Case Study 1: An Evidence-Based Practice Review Report

How effective is Precision Teaching in teaching numeracy skills to school-aged pupils with maths difficulties?

Summary

This literature review aims to evaluate the effectiveness of Precision Teaching (PT) interventions in teaching numeracy skills to school-aged pupils with maths difficulties. PT is an intervention used one-to-one between pupils and teaching staff to encourage pupil progress in basic academic skills. Rather than being a prescribed teaching technique, it is a tool for measuring pupil progress which uses probes and standard celeration charts (SCC) to closely monitor and modify teaching to promote positive learning outcomes for pupils. Following a comprehensive literature search, five relevant studies were identified which met the review's inclusion and exclusion criteria. These five studies were critically reviewed using the Weight of Evidence Framework (Harden & Gough, 2012), and where possible, effect sizes were calculated to help conclude the effectiveness of the intervention. All five studies found PT to improve the numeracy skills of pupils with maths difficulties, with effect sizes indicating a medium to large effect. However, the methodological quality of these studies was poor which has a number of implications when interpreting the findings. The methodological relevance of the studies and their focus was also varied, and thus suggestions for future research have been made which would address the current research limitations.

Introduction

What is Precision Teaching?

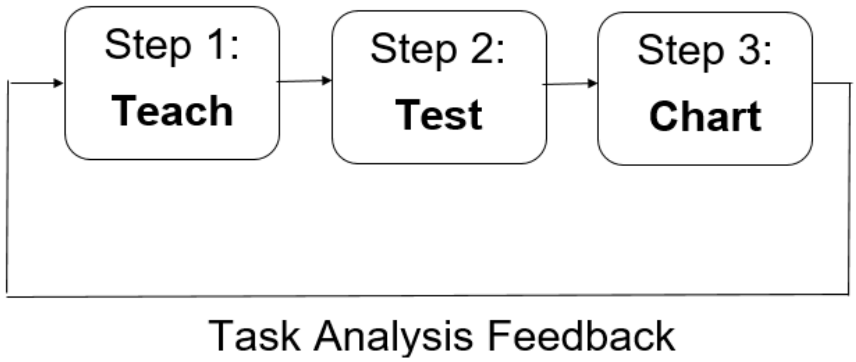
PT is a targeted wave three intervention, usually delivered for approximately ten minutes daily by a teaching assistant (TA) one-to-one with a pupil. Despite its name, PT is best understood as a measurement tool of learning (West, Young & Spooner, 1990), rather than a specific teaching technique. It offers a structured approach to planning, evaluating and modifying teaching strategies to support pupils who may either be struggling with basic academic skills, or who may lack the confidence or motivation to complete such tasks (Binder & Watkins, 2013). The goal of PT is to support pupils in developing accuracy and fluency in specific targeted skills. To achieve this, Solity and Bull (1987) recommended that PT interventions should include five key components (see Table 1).

Table 1: *Solity and Bull (1987) five key components of PT*

1	The desired pupil performance should be precisely defined in observable and measurable terms.
2	Record the pupil's performance on a daily basis using a probe for one minute.
3	Chart the pupil's performance of the probe on a daily basis using a SCC.
4	Record the teaching approach and behaviour in relation to the pupil's performance to enable task analysis.
5	Analyse the data to determine whether progress is satisfactory or whether adaptations are needed to maintain or accelerate learning.

At the beginning of a programme, clear targets are set which specify how many correct and incorrect answers would indicate that the pupil has learnt the taught skills. These targets can either be based upon age expectations or individualised to the pupil, and should be ambitious to help a student achieve their potential (Gallagher, 2006). The intervention is then delivered daily and includes three key stages; teach, probe[1] and chart using a SCC[2] (see Figure 1). These charts provide a 'learning picture' of the pupil and enable task analysis by providing instant feedback on the pupil's progress and evaluation of the teaching methods used (Lindsley, 1972). The intervention is based on the assumption that the learner always knows best (Lindsley, 1972), and therefore if teaching methods are effective they should continue to be used. However, if ineffective, it is deemed the responsibility of the teacher to alter their teaching methods to suit the learner to ensure the learner experiences success (Lindsley,1990).

Figure 1: *Outline of a daily PT intervention*



[1] A probe is an activity which assesses how many correct and incorrect answers a pupil achieves of a specific set of targeted skills. This test phase usually lasts one minute

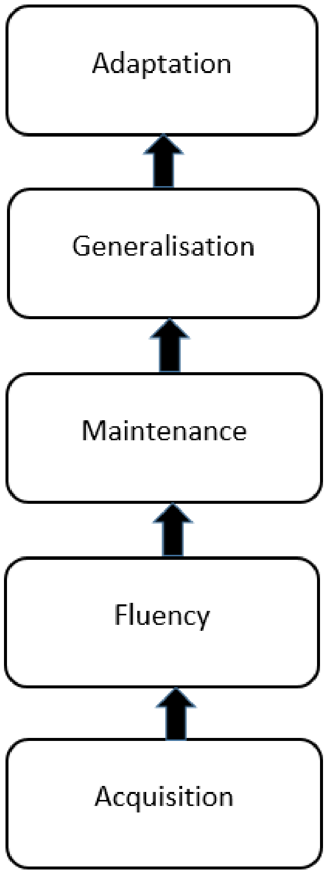
[2] A standard celeration chart (SCC) is a chart used to record the frequency of a pupil's behaviour, which in the case of PT, is the number of correct and incorrect responses a pupil displays in daily PT sessions.

What are the psychological theories underpinning Precision Teaching?

PT draws upon the principles of behavioural analysis (Potts, Eshleman & Cooper,1993) and key ideas and techniques developed by Skinner and the Behaviourist approach to learning. This approach is underpinned by the notion that behaviours can be learned when taught (Skinner, 1985), and techniques such as conditioning and reinforcement encourage the development of new behaviours (Lindsley, 1992). Skinner's pupil, Lindsley, applied these principles to education and working with pupils with special educational needs (SEN), leading to the development of PT (Gallagher, 2006; Lindsley, 1992). Lindsley (1992) emphasised the optimistic approach of the intervention and encouraged the use of positive reinforcement through both verbal praise and visual feedback from the SCC. This positive reinforcement motivates pupils and has been found to increase pupils' self-confidence (Binder & Watkins, 2013).

The Instructional Hierarchy of Learning (Haring, Lovitt, Eaton & Hansen, 1978) which describes how we acquire all skills, is very helpful in understanding how PT works. This model presents five chronological stages (see Figure 2), starting with the 'acquisition', or accuracy, stage, in which the targeted skill needs to be modelled to the pupil. This draws upon the ideas of Vygotsky's Zone of Proximal Development (1987) which emphasises the need for a more able other to support a child in developing their learning using scaffolding techniques such as modelling. Once a pupil knows what the skill is that they are trying to learn, they move to the 'fluency' stage in which they are expected to perform the skills quickly and repeatedly until it is readily accessible and can be completed automatically (Weisenburgh-Snyder, Malmquist, Robbins & Lipshin, 2015). The goal of PT is for pupils to become both accurate and fluent at taught skills, and once they can do both of these simultaneously, they have achieved a level of ‘maintenance’ or ‘mastery’ (Haring et al., 1978).

Figure 2: *The Instructional Hierarchy of Learning (Haring et al., 1978)*



Rationale and relevance to Educational Psychology practice

Mathematics is a core component of the National Curriculum in England, and since 2013 nearly all school pupils must achieve a minimum of level 4 (previously a C grade) at GCSE (Education and Skills Funding Agency, 2014). This initiative was introduced to address the high numbers of pupils leaving school lacking essential numeracy skills for adult life and employment. However, many pupils are still struggling to meet these ambitious standards and therefore there is a role for Educational Psychologists (EPs) to support pupils in making progress in this subject.

