

Case Study 1: Evidence Based Practice Report

Theme: School/setting Based Interventions for Learning.

How effective is precision teaching at increasing reading fluency in school age children?

Summary

Reading is a skill that impacts on all areas of learning. However, one in five children leave primary school without 'good enough' reading skills (DfES, 2015). This is consequently an area that schools need further support in improving. This review attempts to explore the effectiveness of using Precision Teaching to improve children's reading fluency. Precision Teaching is a method of daily intervention that focuses on the use of monitoring of improvement as a structure to increase fluency and maintenance of new skills.

Six studies were found through a literature search and were assessed using Gough's framework of review (Gough, 2007). Four studies were rated an overall score of 'medium' and two studies were rated 'low'. The evidence from the six studies reviewed suggest that precision teaching does improve reading fluency. However, the methodological designs of the studies were not high quality enough to enable any of the studies to be rated as 'high' using Gough's (2007) framework. Further research into the impact of Precision Teaching on reading fluency needs to be designed. Future recommendations are discussed.

Introduction

What is Precision Teaching?

Precision teaching (PT) was developed in 1965 by Ogden Lindsley as a method of improving learning for children with special educational needs (Lindsley, 1964). It can be thought of as a “precise and systematic method of instruction” rather than simply a method of teaching (West, Young & Spooner, 1990, p5). The aim of PT is to achieve fluency instead of just accuracy in a particular skill (Downer, 2007).

An essential part of PT is the use of Standard Celeration Charts (SCC) which are time series graphs, tracking correct and incorrect answers (Brosnan, Moeyaert, Brooks Newsome, Healy, Onghena & Van den Noortgate, 2018; Lindsley, 1992). These charts measure progress over time and display a learning trend (Brosnan et al., 2018; Kubina & Yurich, 2012). This enables the child to see their learning progress immediately, consequently developing their confidence in their own learning. PT also allows teachers to change and develop their teaching methods based on the daily feedback from the SCC.

There are four principles that should be adhered to when using PT as an intervention method: Use daily, the child knows best, use SCC, use self-recording (Lindsley, 1990). If these principles are sustained then the teacher should be aware of the type of instruction that suits the pupil and consequently adapt and change to suit the pupil’s particular learning (Lindsley, 1990). This explains why PT usually occurs on a 1:1 basis as the adaptations for a particular

learning style are essential for success. All children have different learning styles and these should be accommodated when teaching in order to enable success (Denig, 2004).

During a PT intervention, PT takes place daily and is supported by the SCC as well as time probes which detail the targets of learning (Brosnan et al., 2018). The probes are used to test the child's knowledge of a skill within 1 minute. In the case of this review the time probes all contained words to read and the number read both correctly and incorrectly was noted.

Johnson and Street (2013) stated that the PT intervention is usually characterised by five key steps:

1. The learning is pinpointed.
2. Specific resources including a 'time probe' are prepared.
3. The probe is used to monitor the student's performance.
4. The SCC is used to plot the correct and incorrect answers.
5. The SCC is used with student and teacher to celebrate successes and plan future sessions.

Psychological Theory

Ogden Lindsley was a former student of B.F Skinner and precision teaching was designed around the behavioural principles of operant conditioning, (West et al., 1990) the use of rate of response (probe) and cumulative response recording (SCC) (Lindsley, 1991). The behaviourist B. F. Skinner believed that through the use of a consequence to a behaviour learning can develop (Skinner, 1977). Lindsley took these principles and applied them to special

educational needs education and children (Lindsley, 1990) resulting in the design of the PT intervention.

A psychological model of learning that can be applied to PT is the Hierarchy of Learning (Haring, Lovitt, Eaton & Hansen, 1978). This model identifies five stages in the development of skills that a child acquires when learning something new. The five stages are acquisition, fluency, maintenance, generalisation and finally adaption. The first stage is where the child acquires a new skill, the teacher is crucial in this stage by providing the new knowledge in a way that is accessible and relevant to the child. The child then must be provided with opportunities to become fluent in this new skill before moving to the next stage where they are able to maintain the new knowledge they have acquired. The fourth stage is where the child is able to generalise that skill in different conditions before moving to the final fifth stage 'adaption' where the skill can be used independently to solve problems. PT is targeted between the fluency and maintenance stages and aims to enable the child to become fluent and maintain the knowledge (Brosnan et al., 2018).

