ensures that the paper can be understood and evaluated, and that meaning is not lost or distorted through translations.
3 Publication date	The study is first published before 25 th January 2018	The study is first published on or after 25 th January 2018	This allows all published research that is relevant to the topic at the time of review to be included
4 Primary empirical data	The article presents original research that uses an experimental or quasi-experimental design	Review article that does not present original empirical findings	This review aims to collate experimental evidence to evaluate the impact of SRSD on children's reading comprehension
5 Intervention	At least one of the intervention conditions follows the SRSD model	None of the conditions follow the SRSD model	The review question is specific to SRSD and not strategy instruction or reading comprehension training in general
6 Outcome measures	At least one outcome measure evaluates the pupils' reading comprehension skills	The outcome measures do not evaluate the pupils' reading comprehension skills	The review question aims to establish whether SRSD is effective in improving reading comprehension
7 Setting	The research is carried out in a school or other types of educational settings	The research is not based in a school or educational setting, e.g. home, or a clinical or community setting	The review aims to evaluate school/setting based interventions for children with SEN

Criteria	Inclusion	Exclusion	Rationale
8 Participants' age	Participants are of English schooling age (4 to 16 years old)	Participants are in preschool or attend post-16 educational settings	The review question is specifically evaluating the impact of SRSD on school age pupils with SEND
9 Participants' special educational needs (SEN) status	At least half of the participant group have an identified SEN that corresponds to the four broad areas of need specified in the SEND Code of Practice (DfE & DoH, 2015).	Fewer than half of all participants have an eligible SEN, or if the proportion is not specified	The review question seeks to evaluate the impact of SRSD on pupils with SEN

Table 4
 Full references of studies included in the review

Eligible studies	
1	Howorth, S., Lopata, C., Thomeer, M., & Rodgers, J. (2016). Effects of the TWA strategy on expository reading comprehension of students with autism. <i>British Journal of Special Education</i> , 43(1), 39–59. https://doi.org/10.1111/1467-8578.12122
2	Johnson, J. W., Reid, R., & Mason, L. H. (2012). Improving the Reading Recall of High School Students With ADHD. <i>Remedial and Special Education</i> , 33(4), 258–268. https://doi.org/10.1177/0741932511403502
3	Hedin, L. R., Mason, L. H., & Gaffney, J. S. (2011). Comprehension Strategy Instruction for Two Students with Attention-Related Disabilities. <i>Preventing School Failure</i> , 55(3), 148–157. https://doi.org/10.1080/1045988X.2010.49939
4	Rogevich, M. E., & Perin, D. (2008). Effects on science summarization of a reading comprehension intervention for adolescents with behavior and attention disorders. <i>Exceptional Children</i> , 74(2), 135–154. https://doi.org/10.1177/001440290807400201
5	Johnson, L., Graham, S., & Harris, K. R. (1997). The effects of goal setting and self-instruction on learning a reading comprehension strategy: A study of students with learning disabilities. <i>Journal of Learning Disabilities</i> , 30(1), 80–91. https://doi.org/10.1177/002221949703000107

Critical comparisons of selected studies

Weight of Evidence (WoE)

The quality and relevance of the research evidence from the five included studies were evaluated using Gough's (2007) Weight of Evidence (WoE) framework. This approach takes into account the studies' methodological quality (WoE A), methodological relevance (WoE B), and topic relevance (WoE C).

The WoE A rating is based on a generic judgement of the coherence and integrity of the evidence presented in the studies. Kratochwill's (2003) coding protocols for group-based (pp. 45-62) and single-case designs (pp. 84-113) were adapted for use in this category. WoE B and C ratings, on the other hand, are review-specific judgements of the appropriateness of the studies' methodology and topic to the review question. These are averaged to produce an overall weight of evidence (WoE D). The WoE ratings for all five included studies are summarised in Table 5.

The criteria and rationale for all the WoE ratings are outlined in Appendix C, while the amendments made to Kratochwill's (2003) protocols are specified in Appendix D. Completed examples of both coding protocols can be found in Appendices E and F.

Table 5
Weight of evidence of included studies

Authors	WoE A: Methodological quality	WoE B: Methodological relevance	WoE C: Topic relevance	WoE D: Overall weight of evidence
Howorth et al. (2016)	2.4	2	1.8	2.1 medium
Johnson et al. (2012)	2.2	2	1.8	2.0 medium
Hedin et al. (2011)	1.8	0	1.8	1.2 low
Rogevich & Perin (2008)	2.0	2	1.8	1.9 medium
Johnson et al. (1997)	2.4	1	2.2	1.9 medium

WoE ratings are described as 'High' for average scores of 2.5 or above, 'Medium' for scores between 1.5 and 2.4, and 'Low' for scores of 1.4 or below. Table 5 provides a summary of the WoE scores for the reviewed studies.

Participants

Excluding attrition, a total of 119 participants took part in the reviewed studies, with ages ranging from 10 to 16. The inclusion of both group-based and single-case studies is reflected in a substantial variance in sample size, which ranged from two participants (Hedin, Mason, & Gaffney, 2011) to 63 in (Rogevich & Perin, 2008). Both group-based studies were underpowered for their respective statistical analyses (Cohen, 1992).

The participant group is overwhelmingly male-dominated. Only one study (Johnson et al., 1997) involved female participants (n=13), who accounted for 28% of the study's group, and 11% of the overall sample.

A wide range of SEN were presented across the five studies. Three studies (Hedin, Mason, & Gaffney, 2011; Johnson, Reid, & Mason, 2012; Rogevich & Perin, 2008) recruited pupils with ADHD, the majority of whom had a dual diagnosis of speech and language impairment (SLI), learning difficulties (LD), and behavioural difficulties (BD). Pupils with ASD (Howorth, Lopata, Thomeer, & Rodgers, 2016) and LD without comorbid condition (Johnson, Graham, & Harris, 1997) were also included.

Most studies recruited participants with explicit reference to their SEN diagnostic label and reading difficulties in the inclusion criteria. Three of the studies selected their participants by purposive sampling from the school (Rogevich & Perin, 2008) or school district population (Howorth et al., 2016; Johnson et al., 1997) using existing pupil records. Both Hedin et al. (2011) and Johnson et al. (2012) recruited by teacher nomination. However, the former study's lack of an explicit inclusion criteria raises the likelihood of selection bias, and limits the chance for the study to be replicated. This is reflected in a lower WoE C rating for participant selection.

Appendix B provides an overview of the participant characteristics and other key aspects of the reviewed studies.

Research designs

Three studies used single case experimental designs (SCEDs), of which two (Howorth et al., 2016; Johnson et al., 2012) used a multiple baseline design. By demonstrating the intervention effect at least three times between several participants, they were able to achieve experimental control and thus strengthen their findings' internal validity (Horner, Carr, Halle, & Mcgee, 2005).

The choice of multiple baseline across participants design is suitable for reading comprehension intervention research (Snyder et al., 2017), where skills may often not be reversed once they were taught, making it harder for intervention effect to be demonstrated with just one participant even when the intervention is withdrawn.

It is argued that external validity of SCED findings can be attained through systematic replications and syntheses of the research across different settings and populations (Byiers, Reichle, & Symons, 2012). However, in light of the current review question, both studies can only be given a maximum WoE B rating of 'medium' quality, because the generalisability of findings from individual studies is limited.

Hedin et al. (2011) also used a single-case design. However, its use of within-series simple phase change with two participants means that experimental control is lacking, and the effect of SRSD cannot be concluded. Accordingly, the study is given a '0' rating for methodological relevance.

Both group studies (Johnson et al., 1997; Rogevich & Perin, 2008) separated the participants into four groups, and their different research focus is reflected in their choice of control groups. Both studies achieved a 'high' score for the WoE A 'Comparison Group' rating by including control conditions that received equivalent attention.

Rogevich and Perin (2008) tested for the effects of two independent variables, namely the *intervention* (intervention versus control) and the *diagnoses* (BD only versus BD with ADHD). The researchers controlled for Hawthorne effects by providing an equally motivating and challenging placebo condition. However, because this comparison condition is not an alternative intervention, conclusion cannot be drawn about the effectiveness of SRSD in comparison with other literacy interventions. Therefore a 'medium' rating for WoE B is given.

Johnson et al. (1997), on the other hand, was primarily interested in performing a component analysis of SRSD to identify the unique contribution of the *self-instruction* and *goal-setting* components to improvements. The four groups received different versions of SRSD (with and without combinations of additional self-regulation components). This is appropriate for the stated aim of this study as each condition can be considered the other's alternative intervention control, allowing comparisons to identify the relative gains due to each component and lack thereof. This also earned the study a 'high' rating for the WoE A "Identifiable Component" scale.

Johnson et al.'s (1997) use of randomisation for group assignment was also a desirable feature that enables researchers to draw more valid conclusions about the intervention effect on outcome (Barker, Pistrang, & Elliott, 2015). Nevertheless, given that the aim of the current review is to evaluate the effectiveness of SRSD for improving reading comprehension, all four groups would qualify as an intervention condition. Without a no-treatment or alternative-intervention control groups, confounding factors such as natural development through time are not controlled for. Hence, an overall 'low' WoE B score is given.

Intervention content and fidelity

All the studies delivered a reading comprehension strategy using the SRSD approach, but their procedures differed.

Instructions were given on an individual basis in all three SCED studies, and in small groups in both group-based studies. The length of intervention also varied, although the disparity is not a cause for concern because SRSD instruction typically follows a criterion-based progression (Harris et al., 2008). This means that learners only move through to the next stage when they have met a set criterion of mastery. Three of the studies (Howorth et al., 2016; Johnson et al., 2012; Johnson et al., 1997) followed this procedure, whereas the other two had a fixed number sessions.