One theory for why pupils are leaving school with insufficient numeracy skills is because they are failing to develop their basic maths skills fluently (Binder, 1988). Without fluency, pupils are unable to access mathematical information quickly and easily, and they cannot progress on to more advanced aspects of the curriculum which utilise these skills (Haughton, 1972; McDowell & Keenan, 2001). This leads to a growing gap between those pupils struggling with maths and their peers who are not, and has long-term negative consequences on their life outcomes (DfE, 2014). It is therefore of great importance that this difficulty is addressed. This theory of under-attainment in maths would suggest that an intervention like PT, which aims to develop accurate and fluent academic skills, may be effective in supporting pupils with maths difficulties by enabling them to reach a good level of fluency which in turn will allow them to progress in their learning.

To use PT, school staff should receive training from an EP which is usually completed within one training session. In the session, staff are taught the psychological theories (e.g. Haring et al., 1978) and practical skills (e.g. use of SCCs) necessary to deliver the intervention effectively. The intervention is popular with schools as it is cost-effective and addresses current issues surrounding the ineffective use of TAs in schools (Blatchford, Russell & Webster, 2012). EPs also favour the intervention as it upskills staff in psychological interventions, supports pupils with SEN who are struggling with their learning, and prevents the escalation of academic difficulties. By exploring whether PT is effective for teaching numeracy skills, this review aims to extend the literature on this intervention, which to date, is primarily focused on its use for developing literacy skills.

Review Question

How effective is Precision Teaching in teaching numeracy skills to school-aged pupils with maths difficulties?

Section 3: Critical Review of the Evidence Base

Literature Search

A comprehensive literature search was undertaken in December 2017 using four databases which provided access to articles published in peer-reviewed journals relating to psychology and education. Table 2 shows the search terms used.

Table 2: *Database search terms*

Database	Search Terms
PsycInfo	"precision teach*" OR "precision monitor*" + "math*" OR "calcul*" OR "number*" OR "arithmetic"
Education Resources Information Centre (ERIC)	"precision teach*" OR "precision monitor*" + "math*" OR "calcul*" OR "number*" OR "arithmetic"
British Education Index (EBSCO)	"precision teach*" OR "precision monitor*" + "math*" OR "calcul*" OR "number*" OR "arithmetic"
Web of Science	“precision teaching”

This initial search produced 345 results. An additional search of the Journal of Precision Teaching and Celeration produced 21 results and an ancestral search identified two further studies. Therefore, the searches produced a total of 368 studies, although 7 of these were duplications. Figure 3 shows at which stages studies were excluded following the review of their titles, abstracts and full text. Appendix A lists the excluded studies following full text screening and Table 4 shows the five studies which met all eight inclusion/exclusion criteria and thus have been include in the current literature review. For a summary of each paper, see Appendix B ‘Mapping the Field’.

Figure 3: *Flow diagram of literature search*

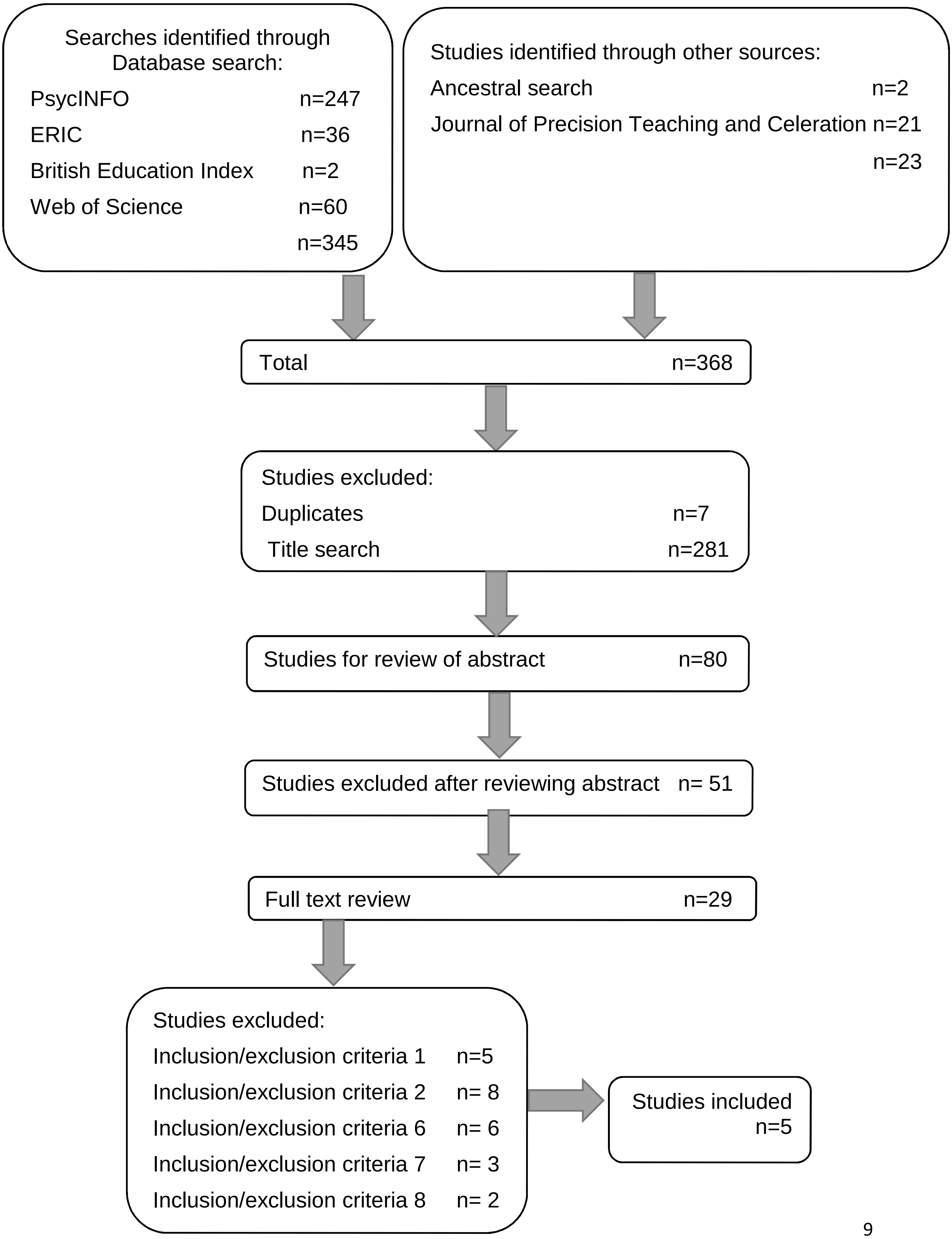


Table 3: *Inclusion and exclusion criteria*

		Inclusion Criteria	Exclusion Criteria	Rationale
1	Type of publication	Articles from peer reviewed journals	Articles published in non-peer reviewed journals	Peer review studies have undergone a higher level of scrutiny, ensuring a minimum standard of quality and methodological rigour has been met.
2	Date of publication	Studies published since 1988	Studies published before 1988	Only studies published after the introduction of the UK National Curriculum in 1998 will be relevant to the current education context.
3	Geographical context and language	Studies can be from any geographical context, but studies must be written in English	Studies not written in English	For convenience as translation services are not available.
4	Setting	Intervention delivered in an education setting.	Study not based in an education setting e.g. at home.	To consider the implication of Precision Teaching in educational settings.