Rationale and relevance to Educational Psychology

The ability to read is a key skill. Findings by the Organisation for Economic Co-operation and Development (2013) revealed that 16.4 % of adults in the United Kingdom function at the lowest level of literacy skills (level 1). This can prevent access to meaningful employment due to the necessities of reading across all areas of independent living such as completing job applications, etc.

The Department for Education revealed in 2015 that one in five children leave primary school unable to read to a level that will enable academic success at secondary school (DfES, 2015).

Therefore, it can be concluded that improving reading skills is a crucial requirement in schools across the UK. Targeting fluency is necessary in any intervention for children with difficulties in reading (Snowling & Hulme, 2010). PT is a cost effective (Binder, 1996) and efficient intervention as it requires only 10 minutes a day and teaching assistants can be trained in a short space of time. PT is a wave three intervention, it can be used across reading, writing and maths to support learning (Lindsay, 1992).

Educational psychologists should be providing advice related to evidence based interventions (Health & Care Professions Council, 2015) and for PT there is a base of evidence for effectiveness. Educational psychologists are a key part of implementing this intervention as they are in a position to provide training to school staff on implementation and resources. Educational psychologists are in the unique position of providing ongoing support to a school and can update training whenever necessary. Radford, Bosanquet, Webster, Blatchford and Rubie-Davies (2014) stated that in order for teaching assistants to be effective they should have been informed by relevant theories of teaching and learning.

Review Question

How effective is precision teaching in increasing reading fluency in school age children?

This review specifically looks at the effect of PT on reading fluency after schools changed their method of teaching reading as a result of the Rose Report (2006). Previous reviews have only evaluated the impact of PT prior to 2006.

Critical Review of the Evidence Base

Literature Search

A comprehensive search of the literature was undertaken in January 2020 using three databases. These provided access to articles from peer reviewed journals relating to psychology, education and the sciences. Table 1 shows the search terms used.

Table 1

Search Terms Used

Database	Search terms
PsychInfo	“precision teaching” OR “precision monitor” + “reading” or “read”
Web of Science	“precision teaching” OR “precision monitor” + “reading” or “read”
Education Resources Information Centre (ERIC)	“precision teaching” OR “precision monitor” + “reading” or “read”

The initial search produced 29 articles with 3 being duplicates leaving 26 studies. Twelve studies were excluded from this review by title and abstract reading. The remaining 14 studies were read in their entirety and 8 were subsequently excluded (Appendix A) based on the inclusion and exclusion criteria in table 2. The final six studies selected for this review are displayed in table 3. Figure 1 displays the searching process. Further information on the included studies can be found in the mapping table in Appendix B.

Table 2

Inclusion and Exclusion Criteria

	Inclusion Criteria	Exclusion Criteria	Rationale
1 Intervention	Precision teaching	Not precision teaching	To investigate whether the changes in outcome are due to the precision teaching intervention
2 Study Design	Group based or single case experimental design	Not group based or single case experimental design	To be able to compare the effectiveness of the intervention using between or within participants
3 Language	Studies written in English	Studies not written in English	Due to translation services not available
4 Published date	Studies published since 2006	Studies published before 2006	This is when the Rose Report was published and schools began teaching reading using a synthetic

	Inclusion Criteria	Exclusion Criteria	Rationale
			phonics approach. Past reviews have only considered the effectiveness of PT prior to this date
5 Type of Publication	Studies from Peer reviewed journals	Studies not from peer reviewed journals	Peer reviewed journals ensure higher quality research than non-peer reviewed journals due to screening for validity and quality control prior to publication
6 Participants	School aged pupils	Not school aged pupils	The purpose is to investigate the effectiveness of the intervention on school aged pupils (minimum age 5, maximum age 16)