All studies reported basing their SRSD instruction on a published protocol. Only Johnson et al. (1997), through their specific component analysis described above, qualified for a 'high' rating for the WoE A "Identifiable Component" scale. However, all studies provided a clear outline of the SRSD instructional stages and lesson structures, as well as the adaptations made to the protocol. Of these, Johnson et al. (1997), Johnson et al. (2011), and Rogevich and Perin (2016) provided a greater level of detail, and were thus scored higher on the relevant WoE C measure.

Four studies taught a version of the nine-step "*Think before reading, think while reading, and think after reading*" (TWA; Mason, 2004) protocol (see Table 6), and involved the reading of informational passages from textbooks or websites (Howorth et al., 2016). Johnson et al. (1997), on the other hand, used a research-validated story grammar strategy to support narrative comprehension. All studies described steps taken to ensure a functional equivalence between passages (Snyder, 2017), and that the reading materials were appropriate to the participants' reading level. All except Howorth et al. (2016) mentioned the use of randomisation or counterbalancing to minimise order effects.

Adherence to the protocol was checked in three studies (Hedin et al., 2011; Howorth et al., 2016; Johnson et al., 2012) by coders using checklist and recordings of the sessions, which led to higher WoE A score for "Implementation Fidelity".

Table 6
Components of the “Think before reading, think while reading, and think after reading” (TWA) strategy (adapted from Mason, 2004)

Stage of reading	Steps to think about
Think before reading	<ul style="list-style-type: none"> • The author’s purpose • What you know • What you want to know
Think while reading	<ul style="list-style-type: none"> • Reading speed • Linking knowledge • Rereading parts
Think after reading	<ul style="list-style-type: none"> • The main idea • Summarising information • What you learned

Outcome measures

Most studies used oral retell as the primary outcome measure of the intervention. This is considered a valid measure of reading comprehension, as research has demonstrated a positive association with the accuracy, sequence and discussion of main ideas and supporting details recalled after reading a text (Mason, 2013; Whalon, Al Otaiba, & Delano, 2009). The level of comprehension is quantified in a variety of ways, including by counting the number of pre-determined main ideas, percentages of supporting detail, or by a research-specific rubric.

Rogevich and Perin (2008) was the only study to use written summarisation as their main outcome. The additional writing skills involved may mean that it is a less direct measure of reading comprehension, and less appropriate in replication studies if the potential groups of participants have difficulties with

writing. Nevertheless, it is an educationally meaningful and appropriate measure that can demonstrate the intervention effect of strategies learned through SRSD.

Most studies have reported procedures of inter-rater reliability and validity checks of their primary outcome measures, but the omission of a reliability indicator in Hedin et al. (2011) led to a low rating on the WoE A 'Measurement' scale.

Findings

Tables 9 and 10 present a summary of effect sizes of the single-case and group-based studies respectively. Interpretations of these values were based on Cohen's (1992) and Scruggs and Mastropieri's (1998) descriptors (see Tables 7 and 8).

All three single-case studies presented their results on line graphs, and discussed their findings with reference to visual analyses of trend, level, and stability of data within and between conditions (Lane & Gast, 2014). Howorth et al. (2016) and Johnson et al. (2012) also provided non-overlap effect size estimates in the Percentage of Nonoverlapping Data (PND; Scruggs & Mastropieri, 1998, 2013) the Percentage of Data Exceeding the Median (PEM; Ma, 2006).

The lack of precise data means more sensitive effect size estimates such as Tau-U (Parker, Vannest, Davis, & Sauber, 2011) cannot be accurately

calculated. PND is chosen instead as the main evaluative effect size measure for SCEDs in this review for its correlation with visual analysis judgement, and its wide use in the literature means that the findings can be directly compared with other research (Lenz, 2013; Scruggs & Mastropieri, 2013). PND was hand-calculated by the reviewer for results in Hedin et al. (2011) using the graphical data.

Table 10 includes the Rogevich and Perin's (2008) effect sizes (d) for statistically significant pairwise between-group comparisons ($p < .01$ after Bonferroni corrections), as presented in the paper. These values are based on between-group mean differences in gain scores (unadjusted scores for each post-intervention measure subtracted by pre-intervention score) divided by pooled standard deviation for the unadjusted post-intervention scores.

Johnson et al. (1997) also reported that post-hoc analyses indicated a significant difference for posttest and maintenance scores compared to pretest, but did not include effect size estimates. In the absence of no-treatment control groups, Cohen's d (Cohen, 1988) were calculated based on within-group mean score changes by the reviewer using an online effect size calculator (Wilson, n.d.). Since the authors reported no significant main effects for instructional conditions, scores from all four treatment conditions were aggregated for this calculation, with between sub-group variance taken into account in the pooled SD.

Table 7
Effect size descriptors for PND (Scruggs & Mastropieri, 1998)

PND range	Descriptive label
Over 90%	Very effective
70 - 89%	Effective
50 – 69%	Questionable
Below 50%	Ineffective

Table 8
Effect size descriptors for Cohen's d (Cohen, 1992)

Cohen's d	Descriptive label
.8	Large
.5	Medium
.2	Small

Table 9
Summary of effect sizes from single-case design studies

Author	Design	Outcome measure	Participants	T2 (Intervention)		T3 (Maintenance)		WoE D
				PND	Descriptor	PND	Descriptor	
Howorth et al. (2016)	Multiple Baseline across Participants	Oral retell: researcher-developed rubric scoring	1	100%	very effective	100%	very effective	2.1 medium
			2	66.70%	questionable	100%	very effective	
			3	50%	questionable	100%	very effective	
			4	83.30%	effective	100%	very effective	
		Comprehension questions	1	100%	very effective	66.70%	questionable	
			2	NC	NC	NC	NC	
			3	83.30%	effective	66.70%	questionable	
			4	83.30%	effective	100%	very effective	
Johnson et al. (2012)	Multiple baseline across participants + multiple probe	Oral retell: number of main ideas	1	80%	very effective	100%	very effective	2.0 medium
			2	100%	very effective	100%	very effective	
			3	100%	very effective	100%	very effective	
		Oral retell: percentage of details	1	66.70%	questionable	100%	very effective	
			2	100%	very effective	100%	very effective	
			3	100%	very effective	100%	very effective	
Hedin et al. (2011)	Within-case AB Design	Oral retell: number of main ideas	1	100%*	very effective	NC	1 data point	1.2 low
			2	100%*	very effective	NC	1 data point	
		Oral retell: overall quality (rubric scoring)	1	100%*	very effective	NC	1 data point	
			2	100%*	very effective	NC	1 data point	

Note. *calculated by reviewer; NC = not calculated due to a single baseline data point at ceiling level; PND = Percentage of Nonoverlapping Data

Table 10
 Summary of effect sizes from group studies

Author	Design	Outcome measure	Comparisons	ES estimate	Effect Size	Descriptor	WoE D		
Rogevich & Perin (2008)	Quasi-experimental non-equivalent groups pretest, posttest, follow-up design	Written summarisation	<u>Interaction Effect</u>	η^2	.42	moderate	1.9 medium		
			Task (T1/T2/T2n/T2f/T3) x Group						
			<u>Task</u>	<u>Post-hoc between group contrasts</u>					
			T2	BD/Intervention v. BD/Practice	d	4.5	large		
			(Posttest v. T1)	BD/Intervention v. BD+ADHD/Practice	d	4.3	large		
				BD+ADHD/Intervention v. BD/Practice	d	2.97	large		
				BD+ADHD/Intervention v. BD+ADHD/Practice	d	2.85	large		
			T2n (#Near transfer) v. T1	BD/Intervention v. BD+ADHD/Intervention	d	0.73	medium-large		
				BD/Intervention v. BD/Practice	d	3.51	large		
				BD/Intervention v. BD+ADHD/Practice	d	3.0	large		
				BD+ADHD/Intervention v. BD/Practice	d	2.01	large		
				BD+ADHD/Intervention v. BD+ADHD/Practice	d	1.71	large		
			T2f (##Far transfer) v. T1	BD/Intervention v. BD+ADHD/Intervention	d	1.49	large		
				BD/Intervention v. BD/Practice	d	5.73	large		
				BD/Intervention v. BD+ADHD/Practice	d	4.66	large		
				BD+ADHD/Intervention v. BD/Practice	d	2.18	large		
				BD+ADHD/Intervention v. BD+ADHD/Practice	d	1.89	large		
			T3. (Maintenance) v. T1	BD/Intervention v. BD+ADHD/Intervention	d	1.09	large		
				BD/Intervention v. BD/Practice	d	3.37	large		
				BD/Intervention v. BD+ADHD/Practice	d	3.02	large		
				BD+ADHD/Intervention v. BD/Practice	d	1.21	large		
				BD+ADHD/Intervention v. BD+ADHD/Practice	d	1.09	large		

Author	Design	Outcome measure	Comparisons	ES estimate	Effect Size	Descriptor	WoE D
Johnson et al. (1997)	Randomised groups pretest, posttest, follow-up design	Oral retell: proportion of main ideas	T2 Posttest v. T1 Pretest	d	1.11	large	1.9 medium
			T3 Maintenance v. T1 Pretest	d	1.00	large	
		Oral retell: proportion of story details	T2 Posttest v. T1 Pretest	d	1.02	large	
			T3 Maintenance v. T1 Pretest	d	.95	large	
		Oral retell: story grammar parts rating	T2 Posttest v. T1 Pretest	d	1.01	large	
			T3 Maintenance v. T1 Pretest	d	1.00	large	
Percentage correct in generalisation probes	T2 Posttest v. T1 Pretest	d	.40	small-medium			

Note. ADHD = attention deficit hyperactivity disorder; BD = behaviour disorder; ES = Effect size; # Near Transfer = generalisation as measured by summarisation of a social science passage rather than a science passage; ## Far Transfer = generalisation as measured by summarisation and synthesis of information from two passages rather than one

Intervention effect

Overall, the studies seemed to support the effectiveness of SRSD in improving reading comprehension in the study populations. This effect is reliable as far as performance within the instructional context is concerned, with all studies documenting improvement immediately after intervention.