		Inclusion Criteria	Exclusion Criteria	Rationale
5a)	Participants	School-aged pupils	Not school-aged pupils e.g. University students	To consider the implications of Precision Teaching on statutory school-aged pupils.
5b)	Participants	Pupils with maths difficulties	No pupils with maths difficulties	To consider the use of Precision Teaching in supporting pupils with maths difficulties.
6	Intervention	Precision Teaching	Not Precision Teaching	To investigate whether pupil outcomes are due to a programme of Precision Teaching.
7	Type of design	Group-based or case study design	Not group-based or case study design	To compare the effectiveness of the intervention either between or within participants.
8	Measures	The study must have pre and post intervention data for pupils' attainment in a maths skill.	The study does not have pre and post intervention data for pupils' attainment in a maths skill.	Pre and post data will help to measure the effectiveness of Precision Teaching.

Table 4: *Studies included in the review*

Included studies
1. Gallagher, E. (2006). Improving a mathematical key skill using precision teaching. *Irish Educational Studies, 25*(3), 303-319.
2. Chiesa, M. & Robertson, A. (2000). Precision teaching and fluency training: Making maths easier for pupils and teachers. *Educational Psychology in Practice, 16*(3), 297-310.
3. Strømgren, B., Berg-Mortensen, C., & Tangen, L. (2014). The use of precision teaching to teach basic math facts. *European Journal of Behavior Analysis, 15*(2), 225-240.
4. Weisenburgh-Snyder, A., Malmquist, S., Robbins, J., & Lipshin, A. (2015). A model of MTSS: Integrating precision teaching of mathematics and a multi-level assessment system in a generative classroom. *Learning Disabilities: A Contemporary Journal, 13*(1), 21-41.
5. Bullara, D., Kimball, J., & Cooper, J. (1993). An assessment of beginning addition skills following three months without instruction or practice. *Journal of Precision Teaching, 6*(1), 11-16.

Weight of Evidence

The five studies selected for review were evaluated using Harden and Gough's (2012) Weight of Evidence (WoE) Framework. Each study was judged on methodological quality (WoE A), relevance of methodology (WoE B) and relevance of the focus of the study to the review question (WoE C). These three weightings were summed and divided by three to produce an overall WoE for each study (WoE D). The highest possible rating for each WoE was three. Table 5 shows each WoE for each study, and Appendix C gives further details of the criteria and rationale used to reach these final scores.

Table 5: *WoE A, B, C and D for all five studies*

	WoE A	WoE B	WoE C	WoE D
Gallagher (2006)	1 (Low)	2 (Medium)	2.3 (Medium)	1.8 (Medium)
Chiesa & Robertson (2000)	1 (Low)	2 (Medium)	2.7 (High)	1.9 (Medium)
Strømgren, Berg-Mortensen & Tangen (2014)	1 (Low)	3 (High)	2 (Medium)	2 (Medium)
Weisenburgh-Snyder, Malmquist, Robbins & Lipshin (2015)	1.4 (Low)	1 (Low)	2.3 (Medium)	1.6 (Low)
Bullara, Kimball & Cooper (1993)	0.9 (Low)	1 (Low)	1.3 (Low)	1.1 (Low)

Ratings between 0.9 -1.6 are “low”, 1.7- 2.3 are “medium” and 2.4 - 3 are “high”

Participants

All five of the studies had relatively small sample sizes, with a total of 107 participants included in the review. In total, across all the studies, 48 pupils received a PT intervention, with the remaining 59 participants acting as controls. Of those receiving PT, 31 were male and 17 were female, which may limit the generalisability of these study findings to female pupils.

All participants who received the PT intervention were identified as having some level of maths difficulty. Gallagher (2006), Chiesa and Robertson (2000), and Strømgren, Berg-Mortensen and Tangen (2014) relied on teachers to identify pupils in their class struggling with maths to be participants. This method of sampling was subjective, lacked reliability and was not supported with relevant assessment data. Therefore, these studies scored a lower rating on both WoE A and WoE C. Bullara et al.'s (1993) case study similarly lacked sufficient detail about their pupil, with no assessment data provided, meaning it also received a lower WoE A and WoE C rating. Weisenburgh-Snyder et al. (2015) was the only study which provided sufficient detail of the participants, including age, gender, ethnicity, disability diagnosis and assessment data, and therefore received the highest WoE A rating out of all the studies.

Location

Four of the studies explored the use of PT in primary school settings, whereas Bullara et al. (1993) examined the use of PT in a Psychoeducational Clinic, for a pupil who usually attended an urban city school. This study therefore scored a lower rating for WoE C as its findings are less generalisable to a classroom setting in which PT is most commonly delivered.

As well as both focusing on PT in school settings, Chiesa and Robertson (2000) and Gallagher (2006) studies were both conducted in the UK. This meant they received a higher WoE C rating due to having higher ecological validity and relevance to the UK education system. The two small-N case studies by Weisenburgh-Snyder et al.

(2015) and Bullara et al. (1993) were both conducted in the USA, and the study by Strømgren et al. (2014) was in Norway, leading them to receive a lower score for the WoE C criteria.

Design

The five studies adopted a number of different designs to explore the use of PT in teaching numeracy skills. Strømgren et al. (2014) received the highest WoE B out of all the studies as it used a randomised experimental design with a comparable treatment as usual (TAU) group as the control. This more rigorous experimental design increases the internal validity of the study and reduces the effect of extraneous variables, allowing more reliable and valid conclusions to be made about the effects of PT.

Gallagher (2006) and Chiesa and Robertson (2000) used similar quasi-experimental designs, with the PT group made up of pupils identified by their class teacher as having maths difficulties, and the rest of the class acting as the control group. However, the control group in the Gallagher(2006) study was not comparable to the PT group and these differences were not accounted for when interpreting the results. In both studies, the PT groups were much smaller than the control groups, reducing the power of the findings and leading to reduced WoE A ratings.

The methodological quality of the group experimental and quasi-experimental studies was examined used the Gersten et al. (2005) protocol, specifically designed for research in special education. See Appendix C for the criteria and rationale for its use, and Appendix D for a copy of a coded protocol. As Weisenburgh-Snyder et al. (2015) and Bullara et al. (1993) both used small N-case design studies, the Horner et al. (2005) protocol was used to assess the methodological quality of these design types. See Appendix C for the criteria and rationale for its use, and Appendix E for a copy of a coded protocol. To receive a higher WoE A rating these two studies needed to demonstrate meeting a number of essential and desirable criteria, however, both scored a low rating due to many of these criteria being unmet. For example, both studies scored zero on ‘experimental control/internal validity’ as they failed to demonstrate experimental effect at three different time points, control for threats to internal validity, and failed to demonstrate experimental control. Lacking these criteria significantly reduces the methodological quality of the design as the recorded changes in maths ability could have occurred without the PT intervention, such as being due to maturation effects or continuation with the normal curriculum. These significant limitations led both the case studies to receive a low WoE A score. They also both received a low WoE B score as their use of methodology is considered less relevant to the current review question investigating effectiveness.

Intervention content and fidelity

To be able to compare the five studies and generalise their findings, it was expected that the studies would have high fidelity and use a PT intervention which included the five key components of PT outlined by Solity and Bull (1987).

1. *The desired pupil performance should be precisely defined in observable and measurable terms* – All five studies clearly outlined how many correct responses needed to be met to indicate fluency. Weisenburgh-Snyder et al. (2015) and Bullara et al. (1993) also outlined how many incorrect answers they were aiming for which additionally reflected the accuracy of the pupils’ performance.

2. *Record the pupil’s performance on a daily basis with the use of a probe for one minute* – All the studies included a test phase using a probe which assessed how many correct and incorrect answers pupils achieved. All studies used a one minute test time, except Weisenburgh-Snyder et al. (2015) who used multiple 30-second test periods and averaged them to produce one daily ‘correct answers per minute’ score. This method of testing, which still produces a number of correct answers per minute, is consistent with the literature on PT and therefore this component was deemed to be met by all five studies, reflecting high fidelity.

