Case Study 1: An Evidence-Based Practice Review Report

Theme: School/setting Based Interventions for Learning.

How effective is Self-Regulated Learning strategy instruction at improving mathematical performance in students with maths difficulties or at risk of developing maths difficulties?

Summary

Effective teaching of maths skills is of fundamental educational importance (OFSTED, 2012) as these skills are applied extensively within everyday life (Cargnelutti, Tomasetto, & Passolunghi, 2017). Therefore, if a child is identified as experiencing maths difficulties, it is important that they are provided with targeted support to mediate the impact of these difficulties on their later life outcomes (Nelson & Powell, 2018). Self-Regulated Learning (SRL) strategy instruction was originally developed to improve literacy skills, (Graham & Harris 1996) but has also been developing as an intervention to enhance maths performance (Lee & McDonough, 2015). This review evaluated the effectiveness of SRL strategy instruction as an intervention to improve the maths performance of students with, or at risk of developing, maths difficulties. Taken together, the findings provide some promising evidence for the use of SRL to improve some areas of maths performance amongst the target population. However, due to methodological limitations, these conclusions must be viewed with caution. The review recommends the

need for further research, specifically exploring the different components of SRL instruction most strongly related to improved performance, along with the need for more longitudinal research before any firm conclusions regarding effectiveness can be made.

Introduction

Mathematical skills

Maths is considered as one of the most important subjects taught at school due to its wide application in everyday life (Cargnelutti et al., 2017). Within the context of schools, research has demonstrated that children with good maths skills appear to be at an academic advantage both within primary and secondary school (Dahl, Aubrey, & Godfrey, 2006; Hooper, Roberts, Sideris, Burchinal, & Zeisel, 2010). In addition, longitudinal studies have found that early maths skills are one of the strongest predictors of later achievement (Duncan, Schmitt, Burke, & McClelland, 2018) and greater outcomes in adult life (Parsons & Bynner, 2005). Thus, it is no surprise that developing mathematical skills is of fundamental importance within the national curriculum.

Within the United Kingdom (U.K.), an Office for Standards in Education, Children's Services and Skills (OFSTED) report, inspecting mathematics in maintained schools, highlighted concerning inequalities between pupils' experiences and achievements in maths (OFSTED, 2012). Specifically, the report outlined that pupils who started their education with lower mathematical skills continued to remain behind and never 'caught up' with the mathematical knowledge and understanding of their peers. The report

called on schools to take action to close this gap and raise national mathematical ambition (OFSTED, 2012).

Approximately 2-8% of school aged children experience a diagnosed mathematical difficulty (e.g. dyscalculia) (Nelson & Powell, 2018). However, within the literature, mathematical difficulty (MD) is considered much more broadly to include those with low maths attainment scores (Ennis & Losinski, 2019). Therefore, the true number of children experiencing a MD may be much higher. Children experiencing MDs are likely to represent some of the children outlined in the OFSTED (2012) report as 'lagging behind' in terms of their maths attainment.

Although MDs are considered to be persistent, a review by Nelson and Powell (2018) highlighted that early identification and targeted interventions have the potential to reduce the risk of poor outcomes, not only in school, but also in later adult life. This suggests that developing effective methods of improving the mathematical skills and understanding of students presenting with MDs should be a priority within schools.

Educational psychologists (EPs) are well positioned to support schools to develop their understanding of what particular interventions are most strongly related with increasing maths attainment. One of the ways EPs can achieve this, is by systematically reviewing the evidence base through means such as a literature review.

Within the literature, maths intervention research has mainly focused on cognitive approaches to improve mathematical skills (Lucangeli, Fastame, Duca, Hitchcott & Penna, 2019). However, exploring cognitive approaches

alone, does not acknowledge the other components, such as self-regulated learning, that contribute to effective maths performance (Mevarech, 2006). Therefore, the aim of this review is to evaluate the evidence base for maths interventions that focus on the development of self-regulated learning strategies in students with MDs or are at risk of developing MDs.

Self-Regulated Learning

Clark and Zimmerman (1990) believe self-regulated learning (SRL) to be a social cognitive model in which effective learning involves the application of multiple processes including; self-instruction, questioning, monitoring, reinforcement and evaluation.

It is thought that the processes involved in SRL underpin the executive processes and functions involved in metacognition (Montague, 2008). More specifically, although one component of metacognition involves knowing what we know (Flavell, 1979), the other component relies on the ability to control, monitor and evaluate this knowledge. This component is thought to be best facilitated through SRL (Montague, 2008).

The benefits of SRL have been highlighted in studies showing that the application of SRL strategies can positively predict academic achievement across a range of subject domains (Adams, Forsyth, Dollarhide, Miskell, & Ware, 2015; Leon, Nunez, & Liew, 2015; Musso, Boekaerts, Segers, & Cascallar, 2019). As a result, teachers have been encouraged to use SRL strategy instruction as a way of improving student learning outcomes (Ness & Middleton, 2012)

Self-Regulated Learning Strategy Instruction

Self-Regulated Learning strategy instruction can take various forms. However, the primary goal is always to intentionally increase the student's awareness of their thinking, the support strategies that enable them to be an effective learner and the wider application of these skills to other tasks (Ness & Middleton, 2012). In order to achieve this, the core teaching components of SRL strategy instruction are often broken down into three phases; planning, performance and self-evaluation (Ness & Middleton, 2012). The aim of explicitly teaching the phases, is to encourage SRL to become an internalised process for the student.

Table 1

The three phases of SRL strategy instruction (Ness & Middleton, 2012)

Phase	Description
Planning	Instruction focuses on encouraging the student to bring their awareness to ways that support them to be 'ready to learn'. This may include providing them with a prompt sheet to think about sitting appropriately, having their pen and paper ready, looking towards the target stimuli
Performance	Instruction is centred on modelling, the promotion of problem solving strategies and independent practice
Evaluation	Instruction focuses on encouraging the student to self-appraise, think about what helps them to learn and what they could try the next time

A more specific 6 stage SRL strategy teaching framework was developed by Graham and Harris (1996) known as, Self-Regulated Strategy Development (SRSD) and is outlined in Table 2.

Table 2

Stages of SRSD (Graham & Harris, 1996)

Stage	**Description**
Developing and activating background knowledge	Learners prior knowledge is established and guided support is offered where necessary
Discuss the strategy	The strategy is explicitly explained including benefits and expectations
Cognitive modelling of the strategy	Active modelling and application of the task by the lead adult
Memorising the strategy	Learners practice the strategy until fluency is achieved, this can be supported through the use of mnemonics
Collaborative support of the strategy	Learners are supported to use the strategy
Independent practice	Teachers promote independent practice of using the strategy to complete tasks

In terms of SRL strategy instruction as an intervention, much of the research has explored it in relation to developing literacy skills (Cheng, 2016; Harris & Graham, 2009) and although SLR strategy instruction as an intervention for

improving maths performance is developing, (Lee & McDonough, 2014) it appears to have received less attention.

The main focus of the review is to evaluate the use of SRL strategy instruction to improve maths performance in students with MDs or at risk of developing MDs. Although the review will evaluate some studies that utilise the SRSD approach, the review is not restricted to only evaluating SRSD studies and will explore the effectiveness of SRL instruction more broadly.

When this review was conducted, there had been one similar systematic literature review published in 2008 (Montauge, 2008). That review concluded that cognitive strategy instruction to improve maths performance in students with learning difficulties, did qualify as evidence-based practice, so long as it was delivered by an expert SEN teacher, in small groups and for an intense time-limited period. Although insightful in its own right, the researcher felt there was a need to provide an up to date review in light of more recent literature.

Review Question

How effective is Self-Regulated Learning strategy instruction at improving mathematical performance in students with maths difficulties or at risk of developing maths difficulties?

Critical Review of the Evidence Base

Literature Search

A systematic literature search was carried out on the 16th of December 2019. Three different databases were used for the search (PsycINFO, ERIC and Web of Science). Various terms, used for the search, are listed in Table 3.