Both group studies produced statistically significant pre-posttest differences. Rogevich and Perin (2008) reported a moderate effect size of *group by task* interaction effect that was explained largely by superior performance of one of the groups that received SRSD instruction ($\eta^2 = .42$). Nevertheless, the extent of the improvement can be uneven across individuals, with effects ranging from “questionable” (PND = 50%) to “very effective” (PND = 100%) across participants and measures in Howorth et al. (2016). It is noted, however, that PND is known to be more conservative than other non-overlap metrics (Lenz, 2013), and therefore the likelihood of Type I error is small.

While conclusions cannot be drawn with confidence based on pre-post changes in Johnson et al. (1997), which was given a ‘low’ WoE B rating due to the lack of a control group that did not receive SRSD intervention, the research was able to demonstrate the educational significance of SRSD through its inclusion of a social validation group. Following instruction, the performance of children in all four intervention groups were indistinguishable when compared with a normative comparison group matched on age, gender, and ethnicity from the same schools. This suggests that SRSD may

enable struggling readers with SEN to catch up with their typically-achieving peers.

Maintenance and generalisation

The impact of SRSD was more varied when considering the maintenance of the effect, which was examined by all five studies. In Johnson et al. (1997), Johnson et al. (2012), and Howorth et al. (2016), participants continued to perform better than before they received the strategy instruction, with some performances even exceeding levels immediately after intervention. This pattern was not replicated in Hedin et al. (2011), which saw performance dropping several weeks after instruction. However, the latter may be given less weight as Hedin et al. (2011) received a 'low' WoE D rating, whereas all other studies were rated 'medium'. It may, therefore, be argued that there is overall support for a maintenance of intervention effect.

Rogevich and Perin (2008) and Johnson et al. (1997) both reported some evidence for generalisation and maintenance among participants who received intervention. Notably, the former study also found between group differences (BD only versus BD with ADHD). Although both groups made equal levels of improvement immediately after intervention, the latter group demonstrated less improvements in the generalisation tasks immediately after intervention ($d = .73$ and 1.49 respectively) and at the maintenance tasks 3 weeks after intervention ($d = 1.09$). In Johnson et al. (1997), participants' scores in the generalisation probe significantly improved after intervention, but the effect size ($d = .40$) was smaller than in other measures.

Hedin et al. (2011) was the only other study to have examined the generalisation of intervention effect. Participants were given an equivalent task, but was tested by an unfamiliar teacher to test for generalisation across contexts. Neither participants maintained the performance level during and immediately after intervention. However, generalisation was probed once so a stable trend cannot be established, and the 5-day gap between the end of intervention and the day of testing would have introduced the confounding factor of maintenance of skills over time. This was reflected in the study's 'low' WoE rating, and any conclusion based on this should be treated critically.

Conclusion and Recommendations

In conclusion, this review has identified promising evidence supporting the use of SRSD as an instructional approach to teach literacy strategies that address reading comprehension difficulties experienced by school-age pupils with SEN.

The conclusions that can be drawn about the reviewed findings is constrained by the difficulty to compare and contrast across studies that have used a diverse range of experimental methodologies, as well as by the 'medium' to 'low' ratings of their weight of evidence. However, the replication of the intervention effect observed in the each studies' measure of pupils' reading comprehension may lend support to the generality of SRSD's effectiveness in supporting school-age learners with different types of SEN attending a range of educational settings.

All of the studies have provided key participant characteristics which would allow readers to determine the applicability of the findings to the target groups. As all studies were carried out in the United States, the differences in curriculum and the educational systems should be considered when generalising the findings to the UK context. It is also noteworthy that only one study has included girls in their sample, which further limits the findings' relevance to female pupils, who represent 35% of the school population on SEN support in England (DfE, 2017).

Given that SRSD functions as a model that guides how strategy instructions can be delivered, but not what should be taught, conclusions about its effectiveness may not extend to other strategies that were not reviewed here. Four of the studies used a version of Mason's (2004) TWA strategy, hence the use of SRSD to teach this strategy can be recommended with more confidence. Similarly, while the educational relevance of the findings were enhanced with most studies' focus on comprehension of curriculum-related informational text, this review provides weaker support for SRSD's support for understanding narrative text given that only one study investigated this. The generalisability of SRSD's effectiveness can be enhanced by research systematically replicating the intervention effect across different strategy types, types of reading materials, age, and SEN groups.

An area that warrants further investigation relates to the mixed findings regarding the generalisation and maintenance. This is because the practical value of SRSD if improvements made within the instructional context cannot be transferred to everyday curriculum learning. A plausible explanation may be related to the time-based, rather than criterion-based model of delivery. As a result of that, the participants may not necessarily have mastered and internalised the strategies before the intervention's end.

That explanation, however, would not account for the differences observed between participants with and without a comorbid diagnosis of ADHD in Rogevich and Perin (2008). This difference may suggest that the effectiveness of SRSD is moderated by the characteristic needs of some

SEN groups. For example, challenges commonly experienced by learners with ADHD or ASD in thinking flexibly, sustaining attention, and planning and organisation (Corbett, Constantine, Hendren, Rocke, & Ozonoff, 2009) may interfere with their abilities to apply the newly learned literacy skills in new situations, and maintaining gains over time. In light of this, EPs should consider recommending this approach in combination with other evidence-informed practices specific to the needs of the pupil. Future research can also fruitfully explore the interactive relationship between specific SEN groups and the effectiveness of SRSD, as this may help identify ways that the approach can be adapted to suit particular characteristic needs.

Although the effect sizes and quality of evidence vary across studies, the educational significance of their findings is enhanced by the ease of implementation, and the central importance of the targeted skills to curriculum learning. The availability of manualised protocol and lesson plans, low running costs, and the relatively small amount of training required suggest that SRSD has potential to be easily implemented.

Another important consideration for the feasibility of school implementation is the learners' and instructors' acceptance of the intervention. Most pupils reported finding the strategies useful in improving understanding of the text, and many were able to identify aspects of the intervention that were helpful for them. This lends support to the affective mechanism by which SRSD supports the learner's skill development. Whilst self-report measures may be confounded by social desirability bias, the low attrition rates (n=5; 4% overall)

across the studies suggests that SRSD is sufficiently motivating for learners across age groups and presenting needs.

The lacking of the teachers' perspectives is significant, however, as it largely determines the everyday usage in schools. This may be particularly pertinent in evaluating the use of SRSD in the UK context, as some may consider certain elements of the intervention "too 'American'" (Torgerson et al., 2014, p.26). With only one teacher-implementer identified in the reviewed study, there is a clear need for more engagement of school staff in future research evaluating the use of SRSD (British Educational Research Association, 2014). School-led research may also have the unique advantage of being able to track the intervention's long-term impact, which was not investigated in the reviewed studies. Given their professional expertise in integrating theory, research and practice, EPs are uniquely placed to work in collaboration with schools to further develop the evidence base for such promising and educationally significant interventions as SRSD in the UK educational context.

References

- Asaro-Saddler, K. (2016). Writing Instruction and Self-Regulation for Students With Autism Spectrum Disorders A Systematic Review of the Literature. *Topics in Language Disorders*, 36(3), 266–283. <https://doi.org/10.1097/TLD.0000000000000093>
- Barker, C., Pistrang, N., & Elliott, R. (2015). *Research methods in clinical psychology : an introduction for students and practitioners* (3rd ed.). Chichester, West Sussex: Wiley Blackwell.
- Brannen, J. (1992). *Mixing methods : qualitative and quantitative research*. Aldershot ; Brookfield: Avebury.
- British Educational Research Association. (2014). *Research and the Teaching Profession: Building the capacity for a self-improving education system*. London. Retrieved from <https://www.bera.ac.uk/wp-content/uploads/2013/12/BERA-RSA-Research-Teaching-Profession-FULL-REPORT-for-web.pdf>
- Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-Subject Experimental Design for Evidence-Based Practice. *American Journal of Speech-Language Pathology*, 21(4), 397–414. [https://doi.org/10.1044/1058-0360\(2012/11-0036\)](https://doi.org/10.1044/1058-0360(2012/11-0036))
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112(1), 155–159.
- Corbett, B. A., Constantine, L. J., Hendren, R., Rocke, D., & Ozonoff, S. (2009). Examining executive functioning in children with autism spectrum disorder, attention deficit hyperactivity disorder and typical development. *Psychiatry Research*, 166(2–3), 210–22. <https://doi.org/10.1016/j.psychres.2008.02.005>
- Department for Education. (2017). National Statistics: Special educational needs in England: January 2017 (SFR37/2017). Retrieved February 18, 2018, from <https://www.gov.uk/government/statistics/special-educational-needs-in-england-january-2017>
- Department for Education, & Department of Health. Special educational needs and disability code of practice: 0 to 25 years, Pub. L. No. DFE-00205-2013 (2015). <https://doi.org/DFE-00205-2013>
- 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. <https://doi.org/10.1080/02671520701296189>
- Graham, S., & Harris, K. R. (1993). Self-Regulated Strategy Development: Helping Students with Learning Problems Develop as Writers. *The Elementary School Journal*, 94(2), 169–181. <https://doi.org/10.1086/461758>