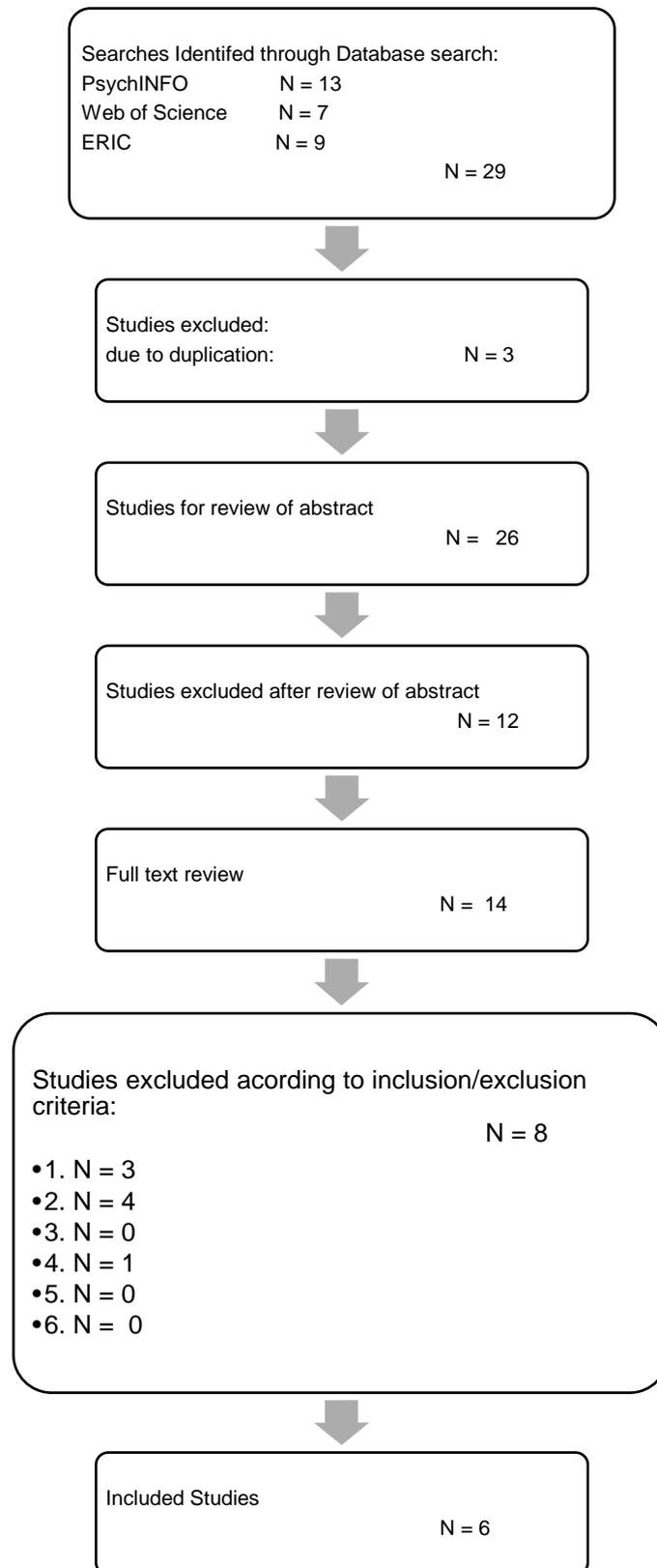
Table 3

List of Included Studies

Included Studies

1. Downer, C. (2007). The National Literacy Strategy Sight Recognition Programme Implemented by Teaching Assistants: A precision teaching approach. *Educational Psychology in Practice*, 23(2), 129-143.
2. Roberts, W., & Norwich, B. (2010). Using precision teaching to enhance the word reading skills and academic self-concept of secondary students: a role for professional educational psychologists. *Educational Psychology in Practice*, 26(3), 279-298.
3. Griffin, C. P., & Murtagh, L. (2015). Increasing the sight vocabulary and reading fluency of children requiring reading support: the use of a Precision Teaching approach. *Educational Psychology in Practice*, 31(2), 186-209.
4. Lambe, D., Murphy, C., & Kelly, M. E. (2015). The impact of a precision teaching intervention on the reading fluency of typically developing children. *Behavioral Interventions*, 30, 364-377.
5. Hughes, J. C., Beverley, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis*, 8, 221-238.
6. Mannion, L., & Griffin, C. (2018). Precision teaching through Irish: effects on isolated sight word reading fluency and contextualised reading fluency. *Irish Educational Studies*, 37(3), 391-410.