Table 3

List of search terms used in the systematic literature search

Database	Search terms
Web of Science & EBSCO	"SELF-REGULATED LEARNING" OR "SLR" OR "SELF-REGULATED STRATEGY DEVELOPMENT" OR "SRSD" OR "SELF-REGULATION" OR SELF-REGULATION TRAINING" AND "MATH*" OR "MATHEMATIC*" OR "CALCU*" OR "ARITHMETIC" OR "NUMERACY" AND "SCHOOL AGED CHILDREN" OR "SCHOOL CHILDREN" OR "ELEMENTARY" OR "SECONDARY" OR "PRIMARY" OR "STUDENTS" OR "PUPILS" OR "CHILDREN" PR "ADOLESCENT*" OR "TEENAGER*" OR "TEEN*" AND "SEN" OR "SPECIAL EDUCATIONAL NEEDS" OR "LEARNING DISABILITY" OR "LEARNING DIFFICULTY" OR "LEARNING NEEDS" OR "DYSCALCULIA" OR "MATH* DIFFICULTY"
PsychINFO	"SELF-REGULATED LEARNING" or "SLR" or "SELF-REGLATED STRATEGY DEVELOPMENT" or "SRSD"AND "MATH*" or "MATHEMATIC*"AND "SCHOOL AGED CHILDREN" or "SCHOOL CHILDREN"AND "SEN" or "SPECIAL EDUCATIONAL NEEDS" or "LEARNING DIFFICULTY" or "MATHS DIFFICUTLY" or "LEARNING DISABILITY"

Screening of articles

The literature search yielded a total of 348 studies. After duplicates were removed the final number of remaining studies was 328. The studies were screened using inclusion and exclusion criteria, outlined in Table 4. Figure 1 depicts the process of exclusion. The rationale for exclusion at full text review is provided in Appendix A. The included studies are listed in Table 5, with further details being provided in Appendix B.

Table 4

Inclusion and Exclusion criteria for the literature search

	Inclusion Criteria	Exclusion Criteria	Rationale
1. Publication type	Studies published in peer reviewed journals.	Not in a peer reviewed journal e.g. a book chapter, conference papers, doctoral thesis or dissertation.	Articles within peer reviewed journals have undergone scrutiny which ensures research integrity.
2. Publication date	Studies published from 2009 onwards.	Studies published before 2009.	This review aims to evaluate research from within the last 10 years.
3. Type of study	Intervention studies that include the evaluation of the effectiveness of teaching students self-regulated learning strategies on maths performance.	Intervention studies that do not include the evaluation of the effectiveness of teaching students self-regulated learning strategies on maths performance.	The review aims to evaluate the effectiveness of teaching self-regulatory strategies on maths performance.
4. Participants age	Participants are of English schooling age, including post GCSE age (between 4 to 19 years).	Participants are younger than age 4 or older than age 19.	The review is evaluating evidence relating to school aged children.
5. Participants learning needs	Participants with maths difficulties or are at risk of developing maths difficulties.	Studies including participants with no maths difficulties and are not at risk of developing maths difficulties.	The review aims to evaluate the effectiveness of self-regulatory teaching strategies on maths performance amongst students with maths difficulties or students at risk of developing maths difficulties.

	Inclusion Criteria	Exclusion Criteria	Rationale
6. Intervention	At least one of the intervention conditions involves self-regulated learning strategy instruction.	None of the intervention conditions involves self-regulated strategy instruction.	The review aims to evaluate the effectiveness self-regulated strategy instruction.
7. Outcomes	Studies that use pre and post outcome data relating to mathematical performance.	Studies that do not use pre and post outcome data relating to mathematical performance.	The review aims to evaluate the effectiveness of self-regulatory learning strategy instruction on maths performance.
8. Setting	The intervention is delivered within an educational setting.	The intervention is not delivered in an educational setting.	The review aims to consider the effectiveness self-regulated learning strategy instruction within an educational setting.
9. Language and geographical context	Studies from any geographical context but are written in English.	Studies not written in English.	As the reviewer only speaks English and translation services are not available only using studies published in the English language ensures the paper can be accurately evaluated.

Figure 1: Flow diagram of literature search and article screening

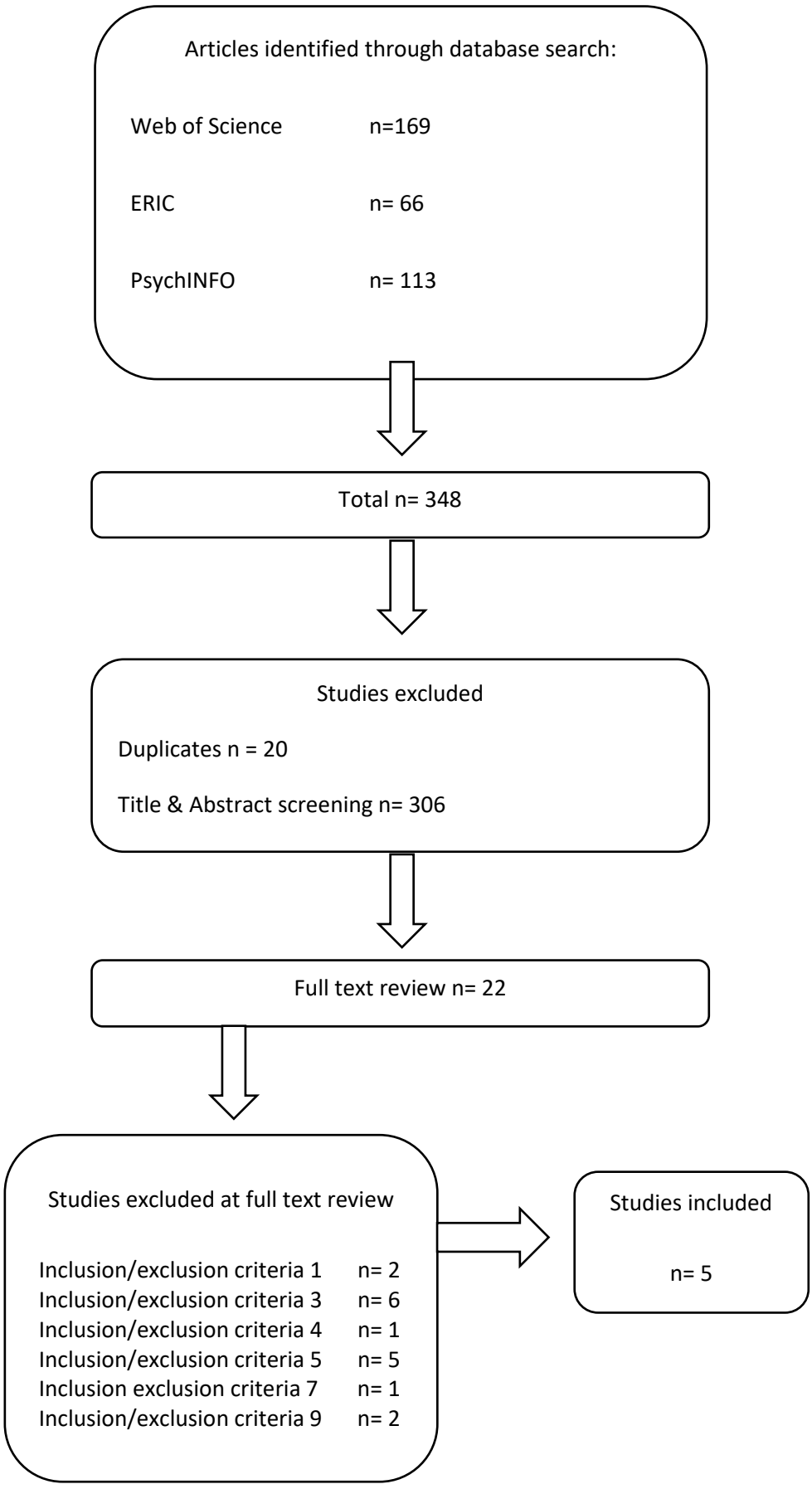

Table 5

List of studies included in the review

	Included studies
1	Cuenca-Carlino, Y., Freeman-Green, S., Stephenson, G. W., & Hauth, C. (2016). Self-regulated strategy development instruction for teaching multi-step equations to middle school students struggling in math. *The Journal of Special Education*, *50*(2), 75-85.
2	Ennis, R. P., & Losinski, M. (2019). SRSD Fractions: Helping Students at Risk for Disabilities Add/Subtract Fractions With Unlike Denominators. *Journal of learning disabilities*, *52*(5), 399-412.
3	Lucangeli, D., Fastame, M. C., Pedron, M., Porru, A., Duca, V., Hitchcott, P. K., & Penna, M. P. (2019). Metacognition and errors: the impact of self-regulatory trainings in children with specific learning disabilities. *ZDM*, *51*(4), 577-585.
4	Ness, B. M., & Middleton, M. J. (2012). A framework for implementing individualized self-regulated learning strategies in the classroom. *Intervention in School and Clinic*, *47*(5), 267-275.
5	Wang, A. Y., Fuchs, L. S., Fuchs, D., Gilbert, J. K., Krowka, S., & Abramson, R. (2019). Embedding self-regulation instruction within fractions intervention for third graders with mathematics difficulties. *Journal of learning disabilities*, *52*(4), 337-348.