- Hagaman, J. L., & Casey, K. J. (2017). Paraphrasing Strategy Instruction in Content Area Text. *Intervention in School and Clinic, 52*(4), 210–217. <https://doi.org/10.1177/1053451216659468>
- Harris, K. R., & Graham, S. (1996). *Making the writing process work : strategies for composition and self-regulation* (2nd ed.). Cambridge, MA: Brookline.
- Harris, K. R., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. In *British Journal of Educational Psychology Monograph Series II, Number 6 - Teaching and Learning Writing* (pp. 113–135). <https://doi.org/10.1348/978185409X422542>
- Harris, K. R., Graham, S., Mason, L. H., & Friedlander, B. (2008). *Powerful writing strategies for all students*. Baltimore; London: Brookes.
- Hedin, L. R., Mason, L. H., & Gaffney, J. S. (2011). Comprehension Strategy Instruction for Two Students with Attention-Related Disabilities. *Preventing School Failure, 55*(3), 148–157. <https://doi.org/10.1080/1045988X.2010.499393>
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education. *Exceptional Children, 71*(2), 165–179. <https://doi.org/10.1177/001440290507100203>
- Howorth, S., Lopata, C., Thomeer, M., & Rodgers, J. (2016). Effects of the TWA strategy on expository reading comprehension of students with autism. *British Journal of Special Education, 43*(1), 39–59. <https://doi.org/10.1111/1467-8578.12122>
- Jitendra, A. K., Petersen-Brown, S., Lein, A. E., Zaslofsky, A. F., Kunkel, A. K., Jung, P.-G., & Egan, A. M. (2015). Teaching Mathematical Word Problem Solving. *Journal of Learning Disabilities, 48*(1), 51–72. <https://doi.org/10.1177/0022219413487408>
- Johnson, J. W., Reid, R., & Mason, L. H. (2012). Improving the Reading Recall of High School Students With ADHD. *Remedial and Special Education, 33*(4), 258–268. <https://doi.org/10.1177/0741932511403502>
- Johnson, L., Graham, S., & Harris, K. R. (1997). The effects of goal setting and self-instruction on learning a reading comprehension strategy: A study of students with learning disabilities. *Journal of Learning Disabilities, 30*(1), 80–91. <https://doi.org/10.1177/002221949703000107>
- Kratochwill, T. R. (2003). *Task Force on Evidence Based Interventions in School Psychology*. American Psychological Association.
- Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 3*(4), 445–463. <https://doi.org/10.1080/09602011.2013.815636>

- Lenz, A. S. (2013). Calculating Effect Size in Single-Case Research: A Comparison of Nonoverlap Methods. *Measurement and Evaluation in Counseling and Development*, 46(1), 64–73.
<https://doi.org/10.1177/0748175612456401>
- Logan, S., Medford, E., & Hughes, N. (2011). The importance of intrinsic motivation for high and low ability readers' reading comprehension performance. *Learning and Individual Differences*, 21(1), 124–128.
<https://doi.org/10.1016/J.LINDIF.2010.09.011>
- Mason, L. H. (2004). Explicit Self-Regulated Strategy Development Versus Reciprocal Questioning: Effects on Expository Reading Comprehension Among Struggling Readers. *Journal of Educational Psychology*, 96(2), 283–296. <https://doi.org/10.1037/0022-0663.96.2.283>
- Mason, L. H. (2013). Teaching Students Who Struggle With Learning to Think Before, While, and After Reading: Effects of Self-Regulated Strategy Development Instruction. *Reading & Writing Quarterly*, 29(2, SI), 124–144. <https://doi.org/10.1080/10573569.2013.758561>
- Mastropieri, M. A., Scruggs, T. E., Graetz, J. E., & Graet, J. E. (2003). Reading Comprehension Instruction for Secondary Students: Challenges for Struggling Students and Teachers. *Learning Disability Quarterly*, 26(2), 103–116. <https://doi.org/10.2307/1593593>
- Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, Rimes, Vocabulary, and Grammatical Skills as Foundations of Early Reading Development: Evidence From a Longitudinal Study. *Developmental Psychology*, 40(5), 665–681.
<https://doi.org/10.1037/0012-1649.40.5.665>
- Nation, K., Clarke, P., Wright, B., & Williams, C. (2006). Patterns of reading ability in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 36(7), 911–919.
<https://doi.org/10.1007/s10803-006-0130-1>
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining Nonoverlap and Trend for Single-Case Research: Tau-U. *Behavior Therapy*, 42, 284–299. <https://doi.org/10.1016/j.beth.2010.08.006>
- Reid, R., Hagaman, J. L., & Graham, S. (2014). Using Self-Regulated Strategy Development for Written Expression with Students with Attention Deficit Hyperactivity Disorder. *Learning Disabilities: A Contemporary Journal*, 12(1), 21–42.
<https://doi.org/10.1177/001440290607300103>
- Rogevich, M. E., & Perin, D. (2008). Effects on science summarization of a reading comprehension intervention for adolescents with behavior and attention disorders. *Exceptional Children*, 74(2), 135–154.
<https://doi.org/10.1177/001440290807400201>
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing Single-Subject Research. *Behavior Modification*, 22(3), 221–242.
<https://doi.org/10.1177/01454455980223001>

- Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, Present, and Future Trends in Summarizing Single-Subject Research. *Remedial and Special Education, 34*(1), 9–19.
<https://doi.org/10.1177/0741932512440730>
- Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology, 81*, 1–23.
<https://doi.org/10.1111/j.2044-8279.2010.02014.x>
- Snyder, S. M., Knight, V. F., Ayres, K. M., Mims, P. J., & Sartini, E. C. (2017). Single Case Design Elements in Text Comprehension Research for Students with Developmental Disabilities. *Education and Training in Autism and Developmental Disabilities, 52*(4), 405–421.
- Sreckovic, M. A., Knowles, M. M., & Lane, K. L. (2014). A Review of Self-Regulated Strategy Development for Writing for Students with EBD. *Behavioral Disorders, 39*(2), 56–77.
<https://doi.org/10.1177/019874291303900203>
- Torgerson, D., Torgerson, C., Ainsworth, H., Buckley, H., Heaps, C., Hewitt, C., & Mitchell, N. (2014). *Improving Writing Quality: Evaluation Report and Executive Summary*. London. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Evaluation_Report_-_Improving_Writing_Quality.pdf
- Whalon, K., Al Otaiba, S., & Delano, M. E. (2009). Evidence-Based Reading Instruction for Individuals With Autism Spectrum Disorders. *Focus on Autism and Other Developmental Disabilities, 24*(1), 3–16.
<https://doi.org/10.1177/1088357608328515>
- What Works Clearinghouse. (2017). *WWC Intervention Report: Self-Regulated Strategy Development*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_srsd_111417.pdf
- Wilson, D. B. (n.d.). Practical Meta-Analysis Effect Size Calculator [Online calculator]. Retrieved April 7, 2018, from <https://www.campbellcollaboration.org/effect-size-calculato.html>

Appendix A

Articles excluded at full text screening

Excluded Studies	Excluded by criteria:
Hagaman, J. L., Casey, K. J., & Reid, R. (2016). Paraphrasing Strategy Instruction for Struggling Readers. <i>Preventing School Failure</i> , 60(1), 43–52.	9. Fewer than half of all participants have an eligible SEN, or the proportion is not specified
Ilter, I. (2017). Improving the Reading Comprehension of Primary-School Students at Frustration-Level Reading through the Paraphrasing Strategy Training: A Multiple-Probe Design Study. <i>International Electronic Journal of Elementary Education</i> , 10(1), 147–161.	9. Fewer than half of all participants have an eligible SEN, or the proportion is not specified
Hagaman, J. L., Casey, K. J., & Reid, R. (2012). The Effects of the Paraphrasing Strategy on the Reading Comprehension of Young Students. <i>Remedial and Special Education</i> , 33(2), 110–123.	5. None of the conditions follow the SRSD model
Li, M., Murphy, P. K., Wang, J., Mason, L. H., Firetto, C. M., Wei, L., & Chung, K. S. (2016). Promoting reading comprehension and critical-analytic thinking: A comparison of three approaches with fourth and fifth graders. <i>Contemporary Educational Psychology</i> , 46, 101–115.	6. The outcome measures do not evaluate the pupils’ reading comprehension skills
McKeown, D., Brindle, M., Harris, K. R., Graham, S., Collins, A. A., & Brown, M. (2016). Illuminating Growth and Struggles Using Mixed Methods: Practice-Based Professional Development and Coaching for Differentiating SRSD Instruction in Writing. <i>Reading and Writing</i> , 29(6), 1105–1140.	9. Fewer than half of all participants have an eligible SEN, or the proportion is not specified

Excluded Studies	Excluded by criteria:
<p>Nevo, E., Brande, S., & Shaul, S. (2016). The Effects of Two Different Reading Acceleration Training Programs on Improving Reading Skills of Second Graders. <i>Reading Psychology</i>, 37(4), 533–546.</p>	<p>4. The article does not present an original research that uses an experimental or quasi-experimental design</p>
<p>Niedo, J., Lee, Y.-L., Breznitz, Z., & Berninger, V. W. (2014). Computerized Silent Reading Rate and Strategy Instruction for Fourth Graders at Risk in Silent Reading Rate. <i>Learning Disability Quarterly</i>, 37(2), 100–110.</p>	<p>9. Fewer than half of all participants have an eligible SEN, or the proportion is not specified</p>
<p>Roohani, A., & Asiabani, S. (2015). Effects of Self-Regulated Strategy Development on EFL Learners’ Reading Comprehension and Metacognition. <i>GEMA Online Journal of Language Studies</i>, 15(3), 31–49.</p>	<p>9. Fewer than half of all participants have an eligible SEN, or the proportion is not specified</p>
<p>Mason, L. H., Meadan-Kaplansky, H., Hedin, L., & Taft, R. (2013). Self-Regulating Informational Text Reading Comprehension: Perceptions of Low-Achieving Students. <i>Exceptionality</i>, 21(2), 69–86.</p>	<p>9. Fewer than half of all participants have an eligible SEN, or the proportion is not specified</p>
<p>Hagaman, J. L., Casey, K. J., & Reid, R. (2012). The Effects of the Paraphrasing Strategy on the Reading Comprehension of Young Students. <i>Remedial and Special Education</i>, 33(2), 110–123.</p>	<p>5. None of the conditions follow the SRSD model</p>
<p>Hagaman, J. L., Luschen, K., & Reid, R. (2010). The “RAP” on Reading Comprehension. <i>TEACHING Exceptional Children</i>, 43(1), 22–29.</p>	<p>5. None of the conditions follow the SRSD model</p>
<p>Lewis, W. E., & Ferretti, R. P. (2009). Defending interpretations of literary texts: The effects of topoi instruction on the literary arguments of high school students. <i>Reading & Writing Quarterly</i>, 25(4), 250–270.</p>	<p>4. The article does not present an original research that uses an experimental or quasi-experimental design</p>