Figure 1: *Flowchart of the searching and screening process*



Weight of Evidence

This review used the framework by Gough (2007) to appraise the six included studies. Gough's framework includes 3 dimensions. The first dimension is weight of evidence A (WoE A) which evaluates the quality of the study according to an agreed criteria. This review uses Gersten, Compton, Coyne, Greenwood and Innocenti's (2005) quality indicators for the quasi-experimental design studies. Horner, Carr, Halle, Mcgee, Odom and Wolery's (2005) criteria is used to review the WoE A for the single case experimental design studies. The Weight of Evidence B (WoE B) dimension assesses the relevance of the methodological design to the research question. Weight of Evidence C (WoE C) considers how relevant the focus of the study is to the review question. Weight of Evidence D (WoE D) takes an average of the first three criteria and gives an overall numbered rating. Further information about the coding protocols used and rationale for WoE B and C can be found in Appendix C and D. Table 4 displays the WoEs for all 4 areas.

Table 4

WoE A, B, C & D for all 6 Studies

Study	WoE A	WoE B	WoE C	WoE D
Downer (2007)	1 Low	2 Medium	1.8 Medium	1.6 Low
Roberts & Norwich (2010)	2 Medium	2 Medium	2.5 High	2.2 Medium
Griffin & Murtagh (2015)	2 Medium	2 Medium	2 Medium	2 Medium
Lambe, Murphy & Kelly (2015)	2.1 Medium	2 Medium	1.5 Low	1.9 Medium
Hughes, Beverley & Whitehead (2007)	1 Low	2 Medium	2 Medium	1.7 Medium
Mannion & Griffin (2018)	1 Low	2 Medium	1.5 Low	1.5 Low

Note. 0.9
-1.6 =

“low”, 1.7- 2.3 = “medium” and 2.4 - 3 = “high”

Participants

214 participants were included in the six included studies. The ages of participants ranged between 5 and 16 years of age. This demonstrates that the studies captured children across both primary and secondary settings. The studies included children from across England and Ireland suggesting that the intervention can be generalised to support reading in westernised countries. Studies by Downer (2007) and Roberts and Norwich (2010) both used over double the amount of male participants to female with Downer (2007) using 32

male participants and 15 female, while Roberts and Norwich (2010) used 53 males and 24 female participants. The studies by Griffin and Murtagh (2015) and Hughes, Beverley and Whitehead (2007) did not state the gender of participants included. Both Lambe, Murphy and Kelly (2015) and Mannion and Griffin (2018) used more female participants, however their studies included participants without reading difficulties. Consequently, this may limit the generalisability of the precision teaching intervention to females with reading difficulties. None of the studies characterised the sample chosen, therefore making it difficult to generalise the study to ethnicity or culture.

Four of the studies reviewed included participants with reading difficulties, these participants were identified by school staff (Downer, 2007; Roberts & Norwich, 2010; Griffin & Murtagh, 2015; Hughes et al., 2007). The study by Griffin and Murtagh (2015) ran standardised tests to capture the reading age of participants during selection. However they still scored low in the 'describing participants' criteria for WoE A as the participants were not comparable. Children without reading difficulties were selected for the control group.

Mannion and Griffin (2018) scored higher on the WoE A criteria for similar characteristics across samples due to randomly assigning participants to control or experimental groups, therefore enabling it to be assumed participants were comparable across conditions.

Research Design

Four of the six studies included in this review used a quasi-experimental design. The study by Downer (2007) does not include a control group therefore lowering the score in the WoE A criteria based on describing comparison conditions. The studies by Roberts and Norwich (2010) and Mannion and Griffin (2018) both included a control and experimental group and the participants were randomly assigned after selection resulting in higher scores for WoE A.

As discussed earlier Griffin and Murtagh (2015) did not have comparable conditions across their control and experimental groups therefore resulting in less superior scores for this WoE A dimension.

The final two studies included in the review were single case experimental designs. Horner et al., (2005) coding protocol was used to measure WoE A for these studies. The study by Lambe et al., (2015) did not include a control group unlike Hughes et al., (2007). However, the design used a multiple baseline effect and showed experimental effect over three time periods therefore scoring higher in WoE A.

As studies that include an evaluation of maintenance of effects of an intervention (Gersten et al., 2005) are considered better quality than those that do not, one study included was considered methodologically superior according to WoE A than the other five studies. This study was Roberts and Norwich (2010) who completed a follow up that was statistically analysed for one of the cohorts measured in the study.