Weight of Evidence

The selected studies were reviewed using Harden and Gough's (2012) Weight of Evidence (WoE) framework, evaluating them in terms of (a) Methodological quality (WoE A), (b) Methodological Relevance (WoE B) and (c) Topic Relevance (WoE C). Table 6 provides details regarding the WoE scores. Appendix C gives information relating to criteria and rationale for the scores assigned.

Table 6

WoE scores for A, B, C and D

Authors	WoE A Methodological quality	WoE B Methodological relevance	WoE C Topic relevance	WoE D Overall weight of evidence
Cuenca-Carlio et al. (2015)	1.8	2	2.6	2.1 Medium
Ennis and Losinski (2019)	1.6	2	2.4	2 Medium
Lucangeli et al. (2019)	0.6	2	2.6	1.7 Medium
Ness and Middleton (2012)	0	1	2.2	1.1 Low
Wang et al. (2019)	1.4	2	3	2.1 Medium

Note1: WoE ratings receive a rating 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.
Note 2: WoE D was calculated by summing each of the individual WoE A, B & C scores then calculating the average rating score.

Participants

Collectively, 152 participants took part across the 5 studies. However, in the Wang et al. (2019) study, 24 participants were involved in a 'base intervention' which did not include a SRL component. Therefore, a revised total of 132 participants were considered within this review.

The inclusion of group based studies, single case studies and a case study meant that there was a large variation in sample size, ranging from 1 (Ness & Middleton, 2012) to 68 participants (Lucangeli et al., 2019). In terms of statistical power, regarding the group studies, the Lucangeli et al.(2019) study was appropriately powered. However, the Wang et al. (2019) study was slightly underpowered.

Taking four of the studies collectively, there did not appear to be a gender bias. Specifically, the Lucangeli et al.(2019) study provided explicit details on how gender was counterbalanced across the groups. Conversely however, the Wang et al. (2019) study did not report on the gender demographics of the individual participants included in the study. Therefore, it is unclear if and how gender was counterbalanced.

The studies varied on whether they reported the participants' age or year grade. For the 4 studies that reported grade, these ranged from 2nd grade to 1st year in high school.

All participants involved in the studies, were identified as having MDs or being at risk of developing a MD Wang et al.(2019). As the studies were clear in explaining how participants were identified as having MDs, they all received the highest score in the WoE C section A, relating to the sample. Further details of this can be found in Appendix C.

Design

Two of the studies were group based designs. One of these Wang et al., (2019) used a block randomisation design whereas the other used a non-randomised block design Lucangeli et al. (2019). The use of block designs is appropriate as the participants were initially selected due to their MD. The use of randomisation in the Wang et al.,(2019) study is considered superior as randomisation reduces selection bias which promotes greater internal validity (Barker & Pistrang, 2016, p. 152).

The methodological quality (WoE A) of the group based designs were evaluated using Kratochwill's (2003) Group Based Design Protocol. Section II B of the protocol examines the comparison group used, which in both cases was a 'typical comparison'. The use of a typical contact comparison is appropriate for the purpose of evaluation as it allows for the intervention to be compared to teaching as usual. The Wang et al. (2019) study, received a 'promising' rating regarding the comparison group used compared to the Lucangeli et al.(2019) study which received a 'weak' comparison group rating. The reason for this difference was because the use of a 'typical contact' was more clearly stated in the Wang et al. (2019) paper. Neither study received the high comparison group rating, however, as they did not outline if or how they controlled for counterbalance of change. This impacts on the strength of the inferences that can be made about the intervention effectiveness.

Two other studies used single case experimental designs (SCEDs) (Cuenca-Carlino, Freeman-Green, Stephenson, & Hauth, 2016; Ennis & Losinski, 2019). More specifically, they used multiple baseline across participant designs. According to the Kratochwill's (2003) Single Participant Design coding protocol, this indicates a strong design as it replicates the intervention across participants. The staggering of the interventions across time also increases the internal validity. In addition, baseline data was collected at, at least 3 time points. For this reason, both SCEDs scored strongly in terms of baseline quality. In contrast to this, the case study design (Ness & Middleton,

2012) received a 'weak' baseline quality rating as baseline data was only collected at one point.

The SCEDs were rated higher than the other studies in relation to generalisation of findings as both provided follow-up data which allowed for an exploration of maintenance effects. However, because the follow-up data was not collected over multiple intervals, this meant their rating regarding 'follow up assessment' was not regarded as 'strong' and they received a 'promising' rating instead.

All of the studies WoE A ratings were negatively impacted by not analysing identifiable components of the SRL strategy instruction. This made it difficult to understand if there were particular components that were more strongly linked to change than others for example, the planning component versus the evaluation component.

Intervention content and fidelity

As this review was not restricted to evaluating studies that used a specific SRL strategy intervention, the studies differed in terms of their procedure (See Table 7).

Table 7

Intervention procedures

Study	Instruction delivery	Intervention sessions	Delivered by
Cuenca-Carlino et al. (2015)	Pairs	16 (first and second pair)13 (third pair)	Qualified maths teacher
Ennis and Losinski (2019)	Small groups (3 students, 3 students, 2 students)	Unknown	Trained researcher

Study	Instruction delivery	Intervention sessions	Delivered by
Lucangeli et al. (2019)	Group	16 (1 x 1 hour session per week for 16 weeks)	Psychologists specialised in the treatment of specific learning disabilities
Ness and Middleton (2012)	Individual	Until it was evident that the individual had memorised the SRL strategy	Special education teacher
Wang et al. (2019)	Group	39 (3 x 35 minute sessions per week for 13 weeks)	Trained researchers

Due to the variability within procedures used, in order to effectively compare the five studies it was expected that they would use SRL strategy interventions that included the 3 components outlined by Ness and Middleton, (2012).

1. *Planning*- All of the interventions outlined a planning phase in which the participants were instructed to focus on 'getting ready to learn'.
2. *Performance*- Each study included a performance phase whereby the student were engaged in actively 'doing' the maths activity. However, the amount of description given to this phase varied in each of the studies. The studies that used the specific SRSD strategy instruction framework were explicit in outlining how the performance phase was implemented, including details on modelling and independent practice. The Lucangeli et al. (2019) study also provided clear details regarding independent practice. However, the other studies, Ness and Middleton, (2012) and Wang et al. (2019), gave less details on what the performance phase entailed.

3. *Evaluation*- Four of the studies were clear in outlining how participants were encouraged to evaluate their work. However, in the Ness and Middleton, (2012) study, although evaluation was mentioned as part of the intervention, the details of this were not evident to the reader.

The overall rating for the intervention components can be found the the WoE C summary in Appendix C. The only study not receiving a high rating was the Lucangeli et al. (2019) study as the SRL strategy instruction was not the primary intervention component.

Procedures for adherence to the intervention protocols were robust in the SCEDs and Wang et al.(2019) study, including measures such as ongoing supervision, recoding and coding of sessions and the use of a manual. This led to a high WoE A for implementation fidelity. Although the Lucangeli et al. (2019) study used supervision and a manual, the researchers did not provide details on how each SRL strategy instruction was adapted for individual students. This resulted in a weaker WoE A rating for implementation fidelity. The Ness and Middleton, (2012) study did not provide information on how the fidelity was monitored or maintained.

Outcome Measures

All of the studies used primary outcome measures that aimed to assess maths performance. However, the specific measures used by each study differed as the studies were either concerned with maths performance generally (Lucangeli et al., 2019; Ness & Middleton, 2012) or with a particular area of maths, such as fractions (Ennis & Losinski, 2019; Wang et al., 2019) or equations (Cuenca-Carlino et al., 2016).