Excluded Studies	Excluded by criteria:
Mason, L. H., Benedek-Wood, E., & Valasa, L. (2009). Teaching Low-Achieving Students to Self-Regulate Persuasive Quick Write Responses. <i>Journal of Adolescent & Adult Literacy</i> , 53(4), 303–312.	6. The outcome measures do not evaluate the pupils' reading comprehension skills
Patel, P., & Laud, L. (2009). Helping Students to Add Detail and Flair to Their Stories. <i>Preventing School Failure</i> , 54(1), 2–10.	6. The outcome measures do not evaluate the pupils' reading comprehension skills
Hagaman, J. L., Casey, K. J., & Reid, R. (2012). The Effects of the Paraphrasing Strategy on the Reading Comprehension of Young Students. <i>Remedial and Special Education</i> , 33(2), 110–123.	4. The article does not present an original research that uses an experimental or quasi-experimental design
Meadan, H., & Mason, L. H. (2007). Reading Instruction for a Student with Emotional Disturbance: Facilitating Understanding of Expository Text. <i>Beyond Behavior</i> , 16(2), 18–26.	9. Fewer than half of all participants have an eligible SEN, or the proportion is not specified
Mason, L. H., Snyder, K. H., Sukhram, D. P., & Kedem, Y. (2006). TWA plus PLANS strategies for expository reading and writing: Effects for nine fourth-grade students. <i>Exceptional Children</i> , 73(1), 69–89.	6. The outcome measures do not evaluate the pupils' reading comprehension skills
Mason, L. H. (2004). Explicit self-regulated strategy development versus reciprocal questioning: Effects on expository reading comprehension among struggling readers. <i>Journal of Educational Psychology</i> , 96(2), 283–296.	4. The article does not present an original research that uses an experimental or quasi-experimental design
Lauterbach, S. L., & Bender, W. N. (1995). Cognitive strategy instruction for reading comprehension: A success for high school freshmen. <i>The High School Journal</i> , 79(1), 58–64.	9. Fewer than half of all participants have an eligible SEN, or the proportion is not specified

Appendix B

Summary of studies included in the review

Study	Research design	Participants	SRSD intervention details	Outcome measures	Key findings
Howorth et al. (2016)	<p>Single-case design: Multiple-baseline across participants</p> <p>Phases of data collection: T1 Baseline T2 Intervention T3 Maintenance (2 week)</p>	<p>Sample size: 4 Gender: 100% male Age: 10-11 (Grades 5-6) SEN: ASD</p> <p>Country: US Setting: mainstream & special schools</p>	<p>Reading strategy: “Think before reading, think while reading, and think after reading” (TWA)</p> <p>Reading Material: Expository text – science and social studies passages from website</p> <p>SRSD protocol: Based on Mason et al. (2012)</p> <p>Intervention delivery: 1:1 by researcher 6 x 45 minute sessions (criterion-based)</p>	<p>Oral retell</p> <ul style="list-style-type: none"> scored by researcher-developed rubric <p>Comprehension questions</p> <ul style="list-style-type: none"> researcher-developed questions based on passages <p>Intervention acceptability (child)</p>	<p>Effect of intervention</p> <ul style="list-style-type: none"> At T1, participants had slightly variable, but non-trended baselines. One participant reached ceiling level at one data point for comprehension question. Intervention effect on oral retell was observed through generally higher but variable levels at T2 (PND = 50-100%); and a strong maintenance was demonstrated at T3 with all data points staying above baseline level, and some performance exceeding T2. This may suggest that participants need time to consolidate strategies learned. Intervention effect appeared to be more consistent in reading comprehension questions at T2; performance remained slightly better at T3 than T1.

Study	Research design	Participants	SRSD intervention details	Outcome measures	Key findings
Johnson et al. (2012)	<p>Single-case design: Multiple-baseline across participants, with multiple probes</p> <p>Phases of data collection: T1 Baseline T2a Intervention T2b Postintervention T3 Maintenance (2-4 weeks)</p>	<p>Sample size: 3 Gender: 100% male Age: 14-15 (Grade 9) SEN: P1. - ADHD P2. - ADHD + SLI P3. - ADHD + LD Country: US Setting: mainstream school</p>	<p>Reading strategy: "Think before reading, think while reading, and think after reading" (TWA)</p> <p>Reading Material: Expository text – social studies passages from textbooks</p> <p>SRSD protocol: Based on Harris et al. (2008)</p> <p>Intervention delivery: 1:1 by researcher 6 – 7 x 30 minute sessions (criterion-based)</p>	<p>Oral retell</p> <ul style="list-style-type: none"> • number of main ideas • percentage of supporting details 	<p>Effect of intervention</p> <ul style="list-style-type: none"> • The level of recall of main ideas and supporting details were low and stable in all participants at T1 • At T2a, there were immediate and marked improvement in recall • This effect is maintained at T2b and T3, where levels were stable and remained above baseline • The replication of intervention effect across participants suggest that there is a true and reliable effect of SRSD on reading comprehension
Hedin et al. (2011)	<p>Single-case design: AB design</p> <p>Phases of data collection: T1 Baseline T2 Intervention Single data points for: <ul style="list-style-type: none"> • Post-intervention • 5-day delay • Generalisation • Maintenance (4-8 weeks) </p>	<p>Sample size: 2 Gender: 100% male Age: 10-11 (Grade 4-5) SEN: P1: ADHD + SLI P2: ADHD + LD Country: US Setting: mainstream</p>	<p>Reading strategy: "Think before reading, think while reading, and think after reading" (TWA) + prompted discourse (oral questioning)</p> <p>Reading Material: Expository text – science passages from textbook</p> <p>SRSD protocol: Based on Mason (2004) etc.</p> <p>Intervention delivery: 1:1 by teacher 10 x 30 minute sessions (time-based)</p>	<p>Oral retell</p> <ul style="list-style-type: none"> • number of main ideas • overall quality - scored by rubric adapted from previous research (main ideas, details, and organisation) <p>Intervention acceptability (child)</p>	<p>Effect of intervention</p> <ul style="list-style-type: none"> • Both participants' performance at T2 on both measures were above baseline level, which may suggest an intervention effect* • The improvement did not appear to be maintained or generalised, with some 5-day delay and maintenance measures returning close to pre-intervention level <p><i>*These findings need to be interpreted with caution due to the lack of acceptable experimental control through multiple demonstrations of effect, an upward baseline trend for one participant, and only single data point measures for follow-up measurements</i></p>

Study	Research design	Participants	SRSD intervention details	Outcome measures	Key findings
Rogevich & Perin (2008)	<p>Group-based design: Quasi-experimental non-equivalent groups pretest, posttest, follow-up design</p> <p>Time points of assessment: T1 Pre-intervention T2 Post-intervention T2n Post-intervention: near transfer T2f Post-intervention: far transfer T3 Maintenance (3 weeks)</p>	<p>Sample size: 63 Gender: 100% male Age: 13-16 SEN: BD (n=32) BD + ADHD (n=31)</p> <p>Comparison groups: Intervention / BD Intervention / BD+ADHD Control / BD Control / BD+ADHD</p> <p>Country: US Setting: residential treatment facilities for young people with behavioural and emotional needs</p>	<p>Reading strategy: “Think before reading, think while reading, and think after reading” with Written Summarisation (TWA-WS)</p> <p>Reading Material: Expository text – science passages from textbook (& social studies passages for generalisation task)</p> <p>SRSD protocol: Based on Mason (2004) & Mason et al. (2006)</p> <p>Intervention delivery: Group (n=3-4) by researcher 5 x 45 minute sessions (time-based)</p>	<p>Written summarisation</p> <ul style="list-style-type: none"> • number of main ideas • includes generalisation assessment of <i>near transfer</i> (reading social studies text) and <i>far transfer</i> (summarising from 2 text sources) <p>Intervention acceptability (child)</p>	<p>Effect of intervention</p> <ul style="list-style-type: none"> • Both Intervention groups performed significantly better in all post-intervention measures (T2, T2-N, T2-F, T3) than Control groups, suggesting that the strategies are effective, generalisable, and can be maintained • This is demonstrated by significant interaction effect (group x task). • Effect size was moderate $\eta^2 = .42$ <p>Effect of diagnoses</p> <ul style="list-style-type: none"> • No difference between BD and BD+ADHD Intervention groups for T2 • but BD Intervention group made more gains than BD+ADHD Intervention group at T2n, T2f, T3, suggesting the latter group be less able to generalise strategy use and maintain intervention gains

Study	Research design	Participants	SRSD intervention details	Outcome measures	Key findings
Johnson et al. (1997)	<p>Group-based design: Randomised groups pretest, posttest, follow-up design</p> <p>Time points of assessment: T1 Pre-intervention T2 Post-intervention T3 Maintenance (4 weeks)</p>	<p>Sample size: 52 (47 at end of study) Gender: 100% male Age: 13-16 SEN: LD</p> <p>Comparison groups: SRSD SRSD + GS SRSD + SI SRSD + GS + SI</p> <p>Country: US Setting: residential treatment facilities for young people with behavioural and emotional needs</p>	<p>Reading strategy: Story grammar strategy</p> <p>Reading Material: Stories from previous research</p> <p>SRSD protocol: Based on Bednarczyk (1991)</p> <p>Intervention delivery: Group (n=2–3) by graduate students Varying number of 45 minute sessions (criterion-based)</p>	<p>Oral retell</p> <ul style="list-style-type: none"> percentage of main ideas details story grammar parts <p>Classroom generalisation</p> <ul style="list-style-type: none"> multiple choice questions <p>Intervention acceptability (child)</p>	<p>Effect of Intervention</p> <ul style="list-style-type: none"> All groups made significant improvement for main ideas, details, and total story grammar recalled at both T2 and T3, suggesting that the effect is maintained over time All participants also improved in generalisation measure at T2 <p>Social comparison</p> <ul style="list-style-type: none"> Despite performing significantly worse than typically-achieving pupils at T1, all intervention groups performed at same level as this social comparison group after intervention at T2 <p>Effects of additional components</p> <ul style="list-style-type: none"> No difference between conditions, suggesting that goal-setting and self-instruction provide no incremental increase over SRSD instruction without these components.