Measures

All of the outcome measures across the studies were used at appropriate times (e.g. pre and post intervention). Only two studies used the same measures for testing reading fluency (Lambe et al., 2015; Hughes et al., 2007) who both used a set of words taken from the First 220 Dolch Sight words. The score was calculated from the words read per minute. Mannion and Griffin (2018) used a list of isolated Irish vocabulary, selected from the book *Fuaimeanna agus Focail 5*. The score was calculated from the words read per minute.

Downer (2007) used two different outcome measures depending on the age of the participant. For the infant and primary aged children word lists from a variety of sources were used and for secondary participants the Salford Reading Test. The Salford Reading Test is a standardised assessment providing a reading age but not specifically assessing reading fluency. Therefore this may be considered less superior when considering the outcomes of the study.

The study by Roberts and Norwich (2010) used a list of the 332 most commonly occurring sight words from the national literacy strategy which is not a standardised assessment, although gives a clear measurable pre and post intervention score. Reading fluency was measured by Griffin and Murtagh (2015) using wordlists from SNIP PT-Pack.

Some studies specify that children were given one minute to read as many words correctly as they could. Other studies are not specific about how the measures were applied to the participants (Griffin & Murtagh, 2015; Downer, 2007; Roberts & Norwich, 2010). This prevents replicability of these studies

thus suggesting they are less methodologically superior. None of the studies provide evidence of the construct validity or the criterion-related validity of the measures used, this may be one of the reasons none of the studies scored 'high' for WoE A.

Intervention

The interventions implemented ran from 2 weeks to 26 weeks. All were delivered daily apart from Lambe et al., (2015) and Hughes et al., (2007) who implemented PT 3-4 times a week. Lindsley (1990) stated one of the key components of PT is that the participant receives the intervention daily, thus it may be considered that Lambe et al., (2015) and Hughes et al., (2007) did not keep to the fidelity of the intervention. The interventions lasted between 4 minutes and 20 minutes. All interventions followed the procedure of teach-test-chart- review. All studies included the use of a time probe and SCC as recommended by Lindsley (1990).

All the studies described the intervention in detail, therefore resulting in a higher WoE A rating as the intervention can be easily replicated. However, as the study by Mannion and Griffin (2018) was delivered by pupils in pairs to each other this may result in straying from the foundation of PT suggested by Lindsley (1990) that teachers support in adaption of learning as it may be considered that the children would not know how to adapt the learning for their peer.

Delivery

The precision teaching intervention was delivered in most of the studies by a teaching assistant in a 1:1 situation as is the usual model of delivery. Downer (2007) describe how TAs were shown how to use the materials for each child and a proforma for the probe sheets was provided. An intensive PT training programme was undertaken for TAs in the study by Roberts and Norwich (2010) and ongoing support was offered by the researcher. In the study by Griffin and Murtagh (2015) it is stated that teachers were trained by the researcher and guidelines alongside an individual folder of resources provided for each pupil in the experimental group. The study by Hughes et al., (2007) states that a TA delivered the intervention but there is little information about the training provided for the TA. The researcher delivered the PT intervention in the study by Lambe et al., (2015). In contrast to all of the other studies the intervention in the study by Mannion and Griffin (2018) was performed by the participants in pairs. The children received training and were provided with materials before beginning the intervention. The children used a laptop computer which presented the daily PT probes. As this may be considered as having less fidelity to the intervention the WoE C criteria scored this method of delivery lower.

Roberts and Norwich (2010) developed the intervention for the second cohort study by providing further training to the teaching assistants on how to adapt the learning for the individual child. This resulted in significant improvements in reading for cohort 2.

Findings

All six studies stated that the precision teaching intervention increases reading fluency. Three studies reported effect sizes, thus resulting in higher WoE A scores for data analysis. Mannion and Griffin (2018) and Griffin and Murtagh (2015) both reported partial eta squared effect sizes, these have been converted to Cohen's *d* in order to allow easier comparison to other studies using the Campbell Collaboration calculator (table 5). Roberts and Norwich (2010) reported *r* effect sizes which have also been converted to Cohen's *d* using the Campbell Collaboration calculator. The effect sizes for the intervention compared with the control group were small for cohort 1 (groups A and B). For cohort 2 (after extra TA training) there was one large effect size for a group at follow up testing compared with their baseline score.