3. *Chart the pupil’s performance of the probe on a daily basis, using a SCC* – All five studies used SCC’s to chart pupil’s performance.
4. *Record the teaching approach and behaviour in relation to the pupil performance, allowing for task analysis* – Weisenburgh-Snyder et al. (2015) was the only study which explicitly stated that task analysis was carried out. As the other four studies did not report this, it was assumed this was not done and their WoE C rating was reduced.
5. *Analyse the data to determine whether progress is satisfactory or whether adaptations are needed to maintain or accelerate learning* – All the studies reported analysing the data and reviewing pupil progress by studying the SCCs. Gallagher (2006) and Chiesa and Robertson (2000) reviewed pupil progress weekly, but as authors, they were not blind to the conditions and experimenter bias is possible. To overcome this problem, and to increase the reliability of the intervention findings, Strømgren et al. (2014) report the use of inter-observer coding. Two independent coders scored and recorded the pupils’ correct answers per minute. The Inter-observer Agreement (IOA) scores were all very high; 99.8% at pre-intervention, 99.9% at post-intervention and 99.9% at follow up.

Strømgren et al.’s (2014) inclusion of all five components of PT outlined by Bull and Solity (1987) meant it was assigned the highest possible WoE B rating of 3.

Delivery

In practice, PT is usually delivered by TAs in schools following training by an EP. However, in Gallagher (2006), and Chiesa and Robertson (2000), each pupil in the PT intervention was given a personalised folder containing practice sheets, timed probes and SCCs to be completed on a daily basis, independent of a teacher. Although the pupils had received some initial training in how to complete their folders, the fidelity of how the programme was delivered is questionable. It is not known how closely the instructions were followed by the pupils, or how generalisable the findings are to the normal class setting. Bullara et al. (1993) used a PT teacher in a clinic and Strømgren et al. (2015) used PT instructors; both of which are poorly described and thus difficult to replicate. Weisenburgh-Snyder et al. (2015) was the only study to report that the normal classroom teacher delivered the intervention to the pupils, and this was reflected by the higher WoE A rating.

Measures

Four of the five studies used only the SCC data to measure pupil performance, which reflected their performance on the probe task. The limitation of only using such data is that it was derived from composite tasks created by the class teacher, and thus are not valid and standardised measures, and may not even test the underlying skill being targeted. Weisenburgh-Snyder et al. (2014) was the only study to use a standardised measure, the Iowa Test of Basic Skills (ITBS), which includes a battery of maths subtests and provides an overall maths attainment score. This is a more reliable measure, with an internal reliability coefficient of between .929 and .972 for the ages tested. To validate and further support the data provided in the current studies, additional measures would have ideally been used, and this is reflected in the lower WoE A ratings for all the studies.

Findings

Although all five studies found that PT increased pupils' maths attainment, Strømgren et al. (2014) was the only study to report effect sizes. The effect sizes of pre-post and pre-follow up were reported as Cramer's V, but have been converted into Cohen's d to enable easier comparison with other studies (see Table 6). Gallagher (2006) and Chiesa & Roberston (2000) both reported the individual pupils' pre and post intervention scores, as well as group means and ranges but did not apply statistical analysis to this data. Gallagher (2000) argues that as a Behaviourist inductive approach was adopted, which favours observable behavioural change, statistical analysis was unnecessary. The paper argues that observing increased scores is sufficient evidence of the effectiveness of PT. However, unlike Chiesa and Robertson (2000) who evidenced that the control group was comparable to the PT group, Gallagher (2006) did not ensure the two groups were comparable pre-intervention and thus the results of this paper lack considerable validity, reflected by its low WoE A.

Using the data provided in both papers, it was possible to calculate effect sizes using a pre-test post-test standardised mean difference (PPSMD; Becker, 1988) to establish whether the interventions were effective. These results are presented alongside those of the Strømgren et al. (2014) paper in Table 6 and conclude PT to have a medium to large effect size when used to teach numeracy skills to pupils with maths difficulties. This suggests that PT has positive implications and should be recommended for use in schools as a wave 3 intervention, particularly for those with SEN who may be significantly behind their peers.

However, such results should be interpreted with caution due to a number of design flaws. In addition to those issues already discussed, these studies all had relatively small numbers of participants, with a maximum of 24 participants receiving PT in the Strømgren et al. (2014) paper. Having small sample sizes increases the likelihood of a larger effect size being found (Slavin & Smith, 2009). Indeed, Cohen (1992) advises that in order to detect a large effect size a group size of at least 26 participants is needed and for a medium effect size, 64 participants are required. Therefore, these studies were underpowered and their results must be interpreted with caution. It is likely that if they were replicated with larger sample sizes, such effect sizes would not be found.

The two small-N case design studies reported insufficient data to calculate effect sizes, but their results are presented in Table 7. Both papers claim their results, of large increases in correct answers and minimal errors, demonstrates PT to be an effective intervention in improving pupils' maths performance. However, neither studies conducted sufficient data analysis and this, paired with their low methodological quality, means these studies lack reliability and validity. Confident conclusions cannot be drawn on the limited data provided.

Table 6: *Effect sizes*

		PT			Control						
Authors	**Outcome Measure**	**N**	**Pre M (SD)**	**Post M (SD)**	**N**	**Pre M (SD)**	**Post M (SD)**	**Effect size**	**p**	**Effect Size Descriptor**	**WoE D**
Gallagher (2006)	Composite task consisting of 80 multiplication problems	8	16.38 (5.07)	22.75 (5.92)	15	26.87 (4.75)	27.13 (4.58)	d = 1.26	-	**Large**	1.8 (Medium)
Chiesa and Robertson (2000)	Composite task consisting of division problems	5	1 (0.71)	13.20 (2.05)	20	3.7 (3.01)	4.2 (3.49)	d = 4.25	-	**Large**	1.9 (Medium)
Strømgren, Berg-Mortensen and Tangen (2014)	20 minute test consisting of a total of 250 multiplication and division problems	22	5.16 (2.87)	8.04 (4.59)	23	5.07 (2.21)	6.21 (2.05)	Pre-post: d = 0.65	.043*	**Medium**	2 (Medium)
		19		FU: 9.1 (6)	23		FU: 7.17 (2.86)	Pre-FU: d = 0.98	.005*	**Large**	

*significant at p=<0.05

Note: effect sizes are considered: d=0.2 is small, d=0.5 is medium, and d=0.8 is large.

Table 7: *Outcomes for case study designs*

Authors	Outcome Measure	Outcomes	WoE D
Weisenburgh-Snyder, Malmquist, Robbins and Lipshin (2015)	Iowa Test of Basic Skills (ITBS)	*Pre intervention ITBS score:* Mean maths total standard score = 184.0 Number of students performing at grade level = 2 *Post intervention ITBS score:* Mean maths total standard score = 224.8 Number of students performing at grade level = 10 Mean ITBS standard score difference = +40.8 (equivalent of +2.7 years)	1.6 (Low)
Bullara, Kimball and Cooper (1993)	Composite task consisting of sums up to 10.	*Pre intervention:* 23 correct, 6 errors *During intervention:* Median score = 67 correct, 0 errors *Three month follow up:* Median score = 53 correct, 0 errors.	1.1 (Low)

Section 4: Conclusion and Recommendations

All five studies included in this current literature review found PT to improve the basic numeracy skills of pupils with maths difficulties. Impressive increases in the number of correct answers, age equivalent scores and maths attainment levels were reported across the studies. Long term effects of PT were also evidenced by Bullara et al. (1993) who found their case study pupil to retain the numeracy skills taught in PT following a three month break from the intervention. Although the two small-N case studies provided insufficient data and inappropriate data analysis, the remaining three studies found PT to have a medium to large effect size. These findings suggest that PT is effective in improving numeracy skills for school-aged pupils with maths difficulties, and thus EPs may wish to consider recommending it to schools as a cost-effective intervention for this population of pupils. These findings also provide further evidence that maths difficulties may be due to pupils not developing fluent numeracy skills (Binder, 1988), and PT can be used to effectively address this issue.