Note. ADHD = attention deficit hyperactivity disorder; ASD = autism spectrum disorder; BD = behavioural disorders; GS + goal-setting; LD = learning disabilities; PND = Percentage of nonoverlapping data, an effect size measure for single-case designs; SRSD = Self-Regulated Strategy Development; SI = self-instruction

Appendix C

Criteria and the Rationale for all WoE Ratings

Weight of Evidence A: Methodological Quality

WoE A is a generic judgement about the quality of the evidence based on generally accepted criteria for evaluation of the given types of studies (Gough, 2007). In this review, the WoE A ratings for the five included studies were determined using adapted versions of the Kratochwill (2003) coding protocol for group-based (pp. 45-62) and single-case designs (pp. 84-113). Details of the amendments made to both protocols are listed with rationale in Appendix D. Examples of the completed coding protocols are also provided in Appendices E and F.

Five areas of consideration from Section II of both protocols were used to determine the overall WoE A ratings for the included studies (see table below). Within each area of consideration, numerical ratings were given based on Kratochwill's (2003) coding manual. Ratings ranged from 0 to 3, with 0 indicating 'no evidence', 1 indicating 'weak evidence', 2 indicating 'promising evidence', and 3 indicating 'strong evidence'. These ratings are then averaged to produce an overall WoE A rating.

Summary of Weight of Evidence A ratings of the included studies

Study	A. Measure- ment	B. Baseline / Compariso n Group	E. Identifiable Componen t	F. Implement -ation Fidelity	I. Follow Up Assessme nt Conducted	Overall WoE A
Howorth et al. (2016)	3	3	1	2	3	2.4
Johnson et al. (2012)	3	3	1	2	2	2.2
Hedin et al. (2011)	1	3	1	2	2	1.8
Rogevic h & Perin (2008)	3	3	1	1	2	2.0
Johnson et al. (1997)	3	3	3	1	2	2.4

Weight of Evidence B: Methodological Relevance

WoE B is a review-specific judgement of the appropriateness of the study's methodology in addressing the review question. In the context of this review, WoE B considers the extent to which the methodological design is appropriate for evaluating the impact of SRSD in improving reading comprehension for children with special educational needs.

Different sets of criteria are specified for group-based and single case designs. The evaluated studies would need meet all of the relevant sets of criteria to achieve a particular rating. These criteria are based on Brannen's (1992) evidence hierarchies. In general, group-based research designs that have taken steps to minimise threats to

internal validity (e.g. by including an active control group and making use of randomisation in a group-based design) are rated more highly in the hierarchy, and are therefore given more weight in this WoE dimension. It is also recognised that while some validity issues are unavoidable in single-case experiments, some aspects of the research design can improve the external validity of the findings (e.g. by demonstrating intervention effects over multiple baselines across a number of participants or settings). These are reflected in the criteria differentiating between medium and low ratings for single-case designs.

Additionally, the maintenance of any intervention gains is of particular interest in this review, given the aim of the intervention is for the student to continue to use the strategies to support their reading comprehension after they were instructed. It was therefore important for studies to include a follow-up measure of their primary outcome variables to receive a high or medium WoE B rating.

Criteria for Weight of Evidence B ratings

WoE B ratings	Criteria	
	<u>Group-based design</u>	<u>Single-case design</u>
3 (high)	<ul style="list-style-type: none"> The study has an active control group that are receiving an alternative literacy intervention. Participants are randomly allocated to an intervention or control conditions. Outcome measures are collected at pre- and post-intervention and follow-up Sample size is adequate for all statistical analyses 	

WoE B ratings	Criteria	
2 (medium)	<ul style="list-style-type: none"> The study has a control group that are not receiving SRSD instruction, but may not be receiving an alternative intervention. Participants may not be randomly allocated to intervention or control conditions. Outcome measures are collected at pre- and post-intervention and follow-up Sample size may be lower than required for statistical analyses. 	<ul style="list-style-type: none"> The study includes at least three attempts to demonstrate an intervention effect at different points in time or across different participants Outcome measures are collected at baseline, intervention, and maintenance phases. Outcome measures are taken over at least three data points in all phases
1 (low)	<ul style="list-style-type: none"> The study may not have a control group that is not receiving SRSD instruction. Outcome measures are collected at pre- and post-intervention 	<ul style="list-style-type: none"> The study includes at least three attempts to demonstrate an intervention effect at different points in time or across different participants Outcome measures are taken over at least three data points at baseline and intervention phase
0 (no evidence)	<ul style="list-style-type: none"> None of the criteria are met 	<ul style="list-style-type: none"> The study reports fewer than three attempts to demonstrate an intervention effect

Summary of Weight of Evidence B ratings of the included studies

Study	WoE B
Howorth et al. (2016)	2
Johnson et al. (2012)	2
Hedin et al. (2011)	0
Rogevich & Perin (2008)	2
Johnson et al. (1997)	1

Weight of Evidence C: Topic Relevance

WoE C is a review-specific judgement about the extent to which the focus of the evidence is relevant to the review question. This WoE dimension evaluates whether the studies' contextual information, learner, implementer, and setting characteristics, research procedures, and the use of generalisation allow the findings to be generalised to answer the review question. The following criteria were used to give a rating for WoE C in this review:

Criteria for Weight of Evidence C ratings

Area of consideration	Criteria	Rationale				
A. Participant selection and characteristics	<table border="1"> <tr> <td data-bbox="448 1066 762 1122"><u>Relevant participant characteristics:</u></td> <td data-bbox="770 1066 1110 1279"> <ul style="list-style-type: none"> • Home language • Socio-economic background • Levels of general cognitive abilities • Levels of reading achievement </td> </tr> <tr> <td data-bbox="448 1155 762 1279"> <ul style="list-style-type: none"> • Age / school year • Gender • SEN label • Ethnicity </td> <td></td> </tr> </table>	<u>Relevant participant characteristics:</u>	<ul style="list-style-type: none"> • Home language • Socio-economic background • Levels of general cognitive abilities • Levels of reading achievement 	<ul style="list-style-type: none"> • Age / school year • Gender • SEN label • Ethnicity 		The participants and inclusion criteria need to be described in sufficient detail for readers to determine generalisability to their intended clients, and also for studies to be replicable
<u>Relevant participant characteristics:</u>	<ul style="list-style-type: none"> • Home language • Socio-economic background • Levels of general cognitive abilities • Levels of reading achievement 					
<ul style="list-style-type: none"> • Age / school year • Gender • SEN label • Ethnicity 						
	3 The study clearly describes its inclusion criteria, with reference to the participants' SEN category and reading difficulties.					
	The article provides at least 6 of the relevant demographic and learning characteristics from the list above. (see WoE A coding protocol item III.A2)					
	2 The study clearly describes its inclusion criteria, which may not reference SEN and reading difficulties.					
	The article provides at least 4 of the relevant demographic and learning characteristics from the list above.					
	1 The study may not explicitly state its inclusion criteria, but provides at least 4 of the relevant demographic and learning characteristics from the list above.					
	0 The study does not meet the above criteria, or does not provide information about the pupils' SEN.					

Area of consideration	Criteria	Rationale
B. Setting generalisability	3 The participants were sampled from more than one school.	Most pupils with SEN attend mainstream or special schools. The findings will have higher external validity if they are replicated across multiple settings.
	2 The participants were sampled from one school.	
	1 The participants were sampled from an alternative provision (e.g. pupil referral unit / treatment facilities).	
	0 The participants were not attending any educational setting at the time of the study.	
C. Implementer	3 SRSD instruction was implemented by a staff member of the educational setting.	The intervention is more likely to be adopted if staff implementation is shown to be feasible without specialist personnel or equipment at the setting where it is hoped that it would be delivered
	2 SRSD instruction was implemented by external personnel (e.g. member of the research team) in the educational setting.	
	1 SRSD instruction was implemented by an external instruction outside of the premise of the educational setting	
	0 The identity or characteristics of the SRSD instructor is not clear.	
D. Identifiable SRSD components	3 The study clearly documents the essential SRSD components with replicable instructional steps (or makes clear reference to existing protocols it has closely followed). Procedures for adapting the intervention are described in detail and justified.	SRSD is a versatile framework for strategy instruction, so it may be implemented in a variety of ways. High quality reporting of the implementation protocol allows readers to make an informed judgement of the applicability of the findings, and also enables future research to replicate the studies.
	2 The study provides a clear outline of lesson structure, the number and duration of sessions. Procedures for adapting the intervention are outlined.	
	1 The article provides some information about the intervention procedures, but there is insufficient details to determine the level of support given, or for the study to be replicated.	
	0 No description of the intervention procedures is provided.	

Area of consideration		Criteria	Rationale
E.Generalisation of skills	3	The study has attempted to evaluate the generalisation of effect across tasks <i>and</i> contexts at multiple time points.	The transfer of the strategies learned outside of the instructional context is key to ensure the continued use by the learner in novel situations.
	2	The study has attempted to evaluate the generalisation of effect across tasks <i>and</i> contexts at least once.	
	1	The study has attempted to evaluate the generalisation of effect across tasks <i>or</i> contexts at least once.	
	0	The study has not attempted to evaluate the generalisation of the strategies instructed	

Summary of Weight of Evidence C ratings of the included studies

Study	A. Participant	B. Setting	C. Implementer	D. SRSD components	E. Generalisation	Overall WoE C
Howorth et al. (2016)	2	3	2	2	0	1.8
Johnson et al. (2012)	2	2	2	3	0	1.8
Hedin et al. (2011)	1	2	3	2	1	1.8
Rogevich & Perin (2008)	3	1	2	3	0	1.8
Johnson et al. (1997)	3	3	2	1	2	2.2

Weight of Evidence D: Overall Weight of Evidence

The three WoE dimensions were amalgamated to provide an overall appraisal of the evidence presented in the study in relation to the review question. Each dimension was given equal weighting, and scores from WoE A, B, and C for each study were averaged to produce an overall weight of evidence (WoE D) rating.