The quality of the measures used were assessed within Kratochwill's (2003) group based and single participant design protocol. This revealed that the studies differed in terms of their quality of outcome measurement. The highest scoring study, regarding measurements used, was the Wang et al.

(2019) which used multiple methods of gaining fraction performance data that were both reliable and valid.

Studies that scored a 1 regarding measurement (Cuenca-Carlino et al., 2016; Ennis & Losinski, 2019) did so because they only used one source of data collection. The measures used in the Lucangeli et al. (2019) study, although valid and gained from multiple sources, did not provide details of their reliability which resulted in a low rating. The Ness and Middleton (2012) study received a weak measurement rating as maths performance was only measured via maths grade that did not report on reliability.

Findings

Where possible, the findings were primarily evaluated using effect size summaries which can be found in Table 10 and 11. Originally, effect sizes from the Lucangeli et al. (2019) study were reported as hedges g. Effect sizes from the Wang et al. (2019) study were reported as partial eta squared. However, for the purpose of comparison, all effect sizes were converted into Cohen's d (1992). This was achieved using the means and standard deviations reported within the papers, which were then inputted into the Campbell's online calculator to provide converted Cohen's d statistic. Table 8 provides descriptions of the meaning of these Cohen's d (1992) values.

Effect sizes for the SCED studies were calculated by examining the percentage of non-over lapping data and applying Scruggs and Mastropeiri's (1998) percentage of non-overlapping (PND) descriptors, which can be found in Table 9. The Ennis and Losinski (2019) paper stated the PND within their results write up. The Cuenca-Carlino study did not state their PND percentages so these were identified by examining the date points within the graphs.

Using pre and post scores between the control and SRL groups, the Lucangeli et al. (2019) study found that the SRL strategy intervention led to a significant improvement in performance across some, but not all, of the

maths outcomes when compared to the control group. Specifically, improvements were related to digit transcription and written calculation, with a medium effect. Considering that these are the areas where maths performance appear to rely most on literacy skills, it is plausible to hypothesise that these areas had the most positive effect as the SRL strategy instructional approaches were initially developed with a focus on improving literacy (Harris & Graham, 2009).

Wang el al. (2019) also used pre and post between group scores to analyse performance and found significant improvement in the SRL group compared to the controls in terms of their word problem solving and ordering fractions, with a large effect. In addition, the SRL group demonstrated greater improvements than the controls in relation to their multiplication of fractions and their scores on the NAEP, with a medium effect. As this study scored much more strongly than the Lucangeli et al. (2019) in relation to methodological quality (see appendix 3), it is plausible to assume that the findings from this study carry more weight. However, it is also important to consider that as this study was slightly underpowered, there is a possibility that there may have been further undetected intervention effects.

In addition, both of the findings from the group based studies are limited by the lack of information regarding maintenance data which brought down both of their overall WoE A, methodological quality, ratings. Similarly, their overall WoE B ratings were only considered to be within the 'medium' range indicating flaws within their methodological relevance (See Table 6). Therefore, conclusions regarding the longer-term impact cannot be made.

The 2 SCEDS chose to visually graph the improvements made, of which there were many, and with large effects. The study by Cuenca-Carlino et al. (2016) provided details of the Tau-U statistics which is a robust measure of effect size (Barker & Pistrang, 2016). However, due to insufficient details regarding pre and post scores in the Ennis and Losinski (2019) study, the Tau-U could not be calculated. Thus, for the purpose of comparison and, as

mentioned previously, PND was used as the main evaluative effect size measure.

Evaluation of PND data, demonstrated that in the Cuenca-Carlino et al., (2016) study, the SRL strategy intervention was 'very effective' for all participants at improving their multi-step equation problem solving, with maintenance of improvements above baseline remaining in most cases.

In the Ennis andLosinski (2019) study, PND analysis also showed the SRL intervention was 'very effective' at improving performance of adding and subtracting fractions with unlike denominators for 6 participants and was 'effective' for 1 participant. The improvements, above baseline, were also maintained post-intervention.

The above findings from the SCEDs do provide support for using a SRL strategy intervention to improve maths performance related to fraction and equation solving. However, it is important to note that both studies were constrained by their methodological quality (WoE A) and methodological relevance (WoE B) as neither study reached the threshold of 2.5 or above, which was needed in order for the author of this review to consider them as being of 'high' quality and 'highly' relevant. Therefore, the conclusions that can be drawn from the findings are limited. Similarly, as with all single case designs, the findings are further constrained by the small sample size. In addition, the interpretive strength of the findings would have been improved if the data analysis involved the calculation of reliable change index (RCI) score for each participant. This would allow for the identification of the point of change needed in order to be representing more than random error measurement (Barker & Pistrang, 2016). Unfortunately, this review was unable to calculate this RCI as there was insufficient information provided regarding the reliability and standard deviation of the outcome measures uses.

The case study Ness and Middleton, (2012) also proposed that the SRL intervention was successful at improving the participants maths grade.

However, this assertion must be viewed with caution given the many methodological flaws, which resulted in an overall WoE A rating of 0, indicating that the conclusions drawn from the study had limited or no evidence. This was primarily due to the poor baseline quality, the use of an unreliable outcome measure and the small sample size.

Table 8

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

Cohen's d	Descriptor
.8	Large
.5	Medium
.2	Small

Table 9

Effect size descriptor for percentage of non-overlapping data (PND) (Scruggs & Masteropieri, 1998)

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

Table 10

Effect size table for group based studies

			Intervention Group			**Control Group**					
Authors	Outcome measure	N	Pre M (SD)	Post M (SD)	N	Pre M (SD)	Post M (SD)	ED d*	P	ES Descriptor	WoE D
Lucangeli et al. (2019)	Digits transcription	34	3.7 (1.86)	4.67 (1.47)	34	3.09 (2.04)	3.62 (1.52)	0.70	.019	Medium	1.8 Medium
	Number ordering		6.97 (1.96)	7.73 (1.94)		6.41 (2.02)	7.15 (1.71)	0.32	.198	Small	
	Written calculation		4.09 (2.00)	5.82 (1.28)		4.5 (1.75)	4.88 (1.67)	0.63	.013	Medium	
	Mental calculation speed		126.42 (86.39)	90.85 (53.01)		100.41 (47.17)	104.15 (48.215)	-0.26	.286	Small	
	Mental calculation errors		2.67 (1.96)	1.58 (1.58)		2.03 (1.61)	2.12 (1.68)	-0.33	.166	Small	

			Intervention Group			Control Group					
Authors	Outcome measure	N	Pre M (SD)	Post M (SD)	N	Pre M (SD)	Post M (SD)	ED d*	P	ES Descriptor	WoE D
Wang et al. (2019)	0-1Number line	23	Not reported	0.23 (0.07)	23	Not reported	0.26 (0.10)	-0.35	.209	Small	2.1 Medium
	Word problems		Not reported	7.18 (4.12)		Not reported	3.76 (2.54)	.99	.007	Large	
	Multiplication		Not reported	20.62 (5.55)		Not reported	16.60 (7.21)	.62	.006	Medium	
	Ordering		Not reported	6.52 (3.55)		Not reported	1.57 (1.67)	1.78	<.001	Large	
	NAEP**		Not reported	10.61 (4.18)		Not reported	8.50 (3.73)	.53	.011	Medium	

**Effect size calculated using post mean and standard deviation scores for both groups.*

***NAEP is released fraction items from the National Assessment of Educational Progress*

Table 11

Effect size table for single case experimental design studies

		Intervention effect				
Author	Outcome measure	Participant	PND	Descriptor	Maintenance Achieved (Y/N)*	WoE D
Cuenca-Carlino et al. (2019)	Multi-step equation probes	1	100%	Very effective	Yes	2.3 Medium
		2	100%	Very effective	No	
		3	100%	Very effective	Yes	
		4	100%	Very effective	Yes	
		5	100%	Very effective	Yes	
		6	100%	Very effective	Yes	