However, it has been noted that due to the limited quality and relevance of the methodology used, and the limited relevance of some of the studies to the current review question, these results should be interpreted with caution. All five studies were found to have low methodological quality as reflected by their low WoE A ratings. The studies failed to provide enough information about their participants and control conditions, and all but one study used unreliable and non-standardised measures of maths attainment. Most of the studies were unable to eliminate alternative explanations for the increased scores reported as they did not use rigorous and relevant methodology and lacked internal reliability. These limitations were reflected in the studies WoE A and B. Only two of the studies were conducted in UK schools, and thus the other three studies also lacked ecological validity and generalisability to the UK education system.

Furthermore, a critical review of the studies found that not all the studies were comparable in how PT was delivered, and only Weisenburgh-Snyder et al. (2014) met all five components of PT as presented by Solity and Bull (1987). This is a significant limitation of the studies included in the review as they had reduced fidelity.

Considering the limitations of the studies reviewed, the following recommendations are suggested for future research to improve upon the current findings and to allow for more confident conclusions to be drawn:

- To provide more data about participants included in research studies to identify which pupils would benefit most from a PT intervention.
- PT interventions to include Solity and Bull's (1987) five components of PT to increase intervention fidelity and allow for easier comparison between studies.
- To use more rigorous experimental designs such as randomised control trials to increase the validity and reliability of research findings.
- To use larger sample sizes and more sophisticated and appropriate data analysis, with effect sizes, to address the issues of studies being underpowered and lacking appropriate analysis of findings.

References

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41,* 257–278.

Binder, C. (1988). Precision Teaching: measuring and attaining exemplary academic achievement. *Youth Policy*, *10*(2), 12_15.

Binder, C., & Watkins, C. L. (2013). Precision teaching and direct instruction: Measurably superior instructional technology in schools. *Performance Improvement Quarterly*, *26*(2), 73-115.

Blatchford, P., Russell, A., & Webster, R. (2012). *Reassessing the impact of teaching assistants: How research challenges practice and policy*. Routledge.

Bullara, D., Kimball, J., & Cooper, J. (1993). An assessment of beginning addition skills following three months without instruction or practice. *Journal of Precision Teaching, 6*(1), 11-16.

Chiesa, M., & Robertson, A. (2000). Precision teaching and fluency training: Making maths easier for pupils and teachers. *Educational Psychology in Practice, 16*(3), 297-310.

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

DfE (2014). *Influences on students' GCSE attainment and progress at age 16: Effective Pre-School, Primary and Secondary Educational Project (EPPSE) Research Report.* Retrieved from: https://www.ucl.ac.uk/ioe/research/pdf/16-Influences-Students-GCSE-Attainment-Progress-RR.pdf

Education and Skills Funding Agency (2014*). 16 to 19 funding: maths and English condition of funding*. Retrieved from: https://www.gov.uk/guidance/16-to-19-funding-maths-and-english-condition-of-funding

Gallagher, E. (2006). Improving a mathematical key skill using precision teaching. *Irish Educational Studies, 25*(3), 303-319.

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., & Cook, R. J. (1995). Users' guides to the medical literature. *Jama*, *274*(22), 1800-04.

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.

Haring, N.G., Lovitt, T.C., Eaton, M.D., & Hansen, C.L. (1978). *The fourth R: Research in the classroom.* Columbus, OH: Charles E. Merrill Publishing Co.

Haughton, E. (1972). Aims-growing and sharing. In J. B. Jordan & L. S. Robbins (Eds.), *Let's try doing something else kind of thing: Behavioral principles and the exceptional child. A report from the invisible college conference on application of behavioral principles in exceptional child education, March, 1971*. Arlington, VA:The Council for Exceptional Children.

Lindsley, O. R. (1972). From Skinner to precision teaching: The child knows best In JB, Jordan, LS Robbins,(Eds.), Let's try do-ing something else kind of thing (pp. 1-11). *Arlington, VA: Council on Exceptional Children*.

Lindsley, O. R. (1990). Precision teaching: By teachers for children. *Teaching Exceptional Children*, *22*(3), 10-15.

Lindsley, O. R. (1992). Precision Teaching: discoveries and effects. *Journal of Applied Behavior Analysis, 25,* 51-57.

McDowell, C., & Keenan, M. (2001) Cumulative dysfluency: still evident in our classrooms despite what we know. *Journal of Precision Teaching and Celeration*, *17*(2), 1-6.

Potts, L., Eshleman, J. W., & Cooper, J. O. (1993). Ogden R. Lindsley and the historical development of precision teaching. *The Behavior Analyst*, *16*(2), 177-189.

Skinner, B. F. (1985). Cognitive science and behaviourism. *British Journal of psychology*, *76*(3), 291-301.

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, *31*(4), 500-506.

Solity. J., & Bull, S. (1987). *Special needs: Bridging the curriculum gap.* Milton Keynes: Open University Press.

Strømgren, B., Berg-Mortensen, C., & Tangen, L. (2014). The use of precision teaching to teach basic math facts. *European Journal of Behavior Analysis, 15*(2), 225-240.

Vygotsky, L. (1987). Zone of proximal development. *Mind in society: The development of higher psychological processes*, *5291*, 157.

Weisenburgh-Snyder, A., Malmquist, S., Robbins, J., & Lipshin, A. (2015). A model of MTSS: Integrating precision teaching of mathematics and a multi-level assessment system in a generative classroom. *Learning Disabilities: A Contemporary Journal, 13*(1), 21-41.

West, R. P., Young, R., & Spooner, F. (1990) Precision teaching: an introduction, Teaching Exceptional Children, 22(3), 4-9.

Appendix A: Excluded Studies Following Full Text Review

Reference	Exclusion Criteria
Campbell, K. U., & McCarthy-Jensen, D. (1982). Digit typing for micro-computer math programs: Toward a determination of proficiency standards. *Journal of Precision Teaching, 3*(3), 68, 70-71.	2
Campbell, K. U., & McCarthy-Jensen, D. (2000). Digit typing for micro-computer math programs: Toward a determination of proficiency standards. *Journal of Precision Teaching and Celeration, 17*(1), 14, 16-17.	2
Desjardins, E. A. (1993). Teaching addition and subtraction word problems. *Journal of Precision Teaching, 10*(2), 25-28.	8
Eaton, M., Partlow, C., Hermann, K., & Lankenau, A. (1984). Doubles, neighbors and two houses away: A quick trip to aim on addition facts. *Journal of Precision Teaching, 5*(3), 49-53.	2
Ferrucci, B. J. (1984). *The relationship of precision teaching to the speed, computational skill development, and concept development of working with rational numbers* (Doctoral dissertation, Boston University).	1
Hayden, J., & McLaughlin, T. F. (2004). The effects of cover, copy, and compare and flash card drill on correct rate of math facts for a middle school student with learning disabilities. *Journal of Precision Teaching and Celeration, 20*(1), 17-21.	6