WoE ratings are described as 'High' for average scores of 2.5 or above, 'Medium' for scores between 1.5 and 2.4, and 'Low' for scores of 1.4 or below.

To obtain a high overall weight of evidence, a study would need to be methodologically sound, and have a research question and methodology that are closely relevant to the review topic. Thus, a well-designed and methodologically robust study may still be given a low WoE rating in relation to the review question if its topic and methodology are deemed to have limited relevance to the focus of the review.

Weight of evidence of included studies

Authors	WoE A: Methodological quality	WoE B: Methodological relevance	WoE C: Topic relevance	WoE D: Overall weight of evidence
Howorth et al. (2016)	2.4	2	1.8	2.1 medium
Johnson et al. (2012)	2.2	2	1.8	2.0 medium
Hedin et al. (2011)	1.8	0	1.8	1.2 low

Rogevich & Perin (2008)	2.0	2	1.8	1.9 medium
Johnson et al. (1997)	2.4	1	2.2	1.9 medium

WoE ratings are described as 'High' for average scores of 2.5 or above, 'Medium' for scores between 1.5 and 2.4, and 'Low' for scores of 1.4 or below. Table 5 provides a summary of the WoE scores for the reviewed studies.

Appendix D

Amendments to the Kratochwill (2003) coding protocols

The criteria for Kratochwill (2003) coding protocols for group-based and single-case designs were amended to suit the purpose of the current review. This section provides information about the changes made, and the rationale behind them. These changes were applied to both protocols unless otherwise specified.

Modified section	Rationale
<i>Removed from coding protocol</i>	
I. B8 and B9 (<i>Single-case protocol</i>) & I. B7 and B8 (<i>Group-based protocol</i>) - Coding for qualitative research methods	Only quantitative data were considered in this review.
II. C - Measures Support Primary and Secondary Outcomes	The outcomes are considered separately as part of the review process (see effect size table)
II. D – Educational / Clinical Significance	The educational significance of the outcomes are considered separately in this review
II. G – Replication	All studies, regardless of whether they are replications, are analysed in this review.
II. H - Site of Implementation	All studies were conducted in an educational setting, as per inclusion criteria. The relevance of the studies in relation to the type of provision is also discussed separately in the review.
III. C & D - Dosage	This information is not applicable to the reviewed intervention.
III. F - Characteristics of the Intervener	No information was reported in any of the included studies in relation to the interveners' race, gender, SES etc.

Modified section	Rationale
III. H – Cost Analysis Data	Cost information is not reported in any of the included studies.
III. I & J – Training and Support Resources; Feasibility	Only one study (Hedin et al., 2011) reported involving a typical staff member as implementer of the intervention. No information is provided training and resources required. The implication for feasibility is discussed in the main review.
<i><u>Modification to existing sections</u></i>	
II. F – Fidelity of implementation (<i>Single-case protocol</i>)	This is evaluated on study-level rather than individually for each participant. This is because all included single-case studies delivered the same interventions to all participants, and the fidelity of intervention was reported across all participants, and not on an individual basis.
III. A2-3. Participant Characteristics Specified for Treatment and Control Group	This section is simplified into a checklist of the types of information that have a bearing on the inclusion criteria and outcomes. A summary of relevant participant information is provided in Appendix D.

Appendix E

Coding Protocol for Single-Participant Design

Adapted from Kratochwill, T. R. (2003). *Task Force on Evidence Based Interventions in School Psychology*. American Psychological Association

- Domain:**
- School- and community-based intervention programs for social and behavioural 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: 8/2/2018

Full Study Reference in APA format:

Howorth, S., Lopata, C., Thomeer, M., & Rodgers, J. (2016). Effects of the TWA strategy on expository reading comprehension of students with autism. *British Journal of Special Education*, 43(1), 39–59

Intervention Name (description from study):

Self-Regulated Strategy Development (SRSD):
using the “Think before reading, think during reading, think after reading” (TWA) strategy

Study ID Number (Unique Identifier): 1

- Type of Publication:**
- Book/Monograph
 - Journal article
 - Book chapter
 - Other (specify):

I. General Characteristics

A. General Design Characteristics (Classify studies according to the type of design)

A1. Type of Single-Participant Design (select A1.1, A1.2, A1.3, A1.4, or A1.5)

A1.1 Within-series design (select A1.1.1 or A1.1.2)

A1.1.1 Simple phase change

A1.1.2 Complex phase change

A1.2 Between-series design (select A1.2.1 or A1.2.2)

A1.2.1 Comparing two interventions

A1.2.2 Comparing interventions with no interventions

A1.3 Combined-series design (select A1.3.1., A1.3.2, A1.3.3, or A1.3.4)

A1.3.1 Multiple baseline across participants

A1.3.2 Multiple baseline across behaviours

A1.3.3 Multiple baseline across settings

A1.3.4 Multiple probe design

A1.4 Mixed design (select A1.4.1 or A1.4.2)

A1.4.1 Combined single-participant and group design (see group manual),

A1.4.2 Combined single-participant design (if combined single-participant design, check A1.4.2.1, A1.4.2.2, or A1.4.2.3)

A1.4.2.1 Within-series design (select i or ii)

i. Simple phase change

ii. Complex phase change

A1.4.2.2 Between-series design (select i or ii)

i. Comparing two interventions

ii. Comparing interventions with no interventions

A1.4.2.3 Combined-series design (select i, ii, iii, or iv)

i. Multiple baseline across participants

ii. Multiple baseline across behaviours

iii. Multiple baseline across settings

iv. Multiple probe design

A1.5 Other (specify):

B. Other Design Characteristics (when randomization is used)

B1. Unit of assignment to conditions/groups (select one of the following)

B1.1 Individual

B1.2 Classroom

B1.3 School

B1.4 Other (specify):

B1.5 N/A (randomization not used)

B2. Type of assignment to conditions/groups (select one of the following)

B2.1 Random after matching, stratification, blocking

B2.2 Random, simple (includes systematic sampling)

B2.3 Nonrandom, post hoc matching

B2.4 Nonrandom, other

B2.5 Other (specify):

B2.6 Unknown/insufficient information provided

B2.7 N/A (randomization not used)

B3. Overall confidence of judgment on how participants were assigned to conditions/groups (select one of the following)

B3.1 Very low (little basis)

B3.2 Low (guess)

B3.3 Moderate (weak inference)

B3.4 High (strong inference)

B3.5 Very high (explicitly stated)

B3.6 N/A (randomization not used)

B3.7 Unknown/unable to code

B4. Equivalence of conditions/groups tested at pretest (select one of the following)

B4.1 Yes

B4.2 No

B4.3 Unknown/insufficient information provided

B4.4 N/A (randomization not used)

B5. Total size of sample (start of the study): 4

B6. Intervention sample size ____ N/A (randomization not used)

B7. Control sample size ____ N/A (randomization not used)

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
(3=Strong Evidence 2=Promising Evidence1=Weak Evidence0=No Evidence)

A. Measurement: Issues of Reliability and Validity (answer A1. through A4.)

A1. Use of outcome measures that produce reliable scores (select one of the following)

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

Inter-rater reliability of two trained doctoral student raters was reported at baseline (0.9) and intervention phase (1.0)

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

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

Outcome measures included oral retell, comprehension questions, and student self-report of satisfaction with the

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

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

Measures were obtained from the interventioners/raters and the pupils

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

- A4.1 Yes

Content validity of the oral retell rubric was reviewed by three writing teachers and doctoral-level reading experts

A4.2 No

A4.3 Unknown/unable to code

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

B. Quality of Baseline.

Rate quality of baseline: (a) for each participant (when there is more than one participant), and (b) for each phase (when the study includes more than one phase). These procedures should be followed for each primary outcome under investigation.

Participant 1 (answer B1. through B5.)

B1. Length: At least 3 data points during baseline (select one of the following)

B1.1 Yes

B1.2 No

B1.3 Unknown/insufficient information provided

B2. Stability: Variability in scores does not eliminate the detection of treatment effects (select one of the following)

B2.1 Yes

B2.2 No

B2.3 Unknown/insufficient information provided

B3. Overlap: Extreme scores during baseline do not overlap with most scores during intervention phase (select one of the following)

B3.1 Yes

B3.2 No

B3.3 Unknown/insufficient information provided

B4. Level: Behaviour is serious enough during baseline to warrant an intervention (select one of the following)

B4.1 Yes

B4.2 No

B4.3 Unknown/insufficient information provided

B5. Trend: Behaviour is not systematically increasing or decreasing in the desired direction of intervention effects during baseline.

B5.1 Yes

B5.2 No

B5.3 Unknown/insufficient information provided

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

Participant 2 (answer B1. through B5.)

B1. Length: At least 3 data points during baseline (select one of the following)

B1.1 Yes

B1.2 No

B1.3 Unknown/insufficient information provided

B2. Stability: Variability in scores does not eliminate the detection of treatment effects (select one of the following)

B2.1 Yes

B2.2 No

B2.3 Unknown/insufficient information provided

B3. Overlap: Extreme scores during baseline do not overlap with most scores during intervention phase (select one of the following)

B3.1 Yes

B3.2 No

B3.3 Unknown/insufficient information provided

B4. Level: Behaviour is serious enough during baseline to warrant an intervention (select one of the following)

B4.1 Yes

B4.2 No

B4.3 Unknown/insufficient information provided

B5. Trend: Behaviour is not systematically increasing or decreasing in the desired direction of intervention effects during baseline.