			Intervention effect			
Author	Outcome measure	Participant	PND	Descriptor	Maintenance Achieved (Y/N)*	WoE D
Ennis and Losinski (2019)	Fraction probes	1	72.72%	Effective	Yes	2 Medium
		2	100%	Very effective	Yes	
		3	100%	Very effective	Yes	
		4	100%	Very effective	Yes	
		5	100%	Very effective	Yes	
		6	100%	Very effective	Yes	
		7	100%	Very effective	Yes	
		8	20%	Ineffective	No	

**Maintenance was considered to have been achieved if the follow-up scores did not fall below baseline.*

Table 12

Intervention outcome table for the case study design

		Outcomes		
Authors	Outcome measures	Pre-intervention	Post-intervention	WoE D
Ness and Middleton 2012	Participants maths grade score	C grade	B grade	1.1 Low

Conclusions and Recommendations

The studies in the review were deemed to be of medium (Ennis & Losinski, 2019; Ness & Middleton, 2012) or high (Cuenca-Carlino et al., 2016; Lucangeli et al., 2019; Wang et al., 2019) relevance to the review question. The findings from the majority of the studies provide promising evidence regarding the role that SRL strategy instruction can play in improving more literacy based aspects of maths performance (Lucangeli et al., 2019) alongside improving specific domains of maths performance namely, fractions (Ennis & Losinski, 2019; Wang et al., 2019) and equations (Cuenca-Carlino et al., 2016) in students who experience or are at risk of experiencing maths difficulties.

However, the quality of the evidence and conclusions that can be drawn are constrained by a number of factors. Therefore, the findings must be interpreted with caution.

In terms of methodological quality, this was not deemed as high in any of the studies, with the majority receiving a low rating (Appendix C). A methodological component effecting many of the studies were the measures used, which were often not obtained by using multiple methods.

The lack of follow-up data also impacted on the methodological quality in the context of the generalisability of the findings. The 2 SCEDs did provide follow-up data but this was limited as post-data was not collected over multiple intervals. This impacts on the studies' external validity and conclusions that can be made in terms of the effect of the intervention on sustained maths performance. More robust longitudinal data would provide EPs, in particular, with greater rationale for suggesting SRL strategy intervention for children with MDs. This therefore, should be collected in future studies exploring SRL strategy instruction as an intervention.

The studies provided sufficient participant characteristics to allow the reader to identify how applicable the finds were to the target group. However, none of the studies were conducted in the U.K. This impacts on the relevance of the findings to a U.K. population, which further undermines the external validity and conclusions that can be drawn particularly for EPs working in the U.K.

Another constraining factor within the review, was that the studies used different variations of SRL strategy intervention, which made it difficult for direct comparisons to be made. In addition, none of the studies explicitly explored the impact of the identifiable components of SRL strategy instruction on the outcomes achieved. This impacted on their methodological quality (WoE A) rating. Interestingly however, in the studies that explored receptivity of the intervention, through the use of qualitative questionnaire feedback, it appeared the evaluation component was highly valued. This was due to how the students and teachers felt it encouraged them to be more willing to self-appraise and problem solve (Cuenca-Carlino et al., 2016; Lucangeli et al., 2019).

Further research exploring identifiable components in more detail, both quantitatively and qualitatively, would be extremely relevant to educational providers and EPs alike. This is because, interventions with many components are time consuming to implement. Therefore, an understanding of what, if any, specific component is best received and most strongly related to improved performance, could result in the refinement of the instruction strategy which may improve feasibility in a school context.

Considering feasibility, it must be noted that the majority of the studies (Ennis & Losinski, 2019; Lucangeli et al., 2019; Wang et al., 2019) were implemented by a researcher or psychologists. Given the current financial constraints experienced by schools in the U.K, implementing interventions that involve outside partners are often not financially feasible. Thus, there is a need to further understand the effectiveness of SRL interventions delivered

specifically by teaching staff which may be considered to be more cost-effective.

In summary, a review of the studies did provide promising evidence that SRL strategy interventions are effective at improving performance in some areas of maths skills in students with or at risk of developing maths difficulties. However, given the methodological limitations, the findings from the studies are not conclusive and must be considered with caution.

In addition, there is a need for further research to be conducted within the U.K. context that include designs that allow for the collection of robust follow-up data. Such research is necessary before EPs, working within the U.K., can be confident in recommending SRL strategy development alone as an effective intervention for improving maths attainment amongst this target population.

References

Adams, C. M., Forsyth, P. B., Dollarhide, E., Miskell, R., & Ware, J. (2015). Self-Regulatory Climate: A Social Resource for Student Regulation and Achievement. *Teachers college record, 117*(2).

Barker, C., Pistrang, N., & Elliott, R. (2015). *Research methods in clinical psychology: An introduction for students and practitioners*. John Wiley & Sons.

Cargnelutti, E., Tomasetto, C., & Passolunghi, M. C. (2017). How is anxiety related to math performance in young students? A longitudinal study of Grade 2 to Grade 3 children. *Cognition and Emotion*, *31*(4), 755–764. https://doi.org/10.1080/02699931.2016.1147421

Cheng, M. (2017). 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). Retrieved from [online] Available at: https://www.ucl.ac.uk/educational-psychology/resources/CS1Cheng17.pdf [Accessed 14 Feb. 2020].

Clark, N. M., & Zimmerman, B. J. (1990). A social cognitive view of self-regulated learning about health. *Health Education Research*, *5*(3), 371-379.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Cuenca-Carlino, Y., Freeman-Green, S., Stephenson, G. W., & Hauth, C. (2016). Self-Regulated Strategy Development Instruction for Teaching Multi-Step Equations to Middle School Students Struggling in Math.

Journal of Special Education, *50*(2), 75–85. https://doi.org/10.1177/0022466915622021

Dahl, S., Aubrey, C., & Godfrey, R. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal*, *18*(1), 27–46.

Duncan, R. J., Schmitt, S. A., Burke, M., & McClelland, M. M. (2018). Combining a kindergarten readiness summer program with a self-regulation intervention improves school readiness. *Early Childhood research quarterly*, *42*, 291–300. https://doi.org/10.1016/j.ecresq.2017.10.012

Ennis, R. P., & Losinski, M. (2019). SRSD Fractions: Helping Students at Risk for Disabilities Add/Subtract Fractions With Unlike Denominators. *Journal of Learning Disabilities, 52*(5), 399–412. https://doi.org/10.1177/0022219419859509

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.

Graham, S., & Harris, K. R. (1996). Self-regulation and strategy instruction for students who find writing and learning challenging. *The science of writing: Theories, methods, individual differences, and applications*, 347-360.

Harden, A., & Gough, D. (2012). Quality and Relevance Appraisal. In D. Gough, S. Oliver, & J. Thomas (Eds.), An Introduction to Systematic Reviews (pp. 153–178).London: Sage

Harris, K. R., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. In *BJEP Monograph*

Series II, Number 6-Teaching and Learning Writing (Vol. 113, No. 135, pp. 113-135). British Psychological Society.

Hooper, S. R., Roberts, J., Sideris, J., Burchinal, M., & Zeisel, S. (2010). Longitudinal predictors of reading and math trajectories through middle school for African American versus caucasian students across two samples. *Developmental Psychology*, *46*(5), 1018–1029. https://doi.org/10.1037/a0018877

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

Lee, S., & McDonough, A. (2015). Role of self-talk in the classroom: investigating the relationship of eight-to-nine-year-olds' self-regulatory self-talk strategies with their classroom self-regulatory behaviour and mathematical achievement. *Early Child Development and Care*, *185*(2), 198–208. https://doi.org/10.1080/03004430.2014.915818

Leon, J., Nunez, J. L., & Liew, J. (2015). Self-determination and STEM education: Effects of autonomy, motivation, and self-regulated learning on high school math achievement. Learning and Individual Differences, *43*, 156–163. https://doi.org/10.1016/j.lindif.2015.08.017

Lucangeli, D., Fastame, M. C., Pedron, M., Porru, A., Duca, V., Hitchcott, P. K., & Penna, M. P. (2019). Metacognition and errors: the impact of self-regulatory trainings in children with specific learning disabilities. *ZDM-Mathematics Education, 51*(4, SI), 577–585. https://doi.org/10.1007/s11858-019-01044-w

Mevarech, Z., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning and meta-cognition. *Metacognition and learning*, *1*(1), 85-97.