Reference	Exclusion Criteria
Irwin, K. C. (1991). Teaching children with Down syndrome to add by counting-on. *Education and Treatment of Children*, 128-141.	6
Johnson, K. R., & Layng, T. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American psychologist*, *47*(11), 1475.	7
Johnson, K. R. (1993). How to build fluency in reading, math, and writing: An overview. *Journal of Precision Teaching, 10*(2), 14-15.	7
McCarthy-Jensen, D., & Campbell, K. U. (1983). A computerized math deficit remediation. *Journal of Precision Teaching, 4*(1), 14-15.	2
McTiernan, A. (2014). *Behavioural fluency: An investigation of instructional approaches to build fluency with academic skills and associated outcomes* (Doctoral dissertation). National University of Ireland, Galway.	1
Mortenson, B. (2001). *A behavior analytic approach to developing mathematics skills in second graders: A precision teaching intervention* (Unpublished doctoral dissertation). Louisiana State University, US.	1
Morton, R. C., & Flynt, S. W. (1997). A comparison of constant time delay and prompt fading to teach multiplication facts to students with learning disabilities. *Journal of Instructional Psychology*, *24*(1), 3.	6
Oskar-Groen, K. (2009). *Using repeated practice and error correction procedures combined with Precision Teaching measurement approaches to improve the math calculation fluency of middle school*	1

Reference	Exclusion Criteria
students with emotional and behavioural disorders (Unpublished doctoral dissertation). University of South Dakota, US.	
Peterson, S. K. (1985). Number reversals: An effective intervention. *Journal of Precision Teaching, 6*(1), 20-23.	2
Poe, A. L. (1993). Mad about math (or "Junior ~ExSEL"). *Journal of Precision Teaching, 11*(1), 44-49.	6
Raggio, S., & Bitgood, S. C. (1982). The effect of number of math drills per day on math performance. *Journal of Precision Teaching, 3*(2), 50-51.	2
Raggio, S., & Bitgood, S. C. (1999). The effect of number of math drills per day on math performance. *Journal of Precision Teaching and Celeration, 16*(2), 58-60.	2
Spencer, M. L., & Henderson, J. C. (1975). Santa Fe School Precision Teaching Program, Evaluation Report 1974-75.	2
Spillman, S., & Milyko, K. (2015). I can do this math, yo! *Behavioral Development Bulletin, 20*(2), 173-176.	6
Tobin, M. (1989). *The effects of an intervention using a precision teaching model on grade four students' arithmetic self-concept* (Unpublished doctoral dissertation). University of British Columbia, Canada.	1
Ward, J., Crawford, S., & Solity, J. (2007). Applying assessment through teaching and instructional psychology: An alternative model of	8

Reference	Exclusion Criteria
service delivery to raise attainment in primary schools. *Educational and Child Psychology, 34*(1), 94-109.	
Whalen, K. P., Willis, R. J., & Sweeney, W. J. (1993). Using 1-minute time trials and 4-minute practice sessions to improve a student's performance of fraction problems. J*ournal of Precision Teaching, 11*(1), 2-10.	6
Woollard, W. J. (1986). Behavioural Objectives in Mathematics for Secondary School Slow Learners. *Educational Psychology in Practice*, *2*(1), 23-27.	7

Appendix B: Mapping the Field

	Author	Aim of study	Participants	Intervention details	Measures	Outcomes
1	Gallagher (2006)	To investigate the benefits of PT for teaching pupils multiplication timetables.	All participants were either 10 or 11 years old in a primary school in Northern Ireland. PT group n=8 (5 males, 3 females). Participants in the PT group were identified by their class teacher as being behind their peers in maths and not taking the year 7 Transfer Test. Control group n=15 (9 males, 6 females). This group was the rest of the class year 7 class who were sitting the Transfer Test.	12 week intervention, daily maths input for both groups. The PT group worked through a folder prepared by the author which included practice sheets, timed worksheets to complete in one minute drills, answer sheets and SCC. During this time, the control completed daily revision for the Transfer Test.	A composite task was used with all participants pre and post the intervention. The task consisted of 80 random multiplication problems.	The mean score for the PT group increased from 16.38 to 22.75 correct answers per minute. This is an increase of 6.38 correct answers. The control group increased by 0.27 in the same time. By the end of the intervention, 3 of the 8 pupils in the PT group were able to perform at levels above the mean of the control group indicating they had caught up with peers.
2	Chiesa & Robertson (2000)	To investigate the use of PT and fluency training in teaching pupils maths skills.	All participants were either 9 or 10 years old, and were all the in the same year 5 class in a primary school in the UK. PT group n=5 (3 males, 2 females). Participants in the PT group were identified by their class teacher and learning support teacher as	12 week intervention. Groups were matched so that they spent the same amount of time engaged in maths problems. Each pupil in the PT group had a personal folder which included practice sheets, timed probe sheets and charts. Once a week, the second author met with the PT pupils for 30	A composite task was used pre and post intervention. The task were division sums of two-digit numbers divided by a one digit number (e.g. 75 ÷ 3).	The mean score for the PT group increased from 1 to 13.2 correct answers post intervention. The control group increased from 3.7 to 4.2 correct answers post intervention. Following the intervention, there was a significant difference between the two groups in their correct answers on the composite task.

	Author	Aim of study	Participants	Intervention details	Measures	Outcomes
			being behind their peers in maths. Control group n=20 was the rest of the year 5 class.	minutes each to review their progress.		
3	Strømgren, Berg-Mortensen & Tangen (2014)	To evaluate an eight week PT intervention to teach basic maths facts.	All participants (n=48) were aged between 10 and 13 years, attending primary school in Norway. All pupils had been nominated by their class teacher as lagging behind the rest of their class in basic multiplication and division skills. Participants were randomly allocated to either a PT intervention group (n=24; 8 males, 16 females) or treatment as usual (TAU) group (n=24; 12 males and 12 females).	8 week intervention, undertaken for 25 minutes daily in school. The PT group had their own folder containing timing charts, SCC and worksheets with multiplication and division sums.	A test consisting of 5 worksheets with 250 mixed multiplication and division problems (50 per sheet) was used pre, post and at a one month follow up of the intervention with all participants. Participants had 20 minutes to complete the test.	Both the PT and TAU group made progress, with the PT group making greater progress. The mean correct answers per minute (CAPM) pre intervention for the PT group was 5.16, compared with 5.07 for the TAU group. Post intervention the PT group had a mean CAPM of 8.04 and the TAU group had 6.21. At a one month follow up, the PT group had a mean CAPM of 9.10 and the TAU group had a mean of 7.17. A post intervention questionnaire also reported positive feedback from the PT group about the intervention.
4	Weisenburgh-Snyder, Malmquist, Robbins & Lipshin (2015)	To present a case study of a school using PT to teach maths within a multi-level school assessment system over	Participants (n=10) were all male pupils in a private elementary (primary) school in the USA who were identified as being behind their peers academically. Four participants were identified as having a	The intervention was conducted across one academic year from September to June. The maths intervention was delivered daily during the school week by a school teacher. A PT based programme called 'Diagnostic/Prescriptive Math	Maths subtests from the Iowa Test of Basic Skills (ITBS) were used to measure participants' maths score pre and post intervention.	All participants demonstrated accelerated rates of improvement in the essential maths skills. The participants' mean ITBS score increased from 184.0 pre intervention to 224.8 post intervention (increase of 40.8) which was a

	Author	Aim of study	Participants	Intervention details	Measures	Outcomes
		one academic year.	learning disability, two as having attention deficit disorder and one as having a behaviour disorder.	Computation' was part of the maths intervention, which also included direct teaching from the teacher and working independent through worksheets. Participants were expected to complete at least three one-minute timings a day and record their best performance on standard celeration charts.		grade equivalent of +2.7 years and an increase of 40.8% in percentile ranks. In the computation subtest participants increased by a mean of +4.5 grade equivalent years.
5	Bullara, Kimball & Cooper (1993)	To investigate the use of Precision Teaching to teach a primary aged pupil addition to 10 and whether this is maintained after four months.	This case study was of one pupil aged 11 years in the USA. The pupil was identified as having poor attainment in maths and had not received PT programme before. The pupil spent a portion of his school day in a resource room for pupils with specific learning disabilities, and the rest in mainstream class. The study took place between grades 4 and 5 (UK equivalent of year 5 and 6).	The PT intervention was delivered at The Ohio State University Psychoeducational Clinic twice a week for one hour. The intervention lasted for 10 weeks and then there was a three month break during the summer holidays in which the pupil received no PT. The pupil returned to the clinic after three months and a long term retention assessment was undertaken to investigate how many sums taught on the PT programme were retained by the pupil. The pupil received an additional 10 weeks of PT for maths in this Autumn term.	A one minute test of sums to 10 in which the pupil had to write his answers, was used to measure the pupil's performance pre and post PT intervention.	The pupil's performance of the one-minute addition sums increased from 23 correct and 6 incorrect answers pre intervention, to 74 with no incorrect answers. At the three month follow up, the pupil showed a slight decrease of 56 correct answers and incorrect answers remained near zero.