B5.1 Yes

B5.2 No

B5.3 Unknown/insufficient information provided

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

Participant 3 (answer B1. through B5.)

B1. Length: At least 3 data points during baseline (select one of the following)

B1.1 Yes

B1.2 No

B1.3 Unknown/insufficient information provided

B2. Stability: Variability in scores does not eliminate the detection of treatment effects (select one of the following)

B2.1 Yes

B2.2 No

B2.3 Unknown/insufficient information provided

B3. Overlap: Extreme scores during baseline do not overlap with most scores during intervention phase (select one of the following)

B3.1 Yes

B3.2 No

B3.3 Unknown/insufficient information provided

B4. Level: Behaviour is serious enough during baseline to warrant an intervention (select one of the following)

B4.1 Yes

B4.2 No

B4.3 Unknown/insufficient information provided

B5. Trend: Behaviour is not systematically increasing or decreasing in the desired direction of intervention effects during baseline.

B5.1 Yes

B5.2 No

B5.3 Unknown/insufficient information provided

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

Participant 4 (answer B1. through B5.)

B1. Length: At least 3 data points during baseline (select one of the following)

B1.1 Yes

B1.2 No

B1.3 Unknown/insufficient information provided

B2. Stability: Variability in scores does not eliminate the detection of treatment effects (select one of the following)

B2.1 Yes

B2.2 No

B2.3 Unknown/insufficient information provided

B3. Overlap: Extreme scores during baseline do not overlap with most scores during intervention phase (select one of the following)

B3.1 Yes

B3.2 No

B3.3 Unknown/insufficient information provided

B4. Level: Behaviour is serious enough during baseline to warrant an intervention (select one of the following)

B4.1 Yes

B4.2 No

B4.3 Unknown/insufficient information provided

B5. Trend: Behaviour is not systematically increasing or decreasing in the desired direction of intervention effects during baseline.

B5.1 Yes

B5.2 No

B5.3 Unknown/insufficient information provided

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

Average Quality of Baseline Rating Across Participants:

$$\frac{\sum \text{of } X}{N} = 3$$

X = individual quality of baseline ratings for each participant
N = number of participants in the study

Overall Rating for Quality of Baseline: (select 0, 1, 2, or 3):
 3 2 1 0

(Round up or down to the nearest whole number when providing a mean rating for the study. For example, 2.0 to 2.4 rated as 2; 2.5 to 2.9 rated as 3).

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: *The intervention was delivered as a whole package*

E4. Number of components linked to primary outcomes: *n/a*

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 Fidelity (select 0, 1, 2, or 3): 3 2 1 0

I. Follow Up Assessment

Timing of follow up assessment: *Three maintenance measurements were conducted two weeks after each participant reached mastery.*

Number of participants included in the follow up assessment: *All four participants were assessed at follow-up*

Consistency of assessment method used: *The participants were given new expository text passages of equivalent instructional level to read; the same outcome measures (oral retell and comprehension questions) as in baseline and intervention phases were used.*

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: *Inclusion criteria (based on age, IQ and ASD diagnostic testing, and discrepancy between decoding and reading comprehension skills) were suitable for the research question; the rationale was explained.*

Specify rationale for sample size: *Not specified; opportunity sampling of child participants who were available and fit the inclusion criteria.*

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 Treatment and Control Group
(*modified*)

Age / school year

Gender

SEN diagnostic label

Ethnicity

Home language

Socio-economic background

Levels of general cognitive abilities (e.g. IQ) *WISC-IV*

Levels of reading achievement *Woodcock Johnson III Tests of Achievement*

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

Participants from Treatment Group	Results (What person reported to have gained from participation in programme)	General rating
<input checked="" type="checkbox"/> Child/student <input type="checkbox"/> Parent/carer <input type="checkbox"/> Teacher	<i>All participants indicated that "the TWA strategy helped them become</i>	<input checked="" type="checkbox"/> Participants reported benefited overall from the intervention

<input type="checkbox"/> School <input type="checkbox"/> Other	<i>better readers” (p.52) and each identified elements of the strategy that were most helpful for them.</i>	<input type="checkbox"/> Participants reported not benefitting overall from the intervention
---	---	--

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: *The participants’ improvements in reading comprehension were sustained in the maintenance phase (see Section II. I – Follow Up Assessment)*

A5.1.2 Procedures for maintaining outcomes are specified yes no

Specify: *The participants’ reading comprehension were assessed using the same measures on new text of equivalent level. (see Section II. I – Follow Up Assessment)*

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: *This study has used a relatively homogenous group of children who have the same diagnostic label and similar reading achievement levels, and of the same age range. No information is provided about the children’s linguistic, cultural, and socio-economic background. Thus the findings by themselves provides limited evidence to support the efficacy of SRSD to support reading comprehension in other groups of participants.*

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: ____

B2.2 months: ____

B2.3 years: ____

B2.4 other: ____

6 x 45 min sessions (Intervention was ended only when the mastery criterion was reached)

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
- 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
 - F3. Different from target participants on key variables
- No relevant information provided in article*

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

- G1. Behavioural
- G2. Cognitive-behavioural
- G3. Experiential
- G4. Humanistic/interpersonal
- G5. Psychodynamic/insight oriented
- G6. Other (specify): _____
- G7. Unknown/insufficient information provided

Summary of Evidence for Single-Participant Design Studies

Indicator	Overall Evidence Rating	Description of Evidence
	NNR = No numerical rating	Strong Promising Weak No/limited evidence
	or	or
	0 - 3	Descriptive ratings
Key Areas of Judgement for Weight of Evidence A		
Measurement	3	
Baseline	3	
Identifiable Components	1	
Implementation Fidelity	2	
Follow Up Assessment Conducted	3	

Average Quality of Evidence across the Included Judgement Areas
$\frac{\sum \text{ of } X}{N} = \frac{3 + 3 + 1 + 2 + 3}{6} = 2.4$ <p>X = individual quality of evidence for each judgement area N = number of judgement areas</p> <p>Overall Rating for Weight of Evidence A: 2.4 (medium)</p>

Appendix F

Coding Protocol: Group-Based Design

Based on Kratochwill, T. R. (2003). *Task Force on Evidence Based Interventions in School Psychology*. American Psychological Association

- Domain:**
- School- and community-based intervention programs for social and behavioural 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: 5 / 2 / 2018

Full Study Reference in APA format:

Rogevich, M. E., & Perin, D. (2008). Effects on science summarization of a reading comprehension intervention for adolescents with behavior and attention disorders. *Exceptional Children*, 74(2), 135–154.

Intervention Name (description from study):

Self-Regulated Strategy Development (SRSD):
using the “Think Before Reading, Think During Reading, Think After Reading, with Written Summarisation” (TWA-WS) strategy

Study ID Number (Unique Identifier): 4

- Type of Publication:**
- 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) *participants were matched on age, IQ, and reading level, with near-equal numbers between groups*
- 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 *Bonferroni corrections applied for post-hoc pairwise contrast*
- B3. Sufficiently large *N* yes no
Statistical Test: *Repeated measures MANCOVA*
_ level: *0.05*
ES: *Large*
N required: *18 per group (for 4 groups)*
- B4. Total size of sample (start of the study): *63*

B5. Intervention group sample size: 15 + 15

B6. Control group sample size: 17 + 16

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 Evidence1=Weak Evidence0=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 *Observer agreement: inter-rater reliability at 95% (99% after discussion of discrepancy)*

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 *Participants in control conditions discussed and wrote summaries of passages, completed multiple-choice and true/false tests, but were not taught any strategies*
- 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

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 _____

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: *Intervention delivered as a whole package*

E4. Number of components linked to primary outcomes: *n/a*

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

I. Follow Up Assessment

- Timing of follow up assessment: *3 weeks*
- Number of participants included in the follow up assessment: *63 (all participants)*
- Consistency of assessment method used: *written summarisation (scored by researcher-developed rubric) is used for all conditions and time points*

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: *specified in participant section (p. 139)*
Specify rationale for sample size: *limited due to difficulties seeking parental consent, and the exclusion of large numbers of pupils due to exclusion criteria of other comorbid diagnoses*
- 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 Treatment and Control Group
(modified)

- Age / school year
- Gender
- SEN diagnostic label
- Ethnicity
- Home language
- Socio-economic background *information provided for whole setting but not for each group*
- Levels of general cognitive abilities (e.g. IQ) *WISC-III*
- Levels of reading achievement *Gates-MacGinite Reading Test*

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes yes no

Specify: *age, IQ, and reading scores were correlated to outcomes*

A3.2 Have relevance to inclusion criteria yes no

Specify: *diagnosis and IQ scores were part of inclusion criteria*

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

Participants from Treatment Group	Results (What person reported to have gained from participation in programme)	General rating
<input checked="" type="checkbox"/> Child/student <input type="checkbox"/> Parent/carers <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other	<i>Highly positive in both intervention groups; most participants described specific steps they found useful</i>	<input checked="" type="checkbox"/> Participants reported benefited overall from the intervention <input type="checkbox"/> Participants reported not benefitting overall from the intervention

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: ____

B2.2 months: ____

B2.3 years: ____

B2.4 other: ____

5 x 45 minute sessions (End of intervention was time-based rather than criterion-based)

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

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

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

Summary of Evidence for Group-Based Design Studies

Indicator	Overall Evidence Rating	Description of Evidence
	NNR = No numerical rating or 0 - 3	Strong Promising Weak No/limited evidence or Descriptive ratings
Key Areas of Judgement for Weight of Evidence A		
Measurement	3	
Comparison Group	3	
Identifiable Components	1	
Implementation Fidelity	1	
Follow Up Assessment Conducted	2	

Average Quality of Evidence across the Included Judgement Areas
$\frac{\sum \text{ of } X}{N} = \frac{3 + 3 + 1 + 1 + 2}{6} = 2.0$ <p>X = individual quality of evidence for each judgement area N = number of judgement areas</p> <p>Overall Rating for Weight of Evidence A: 2.0 (medium)</p>