Montague, M. (2008). Self-regulation strategies to improve mathematical problem solving for students with learning disabilities. *Learning Disability Quarterly*, *31*(1), 37–44. https://doi.org/10.2307/30035524

Musso, M. F., Boekaerts, M., Segers, M., & Cascallar, E. C. (2019). Individual differences in basic cognitive processes and self-regulated learning: Their interaction effects on math performance. *Learning and Individual Differences*, *71*, 58–70. https://doi.org/10.1016/j.lindif.2019.03.003

OFSTED. (2012). *Mathematics: made to measure*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/417446/Mathematics_made_to_measure.pdf [Accessed 11 Jan. 2020].

Nelson, G., & Powell, S. R. (2018). A Systematic Review of Longitudinal Studies of Mathematics Difficulty. *Journal of Learning Disabilities*, *51*(6), 523–539. https://doi.org/10.1177/0022219417714773

Ness, B. M., & Middleton, M. J. (2012). A Framework for Implementing Individualized Self-Regulated Learning Strategies in the Classroom. *Intervention in School and Clinic*, *47*(5), 267–275. https://doi.org/10.1177/1053451211430120

Parsons, S., Bynner, J., & Foudouli, V. (2005). Measuring basic skills for longitudinal study. *The Design and Development of Instruments for Use with Cohort Members in the Age*, *34*.

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing Single-Subject Research. Behavior Modification, 22(3), 221–242. https://doi.org/10.1177/01454455980223001

Wang, A. Y., Fuchs, L. S., Fuchs, D., Gilbert, J. K., Krowka, S., & Abramson, R. (2019). Embedding Self-Regulation Instruction Within Fractions Intervention for Third Graders With Mathematics Difficulties. Journal of Learning Disabilities, *52*(4), 337–348. https://doi.org/10.1177/0022219419851750

Appendix A: Articles excluded at full text screening

Articles Excluded	Criteria
Bishara, S. (2016). Self-regulated math instructions for pupils with learning disabilities. *Cogent Education*, *3*.	3. Not an intervention study
Buzza, D. C., & Dol, M. (2015). Goal Setting Support in Alternative Math Classes: Effects on Motivation and Engagement. *Exceptionality Education International*, *25*(1), 35–66.	7. Did not collect pre and post student maths' data
Collingwood, N., & Dewey, J. (2018). "Thinking Your Problems Away": Can maths interventions be developed to address both the academic and affective aspects of learning in primary aged children? *Educational and Child Psychology*, *35*(Spec Iss 2), 76–92.	5. Not all participants had maths difficulties
Cueli, M., Rodriguez, C., Areces, D., Garcia, T., & Gonzalez-Castro, P. (2017). Improvement of self-regulated learning in mathematics through a hypermedia application: Differences based on academic performance and previous knowledge. *The Spanish Journal of Psychology*, *20*.	5. Not all participants had maths difficulties

Articles Excluded	Criteria
Fung, J. J. Y., Yuen, M., & Yuen, A. H. K. (2014). Self-Regulation in Learning Mathematics Online: Implications for Supporting Mathematically Gifted Students with or without Learning Difficulties. *Gifted and Talented International*, *29*(1–2), 113–123.	3. Not an intervention study
Hacker, D. J., Kiuhara, S. A., & Levin, J. R. (2019). A metacognitive intervention for teaching fractions to students with or at-risk for learning disabilities in mathematics. *ZDM-MATHEMATICS EDUCATION*, *51*(4, SI), 601–612.	3. A review study, not an intervention study
Lazakidou, G., & Retalis, S. (2010). Using computer supported collaborative learning strategies for helping students acquire self-regulated problem-solving skills in mathematics. Computers & Education, *54*(1), 3–13.	5. Not all participants had maths difficulties
Lee, S., & McDonough, A. (2015). Role of self-talk in the classroom: investigating the relationship of eight-to-nine-year-olds' self-regulatory self-talk strategies with their classroom self-regulatory behaviour and mathematical achievement.*Early Child Development and Care*, *185*(2), 198–208.	3. Not an intervention study
Leon, J., Nunez, J. L., & Liew, J. (2015). Self-determination and STEM education: Effects of autonomy, motivation, and self-regulated learning on high school math achievement. *Learning and indiviudal differences*, *43*, 156–163.	5. Not all participants had maths difficulties

Articles Excluded	Criteria
Maffei, L. (2012). Self-regulated learning enhancement through self assessment activities. In Chova, LG nad Martinez AL and Torres, IC (Ed.), INTED2012: International technology, education and development conference (pp. 6560-6569).	1. A conference paper, not published in a peer reviewed journal
Martin, C. S., Polly, D., & Kissel, B. (2017). Exploring the impact of written reflections on learning in the elementary mathematics classroom. Journal of Educational Research, *110*(5), 538–553.	3. Not an intervention study
Paraskevi, K., Irini, D., Sotiria, T., & Charoula, S. (2017). Teaching self-regulation strategies with SOLVE IT to two students with learning disabilities: Effects on mathematical problem-solving performance. In Dooley, T and Gueudet, G (Ed.), Proceedings of The Tenth Congress of The European Society for Research in Mathematics Education (Cerme10)(pp. 1114–1121).	1. A proceeding paper, not a published in a peer reviewed journal

Articles Excluded	**Criteria**
Sanabria Rodriguez, L. B., Valencia Vallejo, N. G., & Ibanez Ibanez, J. (2017). *Effect of Training in Self-Regulation for Learning Mathematics. Praxis & Saber*, *8*(16), 35–56.	5. Not school aged participants
Santos, L., & Pinto, J. (2009). Self-regulating assessment in Mathematics: Saying befo*re doing. Bolema Mathematics Education Bulletin Boletim De Educaacao Matematica*, *22*(33), 51–68.	9. Study not originally written in English
Rutherford, T. (2017). The measurement of calibration in real contexts. Learning and Instruction, *47*, 33–42.	5. Not all participants had maths difficulties
Van Boxtel, J. M. (2016). REASON: A Self-Instruction Strategy for Twice-Exceptional Learners Struggling With Common Core Mathematics. *Teaching Exceptional Children*, *49*(1), 66–73.	7. Did not collect pre and post student maths' data

Appendix B: Mapping the field

Author	Aim of the study	Setting	Research Design	Participants	Intervention Details	Measures	Key Findings
Cuenca-Carlino, Freeman-Green, Stepehnson & Hauth (2015)	To examine the effectiveness of Self-Regulated Strategy Development (SRSD) on eighth-grade students learning to solve multi-step equations.	Public middle school in North America.	Single case, multiple baseline across participants design	• 6 students; 4 females, 1 male. • Mean age: 13 • All previously identified as needing specific maths intervention due to their low scores on a maths benchmark assessment and student standardised test results.	The model of Self-Regulated Strategy Development (SRSD) instruction was implemented to teach students how to solve and check multi-step equations. Delivered by a highly qualified maths teacher who was trained on SRSD learning strategies by a research. Duration 4 sessions a week for 45 minutes over 12 weeks.	Equation probes	**Equation Probes** Substantial improvements were observed in the students' abilities to strategize their work and solve equations correctly. 100% PND calculated. Weighted average Tau-U of .98.
Ennis and Losinski (2019)	To extend the work of Losinski et al. (2019) by investigating the FILMS, CUT, and EDIT SRSD Fractions intervention as a Tier 2 intervention with fifth-grade students at risk for mathematics disabilities with and without behavioral challenges	Suburban elementary school in South Eastern United States.	Single case, multiple baselines across participants design	• 8 5th Grade students; 5 males, 3 females. • Met criteria for mathematics and fractions difficulty risk	Fractions package based on the SRSD framework outlined by Harris & Graham (1996). Lessons were divided into teaching specific skill areas based on the mnemonics FILMS, CUT & EDIT. Duration- Each day during rotations time. Engaged in approx. 9 lessons using the intervention approach. Delivery Delivered by the researcher, with extensive	Fraction probes	**Fraction Probes** PND for 7/8 participants suggest a large effect. .