The effect size for the study by Downer (2007) was calculated as pre and post intervention mean scores and standard deviations were published. This revealed a large effect size suggesting the intervention was effective. The effect size for the Griffin and Murtagh (2015) study was large, however this study scored lower in WoE A criteria as the control group was not taken from a similar group therefore trust cannot be put into the effectiveness of the intervention compared to teaching as usual group.

For the purposes of this review the effect sizes descriptors were taken from Cohen's (1988) criteria; 0.2 means a small effect; 0.5 is a medium effect and 0.8 is a large effect.

The single case experimental design studies did not provide enough information for effect sizes to be calculated. The study by Lambe et al., (2015)

provided information by graph but this was not clear enough to calculate information. The data provided (table 6) shows each child made progress in the amount of words they could read within a minute from pre intervention testing to post intervention testing. In the case of Hughes et al., (2007) this was a larger increase in scores than the participants in the control group.

Although all six studies discussed state the intervention is effective at improving reading fluency this should be taken with caution due to the small sample sizes. Cohen (1992) stated that large effect sizes are significant with 23 participants, medium with 64 participants and small with 393 participants. If we take this into consideration then the studies included should be approached with caution. Two studies used exclusively children with reading difficulties in the control group and experimental group. In contrast the other studies used either children without reading difficulties (in both groups or just control) or did not use a control group at all. Consideration must also be given to attrition bias which may have affected the equivalence of the control and experimental group (Barker, Pistrang & Elliott, 2015) which can impact the results of a study.

Table 5

Effect Sizes for Group Design Studies

Study	Outcome Measure	Participant sample	Effect Size	Follow up effect size	Effect size Descriptor	Overall WoE D
Downer (2007)	Y1,2,3,4,5 & 6 Number of new words the child could instantly recognise	Y1 and Y2 N=13	d= 2.169		Large	1.6
		Y3,4,5,& 6 N= 19	d= 1.233		Large	Low
	Y7 & 8 The Salford Reading Test (1991).	Y7 and Y8 N=15	d= 1.378		Large	
Roberts & Norwich (2010)	Reading Accuracy Measure (RAM) - 332 Commonly occurring sight words from the National Literacy Strategy	N = 77	Cohort 1 d= 0.2828		Small	2.2 Medium
			Cohort 2 d= 0.4296	Cohort 2 only Group C	Small Large	

Study	Outcome Measure	Participant sample	Effect Size	Follow up effect size	Effect size Descriptor	Overall WoE D
	Lists as a criterion referenced assessment			d=-1.1242 Group D d=-0.42	Small	
Griffin & Murtagh (2015)	Wordlists from the SNIP-PT-Pack.	N = 40	d= 4.2259		Large	2 Medium
Mannion & Griffin (2018)	Reading from a list of isolated Irish vocabulary, selected from the book <i>Fuaimanna agus Focail 5</i> . Score in words per minute was calculated.	N = 36	d= 1.372		Large	1.5 Low

Table 6

Data from Single Case Experimental Design Studies

Author	Outcome Measure	Outcomes	WoE D
Lambe et al., (2015)	First 220 Dolch Sight words. Words read per minute tested. N = 7	<i>Pre intervention words per minute mean score</i> 46 <i>Post intervention words per minute mean score</i> 96 <i>Follow up words per minute mean score</i> 93	1.9 Medium
Hughes et al., (2007)	First 220 Dolch Sight words. Words read per minute tested. N = 7 (Control = 2)	<i>Pre intervention words per minute score</i> Range: 38-50 correctly read words <i>Post intervention words per minute score</i> Range: 70 – 130 correctly read words per minute	1.7 Medium

Conclusion and Recommendations

This review investigated the effectiveness of Precision Teaching at improving reading fluency in school age children. Each study reviewed displayed impressive improvements in reading fluency as a result of the implementation of a PT intervention.

Gough's (2007) weight of evidence framework was used to evaluate the studies identified through a literature search. Four of the studies were rated as 'medium' for the overall weight of evidence criteria (WoE D) and two studies were rated 'low.' This suggests further research needs to be performed with more adherence to methodological rigour and the focus relevant to the review question.

Of the four studies where effect sizes could be calculated three found large effect sizes (Mannion & Griffin, 2018; Griffin & Murtagh, 2015; Downer, 2007). The study with the highest WoE D score (Roberts & Norwich, 2010) found a mixture of small and large effect size for the outcomes tested.