Appendix C: Weight of Evidence Ratings and Rationale

Weight of Evidence A: Methodological Quality Design

1. *WoE A for group and quasi experimental studies*

The Gersten et al. (2005) protocol for ‘group experimental and quasi-experimental research articles and reports in special education’ was used in full to appraise Gallagher (2006), Chiesa & Robertson (2000) and Strømgren et al. (2014). This protocol was used as it outlines ten essential and eight desirable criteria which are expected to be present in good quality research in this field (see Appendix D for an example of the Gersten et al. (2005) protocol). For each criteria, the coder indicated whether the criteria was met and these were summed to show a total number of essential and desirable criteria met for each study. Table 1 outlines how many criteria needed to be met to produce a WoE A rating of 1, 2 or 3. A rating of 3 indicates that nearly all the criteria were met and the paper used a high methodological quality, making it more valid and reliable in answering the current research question. All three studies scored only 1 for WoE A, reflecting their poor methodological quality.

Table 1: *Gersten et al. (2005) coding criteria and rationale for WoE A*

WoE A Rating	Criteria	Rationale
3	Meets ≥ 9 essential criteria and ≥ 4 desirable criteria	Based on the recommendations made by Gersten et al. (2005)
2	Meets ≥ 9 essential criteria and ≥ 2 desirable criteria	
1	Meets <9 essential criteria and <2 desirable criteria	

Table 2: *WoE A for group and experimental designs*

Study	No. of essential criteria met	No. of desirable criteria met	WoE A Rating
Gallagher (2006)	3	2	1
Chiesa & Robertson (2000)	4	2	1
Strømgren, Berg-Mortensen & Tangen (2014)	5	3	1

2. WoE A for small N studies

The Horner et al. (2005) protocol for small N studies was used in full to calculate WoE A for the Weisenburgh-Snyder et al. (2015) and the Bullara et al. (1993) studies, as both employed a small-N case design. The quality indicators within single-subject research in Table 1 of the Horner et al. (2005, p.174) paper were used as criteria for WoE A. See Table 3 for a list of criteria and ratings for this protocol. The descriptors provided in the paper were used to give numerical ratings of 1 to 3 under seven different criteria headings (A to G). To calculate the total WoE A for these two studies, the seven criteria ratings (A to G) were summed and divided by 7 for each study. As seen in Table 4, both these studies scored a low WoE A, reflecting poor methodological quality.

Table 3*: Horner et al. (2005) coding criteria*

Criteria	Rating descriptors (Horner et al., 2005)
A: Description of participants and settings	3: All 3 of the following criteria are fulfilled: participants are described with sufficient detail to allow others to select individuals with similar characteristics (e.g., age, gender, disability, diagnosis); the process for selecting participants is described with replicable precision; critical features of the physical setting are described with sufficient precision to allow replication.

Criteria	Rating descriptors (Horner et al., 2005)
	2: Two of the criteria above are fulfilled. 1: One of the criteria above is fulfilled.
B: Dependent variable	3: All of the following criteria are fulfilled: dependent variables are described with operational precision; each dependent variable is measured with a procedure that generates a quantifiable index; measurement of the dependent variable is valid and described with replicable precision; dependent variables are measured repeatedly over time; data are collected on the reliability or inter-observer agreement associated with each dependent variable, and IOA levels meet minimal standards (e.g., IOA = 80%; Kappa = 60%). 2: Three to four of the above criteria are fulfilled. 1: One to two of the above criteria are fulfilled.
C: Independent variable	3: All of the following criteria are fulfilled: independent variable is described with replicable precision; independent variable is systematically manipulated and under the control of the experimenter; overt measurement of the fidelity of implementation for the independent variable is highly desirable. 2: Two of the above criteria are fulfilled. 1: One of the above criteria is fulfilled.
D: Baseline	3: All of the following criteria are fulfilled: the study includes a baseline phase that provides repeated measurement of a dependent variable; the study establishes a pattern of responding that can be used to predict the pattern of future performance, if introduction or manipulation of the independent variable did not occur; baseline conditions are described with replicable precision. 2: Two of the above criteria are fulfilled. 1: One of the above criteria is fulfilled.
E: Experimental Control/Internal Validity	3: All of the following criteria are fulfilled: the design provides at least three demonstrations of experimental effect at three different points in time; the design controls for common threats to internal validity (e.g., permits elimination of rival hypotheses); the results document a pattern that demonstrates experimental control.

Criteria	Rating descriptors (Horner et al., 2005)
	2: Two of the above criteria are fulfilled. 1: One of the above criteria is fulfilled.
F: External Validity	3: Experimental effects are replicated across 3 or more participants and also either settings or materials to establish external validity. 2: Experimental effects are replicated across 3 or more participants 1: Experimental effects are replicated by inclusion of at least 2 participants.
G: Social validity	3: All of the following criteria are fulfilled: the dependent variable is socially important; the magnitude of change in the dependent variable resulting from the intervention is socially important; implementation of the independent variable is practical and cost effective; social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts. 2: Two to three of the above criteria are fulfilled. 1: One of the above criteria is fulfilled.

Table 4: *WoE A criteria ratings for small N case studies*

	Criteria							***WoE A***
Study	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**Total**
Weisenburgh-Snyder, Malmquist, Robbins & Lipshin (2015)	2	2	1	0	0	2	2	1.3
Bullara, Kimball & Cooper (1993)	1	2	1	1	0	0	1	0.9

Weight of Evidence B: Relevance of Methodology

WoE B reviewed how relevant the methodology used in each paper was in answering the current research question. Table 5 outlines the ratings assigned to different study designs, based upon published evidence hierarchies which list study designs in order, based on how rigorous and effective they are in answering research questions.

Table 5: *WoE B criteria and rationale*

WoE B Rating	Type of study design	Rationale
3	Randomised control trial studies	Evidence hierarchies (Guyatt et al., 1995) have rated different types based on their effectiveness in answering research questions of effectiveness of interventions.
2	Cohort studies and quasi-experimental studies	
1	Qualitative research, case studies and non-experimental evaluations	

Weight of Evidence C: Relevance of Focus of Study

WoE C was assessed using three criteria which looked at how relevant the five studies were to the review question. Each criterion had a possible three scores. These scores were summed and divided by three for each study, to calculate an overall WoE C score. Table 6 outlines each criteria, rating descriptors and the rationale for why each criteria was used in this current review.

Table 6: *WoE C criteria and rationale*

Criteria	Rating descriptors	Rationale
A: Sample	3- Specified as having mathematical difficulties, supported with assessment data.	To investigate the use of PT with pupils with identified mathematical difficulties.