					experience in implementing SRSD reading and writing interventions.		
Author	**Aim of the study**	**Setting**	**Research design**	**Participants**	**Intervention details**	**Measures**	**Key Findings**
Lucangeli, Fatame, Pedron, Porru, Duca, Hitchcot &Penna (2019)	To investigate the impact of a metacognitive and cognitive training programme developed to enhance various arithmetic skills, self regulatory and control functions in primary and secondary school students exhibiting atypical mathematical development.	Primary or secondary school in Italy.	Group based, Quasi-experimental, pre-test, post-test design.	• 69 Italian secondary and primary school children. • 36 Males; 32 Females. • Mean age 9.3 years. • Meeting criteria for the diagnosis of dyscalculia or specific difficulties in mathematics.	Individualised cognitive and self-regulatory training programme. Delivered by a team of psychologists specialising in the treatment of specific learning disabilities. Duration 60 min sessions Weekly for 16 weeks.	Mathematical tests included in the AC-MT 6-1- and AC-MT 11-14 batteries; developed for the assessment of calculation and problem solving skills of Italian students.	**AC-MT measure** • Experimental group outperformed control group on some (written calculation and digit transcription) but not all measures.
Ness and Middleton (2012)	To illustrate the effects of strategy instruction on academic behaviours consistent with the Self Regulated Learning cycle, using a case example.	Sixth grade classroom	Case Study Design	• Sixth grade boy with learning difficulties, specifically with maths and literacy skills well below expected achievement.	An acronym was used to define 4 self-regulatory concepts (MARS). An external visual aid was developed to prompt the child to remember the 4 concepts and the child was taught individually by the special educational needs teacher to apply this strategy in the classroom. A check in check out paradigm was used to deliver the intervention.	Class grade	**Class grade** • Increased from a C to B

Author	Aim of the study	Setting	Research Design	Participants	Intervention Details	Measures	Key Findings
Wang, Fuchs, Fuchs, Gilber, Krowka & Abramson (2019)	To explore the efficacy of fractions interventions with and without an embedded self-regulation (SR) component for third-grade students at risk for mathematics disabilities	Primary Schools in a large metropolitan school district	Group based, experimental, pre-test, post-test design	• 69 participants • 3rd Grade from 19 classrooms across 6 schools • Met low level maths skills criteria	Base intervention consisted of the third grade super solvers programme. SR condition consisted of the third grade super solvers programme plus embedded SR component. Sessions in both conditions involved 3, 35 minute sessions each week for 13 weeks. Delivery Delivered by tutors who were research grant employees.	Six fraction outcomes	**<u>Fraction outcomes</u>** SR condition produced significantly higher scores V's control on NAEP, Ordering of fractions ,Fraction word problems and Single digit multiplication No significant improvements on the number line task Results indicate an advantage for the SR condition over the base condition.

Appendix C: Weight of Evidence

1. Weight of Evidence A

WoE A for the group based studies was evaluated using the Kratochwill (2003) coding protocol for evaluating group based designs. WoE A for single case study designs and the case study were evaluated using the Kratochwill (2003) protocol for evaluating single case study designs. The rationale for selecting the Kratochwill protocol's over others was that it is a protocol that was developed specifically to support professionals in the evaluation of intervention based research. In addition, using a protocol designed by the same researcher for both group based and single case designs made for a more direct comparison between the studies for the purpose of the review.

The original protocol includes three key areas by which the research is evaluated

I. General Characteristics

II. Follow up assessment

III. Other descriptive or supplemental criteria to consider

Within these areas there are further evaluative subheadings. As this review did not require the use of all subheadings, amendments were made to the protocol. All amendments made accompanied by supporting rationale, can be found in Table 1. These amendments were based on amendments made by Cheng (2017) as part of a systematic literature review exploring the effectiveness of self-

regulated strategy development as an approach to improve reading comprehension in school aged pupils with Special Educational Needs.

At the end of each protocol document, five areas were considered from both protocols is order to determine the WoE A rating. The Kratochwill coding manual provided guidance on how to code each area which resulted in a weighting score ranging from 0-3. Each of these scores across the 5 areas were averaged to provide the study's overall WoE A rating.

Table 1

Modified sections to Kratochwill (2003) coding protocol and rationale

Modified Section	Rationale
Removed from coding protocol	
I. B8 & B9 in the single case design protocol B7 & B8 in the Group-based design protocol	These sections are used for coding qualitative data which was not relevant to the quantitative studies used in this review
II. C Measures support primary and secondary outcomes	This was not included as the outcomes are considered separately in this review
II. D Educational and clinical significance	The Educational significance is considered separately as part of this review
II. G Replication	Replication was not seen as a necessity in order to be included in this review
II. H Site of Implication	Site of implication was considered in the inclusion/exclusion criteria. All sites of implication were in a school setting
III. D Dosage	The use of dosage was not relevant in this review as the review was not looking at dosage treatment effects
III. F Characteristics of the Intervener	No individual characteristics of the intervener were reported in any of the studies.

III. H Cost analysis	Cost analysis was outside the scope of this review. In addition, none of the studies did not provide cost analysis details

Modifications to existing sections	**Rationale**
III. F- Fidelity of intervention (single-case design)	As the fidelity of the intervention was reported on a study level not an individual level
III. A2. 3 Participant characteristics Specified for treatment and control group	Section has been modified into a checklist to simplify the coding process for the coder

Table 2

Overall WoE A scores for included studies

Study	A Measurement	B Baseline/comparison	E Identifiable component	F Implementation Fidelity	I Follow up assessment	Overall WoE A*
Cuenca-Carlio et al. (2015)	1	3	0	3	2	1.8
Ennis and Losinski (2019)	1	3	0	2	2	1.6
Lucangeli et al. (2019)	0	1	0	2	0	0.6
Ness and Middleton (2012)	0	0	0	0	0	0
Wang et al. (2019)	2	2	0	3	0	1.4

* Overall WoE A was calculated by summing each individual rating and then dividing the total by 5.

2. Weight of Evidence B

The relevance of the methodology used in each of the studies to answer the current review question was evaluated using Weight of Evidence B (WoE B) criteria outlined in Table 3. The rationale for the criteria can be found in Table 4. The WoE B rating assigned to each study is outlined in Table 5.

Table 3

WoE B Criteria

WoE B Rating	Criteria
3 **High**	• **Design** Randomised control group whereby the control group is receiving an alternative maths intervention • **Measures** Primary measures used to test effectiveness are described giving both reliability and validity data • **Data reporting** Use of both pre and post data
2 **Medium**	• **Design** Quasi-experimental studies or single case multiple baseline design studies • **Measures** Primary measures used to test effectiveness are described giving either reliability *or* validity information (for at least one of these primary measures) • **Data reporting** Use of both pre and post data
1 **Low**	• **Design** Case studies • **Measures** Primary measures used to test effectiveness are described • **Data reporting** Use of pre and post data
0 **No weighting**	• None of the above criteria are met

Table 4

WoE B Rationale for the criteria used

Criteria	Rationale
Design	Randomised control trials are considered to be a high quality design when evaluating intervention effectiveness (Petticrew & Roberts, 2003)
Measures	The use of measures that are reliable and valid increase the chance that the research is actually measuring what it is stating to do so.
Data reporting	Pre and post data allows for an exploration as to whether there were any changes in scores following an intervention

Table 5

WoE B scores and descriptors for included studies

Study	WoE B	Descriptor
Cuenca-Carlio et al. (2015)	2	Medium
Ennis and Losinski	2	Medium
Lucangeli et al. (2019)	2	Medium
Ness and Middleton	1	Low
Wang et al. (2019)	2	Medium

3. Weight of Evidence C: Topic Relevance

The topic relevance for each study in relation to the review question was reviewed using Weight of Evidence C (WoE C) criteria outlined in Table 6. The rationale for the criteria can also be found in Table 8. The WoE C rating assigned to each study is outlined in Table 8.