	2- Identified as having mathematical difficulties by teacher with no assessment data to support this. 1- No mathematical difficulties reported	
B: Location	3 - UK and school setting 2- Not UK but in school setting 1- Not UK and not in school setting	To identify interventions that could be implemented in UK schools.
C: Fidelity	3 – Includes all of the following five components as outlined by Solity and Bull (1987): desired pupil performance defined in observable and measurable terms, daily use of probe, chart performance on SCC, record teaching approach for task analysis, and analyse the data to determine whether adaptations are needed. 2- Includes 3 or 4 of the components 1 – Includes 2 or less of the components	Solity and Bull (1987) outlined five key components of PT, which must be followed to ensure the fidelity of the programme. These are the five components that would be taught to schools by Educational Psychologists, therefore for the results of the studies to be applicable to practice in schools, the studies should have included all five key components.

Table 7: *WoE C ratings for the five studies included in the review*

Paper	**Score**	**Criteria**		**Average**
	A	**B**	**C**	
Gallagher (2006)	2	3	2	**2.3**
Chiesa & Robertson (2000)	2	3	3	**2.7**

Strømgren, Berg-Mortensen & Tangen (2014)	2	2	2	**2**
Weisenburgh-Snyder, Malmquist, Robbins & Lipshin (2015)	3	2	2	**2.3**
Bullara, Kimball & Cooper (1993)	2	1	1	**1.3**

Appendix D: Use of the Gersten et al. (2005) Protocol

Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education

Gersten, Fuchs, Compton, Coyne, Greenwood & Innocenti (2005)

Name of Coder: ANONYMOUS

Date: 24/01/2018

Full Study Reference: Chiesa, M. & Robertson, A. (2000). Precision teaching and fluency training: Making maths easier for pupils and teachers. *Educational Psychology in Practice, 16*(3), 297-310.

Intervention Name: Precision Teaching

Research design: Quasi experimental

Type of Publication: Peer-reviewed journal

Essential Quality Indicators

Describing Participants

1. Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?

☐ Yes

☒ No

☐ N/A

☐ Unknown/Unable to Code

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

☒ Yes – No randomisation, but the two groups were from the same class (e.g. same age and comparable in other aspects of the curriculum). There were no significant differences between the scores of the two groups pre intervention.

☐ No

☐ N/A

☐ Unknown/Unable to Code

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

☐ Yes

☐ No

☒ N/A - The study sought no active participation of the class teacher. The pupils worked through a prepared folder of activities by the author of the study. There were weekly visits by the author to review the progress of the pupils.

☐ Unknown/Unable to Code

Implementation of the Intervention and Description of Comparison Conditions

4. Was the intervention clearly described and specified?

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

5. Was the fidelity of implementation described and assessed?

☐ Yes
☒ No - The intervention required pupils to work through prepared folders. The author reviewed the pupils progress on a weekly basis, but there were no checks of fidelity e.g. observations with checklist of treatment components.
☐ N/A
☐ Unknown/Unable to Code

6. Was the nature of services provided in comparison conditions described?

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

Outcome Measures

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

☐ Yes
☒ No – Only pre and post intervention measures for the composite task were provided for the PT and control group. The composite task was designed by the
☐ N/A
☐ Unknown/Unable to Code

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

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

Data Analysis

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

☒ Yes – Compared pre and post intervention data for the PT vs control group
☐ No
☐ N/A
☐ Unknown/Unable to Code

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

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

Desirable Quality Indicators

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

☒ Yes – All PT participants completed the 12 week intervention
☐ No
☐ N/A
☐ Unknown/Unable to Code

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

☐ Yes
☒ No – Pupils plotted their own data independently. This was reviewed once a week by the author of the study (not blind).
☐ N/A
☐ Unknown/Unable to Code

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

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

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

☐ Yes
☒ No – Composite task created by teacher. No evidence of validity.
☐ N/A

☐ Unknown/Unable to Code

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

☐ Yes

☒ No - The pupil self-delivered the intervention. No examination of quality of implementation.

☐ N/A

☐ Unknown/Unable to Code

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

☐ Yes

☒ No – No info provided about control condition.

☐ N/A

☐ Unknown/Unable to Code

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

☐ Yes

☒ No

☐ N/A

☐ Unknown/Unable to Code

8. Were results presented in a clear, coherent fashion?

☒ Yes

☐ No

☐ N/A

☐ Unknown/Unable to Code

Number of essential criteria met: 4
Number of desirable criteria met: 2

Overall Rating of Evidence: ☐ 3 ☐ 2 ☒ 1

Appendix E: Use of the Horner et al. (2005) Protocol

Article Reference: Bullara, D., Kimball, J., & Cooper, J. (1993). An assessment of beginning addition skills following three months without instruction and practice. *Journal of Precision Teaching, 6*(1), 11-16.

Section A: Description of Participants and Setting

1) Participants are described with sufficient detail to allow others to select individuals with similar characteristics; (e.g., age, gender, disability, diagnosis).

- Yes
- No ✓ - Age, ethnicity and gender described, but not sufficient information about diagnosis and disability.
- N/A
- Unknown/Unable to Code

2) The process for selecting participants is described with operational precision.

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

3) Critical features of the physical setting are described with sufficient precision to allow replication.

- Yes ✓
- No
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence: 1

Section B: Dependent Variable

1) Dependent variables are described with operational precision.

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

2) Each dependent variable is measured with a procedure that generates a quantifiable index.

- Yes ✓ - correct and incorrect answers
- No
- Unknown/Unable to Code

3) Measurement of the dependent variable is valid and described with replicable precision.

- Yes ✓ - The measurement is correct answers. This can be replicated.
- No
- N/A
- Unknown/Unable to Code

4) Dependent variables are measured repeatedly over time.

- Yes ✓
- No
- N/A
- Unknown/Unable to Code

5) Data are collected on the reliability or inter-observer agreement associated with each dependent variable, and IOA levels meet minimal standards (e.g., IOA = 80%; Kappa = 60%).

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence: 2

Section C: Independent Variable

Independent variable is described with replicable precision.

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

2) Independent variable is systematically manipulated and under the control of the experimenter.

- Yes ✓
- No
- N/A
- Unknown/Unable to Code

3) Overt measurement of the fidelity of implementation for the independent variable is highly desirable.

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence: 1

Section D: Baseline

The majority of single-subject research studies will include a baseline phase that provides repeated measurement of a dependent variable and establishes a pattern of responding that can be used to predict the pattern of future performance, if introduction or manipulation of the independent variable did not occur.

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

2) Baseline conditions are described with replicable precision.

- Yes ✓
- No
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence: 1

Section E: Experimental Control/Internal Validity

1) The design provides at least three demonstrations of experimental effect at three different points in time.

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

2) The design controls for common threats to internal validity (e.g. permits elimination of rival hypotheses).

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

3) The results document a pattern that demonstrates experimental control.

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence: 0

Section F: External Validity

1) Experimental effects are replicated across participants, settings, or materials to establish external validity.

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence: 0

Section G: Social Validity

1) The dependent variable is socially important.

- Yes ✓
- No
- N/A
- Unknown/Unable to Code

2) The magnitude of change in the dependent variable resulting from the intervention is socially important.

- Yes
- No
- N/A
- Unknown/Unable to Code ✓ - presentation of data does not allow visual analysis

3) Implementation of the independent variable is practical and cost effective

- Yes
- No
- N/A
- Unknown/Unable to Code ✓

4) Social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts.

- Yes
- No ✓
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence: 1

Ratings of Evidence:

A: 1

B: 2

C: 1

D: 1

E: 0

F: 0

G: 1

Average WoE A = 6/7 = 0.9

Average WoE A = Sum of X / N

X = individual quality rating for each judgement area N = number of judgement areas