Table 6

WoE C criteria

Criteria	Weighting & Descriptor	Rationale
1. Sample	3: Sample specified as having or at risk of having mathematical difficulties with supporting assessment data 2: Sample specified as having or at risk of having mathematical difficulties but without supporting assessment data 1: Sample not specified as having or at risk of having mathematical difficulties	The review is investigating the effectiveness of teaching self-regulation strategies to students with or at risk of developing mathematical difficulties
2. Setting generalisability	3: Participants sampled from more than one school 2: Participants sampled from one school 1: Participants not sampled from a school setting	Replicated studies across multiple settings, increases the external validity

Criteria	Weighting & Descriptor	Rationale
3. Intervention component	3: Teaching students self-regulation strategies is described as the main component of the intervention and includes the 3 key components (Planning, Performance, Evaluation) 2: Teaching students self-regulation strategies is part of the intervention but is not described as the main component 1: Teaching students self-regulation strategies is not a component of the intervention	The review is investigating the effectiveness of teaching self-regulation strategies on mathematical performance
4. Fidelity	3: Strong evidence of fidelity to the intervention procedure 2: Promising evidence of fidelity to the intervention procedure 1: No/limited or weak evidence of fidelity to the intervention procedure	To ensure the consistency of the intervention implementation across the sample
5. Outcomes	3: More than one maths outcome is measured 2: One maths outcome is measured 1: No maths outcome is measured	The review is investigating the effectiveness of teaching students self-regulation strategies on mathematical performance

Table 7

Qualitative descriptors of WoE C ratings

Overall Quality	Average Score
High	≥ 2.5
Medium	1.5-2.4
Low	≤ 1.4

Table 8

WoE C summary for included studies

Study	A Sample	B Setting generalisability	C Intervention component	D Fidelity	E Outcomes	Overall WoE C
Cuenca-Carlio et al. (2015)	3	2	3	3	2	2.6 High
Ennis and Losinski (2019)	3	2	3	2	2	2.4 Medium
Lucangeli et al. (2019)	3	2	3	1	2	2.2 Medium
Ness and Middleton (2012)	3	2	3	1	2	2.2 Medium
Wang et al. (2019)	3	3	3	3	3	3 High

Appendix D: Coding Protocol for each of the studies

Coding Protocol for Group-Based 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: 18/01/2020

Full study reference in APA format

Lucangeli, D., Fastame, M. C., Pedron, M., Porru, A., Duca, V., Hitchcott, P. K., & Penna, M. P. (2019). Metacognition and errors: the impact of self-regulatory trainings in children with specific learning disabilities. *ZDM*, 1-9.

Study ID Number (Unique Identifier): 1

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)

A3.5☐Very high (explicitly stated)

A3.6☐N/A

A3.7☐Unknown/unable to code

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

B1. Appropriate unit of analysis☒ yes ☐no

B2. Familywise error rate controlled ☐yes ☒no ☐N/A

B3. Sufficiently large N ☒yes ☐no

Statistical Test: ANOVA level: 0.05 ES: Large N required: 26 per group (for 2 groups)

B4. Total size of sample (start of the study): 68 (34 per group)

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* (answer A1 through A4)

*** Note only measurements analysing mathematical performance (primary outcome) were reviewed in this protocol.**

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 ☐

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)

A4.1☒Yes

A4.2☐No

A4.3☐Unknown/unable to code

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

☒0 ☐1 ☐2 ☐3

As reliably measures, although norm referenced, do not provide reliability co-efficient details

B. Comparison Group

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

B1.1☒Typical contact

B1.2☐Typical contact (other) specify:

B1.3☐Attention placebo

B1.4☐Intervention elements placebo

B1.5☐Alternative intervention

B1.6☐PharmacotherapyB1.1

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 Only rating as 1 as confidence in rating the judgement was low

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: Not specified

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 No evidence of using a familywise/experiementalwise error rate control.

E2. Design allows for analysis of identifiable components (select one) ☐yes ☒no

E3. Total number of components: Delivered as a whole package via a computer-assisted task

E4. Number of components linked to primary outcomes: N/A

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 phase

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:

☐Number of participants included in the follow up assessment:

☐Consistency of assessment method used:

Rating for Follow Up Assessment (select 0, 1, 2, or 3): ☐3 ☐2 ☐1 ☒0 No follow up assessment.

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

A1. Sampling procedures described in detail ☒yes ☐no

A1.1Inclusion/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 maths achievement

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes ☐yes ☒no Specify:

A3.2 Have relevance to inclusion criteria☒ yes ☐no Specify: Participants met the criteria for the diagnosis of dyscalculia or presented with transitory difficulties in mathematics.

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

Participants from treatment group	Results (what person reported to have gained from participation in the programme)	General rating
☒Child ☐Parent ☐Teacher ☐Other	Post-training questionnaire outlined that children in the experimental group were more actively involved in the use of SRL strategies and were also more active in evaluating their work and problem solving.	☒Participants reported benefitted overall from the intervention (This was inferred from the questionnaire analysis) ☐Participants reported did not benefit overall from the intervention ☐Participants did not report receptivity/acceptance

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: 16 weeks

B2.2 months: ___

B2.3 years: ___

B2.4 other: ___

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

C1. Unknown/insufficient information provided ☐

C2. Information provided (if information is provided, ☒ specify both of the following:)

C2.1 length of intervention session 60 minutes

C2.2 frequency of intervention session Once a week

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: Psychologists

E9.☐Unknown/insufficient information provided

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) Metacognitive

G7.☐Unknown/insufficient information provided

Summary of Evidence for Group-Based Design Studies

Indicator	Overall Evidence NNR= No numerical rating OR 1-3	Description of Evidence Strong Promising Weak No/limited evidence Or Descriptive rating
Key areas of judgement for weight of evidence A		
Measurement	**0**	**Weak**
Comparison Group	**1**	**Weak**
Identifiable Component	**0**	**Limited**
Implementation Fidelity	**2**	**Promising**
Follow up assessment conducted	**0**	**No evidence**

Average Quality of Evidence across the Included Judgement Areas
Σ of X = 0 + 1 + 0 + 2 + 0 = 0.6 **———— ——————** **N 5** X= Individual quality of evidence for each judgement area N= Number of judgement areas **Overall Rating for Weight of Evidence A: 0.6**

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: 18/01/2020

Full study reference in APA format

Cuenca-Carlino, Y., Freeman-Green, S., Stephenson, G. W., & Hauth, C. (2016). Self-regulated strategy development instruction for teaching multi-step equations to middle school students struggling in math. *The Journal of Special Education*, *50*(2), 75-85.

Study ID Number (Unique Identifier): 4

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 singleparticipant 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): 6

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.)

*** Note only measurements analysing mathematical performance (primary outcome) were reviewed in this protocol.**

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

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)

A4.1☒Yes

A4.2☐No

A4.3☐Unknown/unable to code

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

☐0 ☒1 ☐2 ☐3

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

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

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

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

Participant 5 (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

Participant 6 (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

Average Quality of Baseline Rating Across Participants:

Σ of X = 18/6 = 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: 6

E4. Number of components linked to primary outcomes: Data analysis did not allow for direct links to be made to different components

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: 4 weeks

☒Number of participants included in the follow up assessment: 6

☒Consistency of assessment method used: Similar equation probes used in both baseline and post instruction phase.

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

III. Other Descriptive or Supplemental Criteria to Consider

B. External Validity Indicators

A1. Sampling procedures described in detail ☒ yes ☐no

A1.1Inclusion/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 maths achievement

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

Participants from treatment group	Results (what person reported to have gained from participation in the programme)	General rating
☐Child ☐Parent ☐Teacher ☐Other	Overall students viewed the intervention as positive. The most valuable component reported was evaluation.	☒Participants reported benefitted overall from the intervention

		☐Participants reported did not benefit overall from the intervention ☐Participants did not report receptivity/acceptance

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: ___

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

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) Metacognitive

G7.☐Unknown/insufficient information provided

Summary of Evidence for Single-Participant Design Studies

Indicator	**Overall Evidence** **NNR= No numerical rating** **OR** **1-3**	**Description of Evidence** **Strong** **Promising** **Weak** **No/limited evidence** **Or** **Descriptive rating**
Key areas of judgement for weight of evidence A		
Measurement	**1**	**'Promising'**
Baseline	**3**	**'Strong'**
Identifiable Component	**0**	**'Limited evidence'**
Implementation Fidelity	**3**	**'Strong'**
Follow up assessment conducted	**2**	**'Promissing'**

Average Quality of Evidence across the Included Judgement Areas
Σ of X = 1 + 3 + 0 + 3 + 2 = 1.8 **_____ _____** **N 5** X= Individual quality of evidence for each judgement area N= Number of judgement areas **Overall Rating for Weight of Evidence A: 1.8**