One of the studies included measured the increase in reading fluency for learning a second language (Irish) which showed a large effect size for improvement (Mannion & Griffin, 2018). This generalises the concept that PT improves reading fluency for all learners not just those with reading difficulties.

All of the studies included information on the intervention and delivery, however the lengths of the intervention (both individual sessions and the intervention as a whole) differed considerably (2 weeks to 26 weeks).

The single case experimental design studies did not provide enough data to allow effective statistical analysis although the data provided did show improvements in reading fluency.

All of the studies that are part of this review took place in the UK and Ireland suggesting that the PT intervention could be generalised across English speaking countries. However, limitations of the included studies included the use of different outcome measures, differences in the length of the intervention and the lack of the use of follow up measures or control groups in most of the studies.

The current review only selected studies from 2006 onwards as this was when early reading changed due to the Rose Report (Rose, 2006) therefore there may be more rigorous research performed prior to this date providing an evidence base. In conclusion it is necessary for further research to be conducted on the use of PT to improve reading fluency.

Recommendations for future research would include; firstly ensuring studies include control groups consisting of comparable samples in order to effectively evaluate whether improvement was due specifically to the PT intervention.

There is also a necessity to include information about the ethnicity and culture of participants in order to be able to generalise findings across cultures and groups of participants. The hierarchy of learning (Haring et al., 1978) includes 'maintenance' as the third stage of acquiring a new skill, therefore the use of follow up measures to evaluate whether PT enable 'maintenance' would be an essential part of future research.

Increasing the sample sizes of participants would also carry more weight to the statistical outcomes of future studies. A final consideration would be the use of more rigorous methodological designs including randomised control trials which would help to increase the validity of research.

References

- Barker, C., Pistrang, N., & Elliott, R. (2015). *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners*. John Wiley: Chichester.
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, 19,163–197.
- Brosnan, J., Moeyaert, M., Brooks Newsome, K., Healy, O., Heyvaert, M., Onghena, P., & Van den Noortgate, W. (2018), Multilevel Analysis of Multiple-baseline Data Evaluating Precision Teaching as an Intervention for Improving Fluency in Foundational Reading Skills for at Risk Readers, *Exceptionality*, 26(3), 137-161.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nded.).Hillsdale, NJ: Erlbaum.
- Denig, S. J. (2004). Multiple intelligences and learning styles: two complementary dimensions. *Teachers College Record*, 106(1), 96–111.
- Department of Education. (2015). *Reading the Next Steps*: DfES Publications.
- Downer, C. (2007). The National Literacy Strategy Sight Recognition Programme Implemented by Teaching Assistants: A precision teaching approach. *Educational Psychology in Practice*, 23(2), 129-143.
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71(2), 149-164.
- Griffin, C. P., & Murtagh, L. (2015). Increasing the sight vocabulary and reading fluency of children requiring reading support: the use of a Precision Teaching approach. *Educational Psychology in Practice*, 31(2), 186-209.

- Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2), 37–41.
- Haring, N.G., Lovitt, T.C., Eaton, M.D., & Hansen, C.L. (1978). *The fourth R : Research in the classroom*. Columbus: Charles E. Merrill Publishing Co.
- Health and Care Professions Council. (2015). *Standards of Proficiency: Practitioner Psychologists*. London: Health and Care Professions Council.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education, *Exceptional Children*, 71(2), 165-179.
- Hughes, J. C., Beverley, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis*, 8, 221-238.
- Johnson, K., & Street, E. M. (2013). *Response to intervention and precision teaching: Creating synergy in the classroom*. New York: The Guilford Press.
- Kubina, Jr., R. M., & Yurich, K. L. (2012). *The precision teaching book*. Lemont, PA: Greatness Achieved Publishing Company.
- Lambe, D., Murphy, C., & Kelly, M. E. (2015). The impact of a precision teaching intervention on the reading fluency of typically developing children. *Behavioral Interventions*, 30, 364-377.
- Lindsley, O. R. (1964). Direct measurement and prosthesis of retarded behavior. *Journal of Education*, 147, 62-81.
- Lindsley, O. R. (1990). Precision teaching: By teachers for children. *Teaching Exceptional Children*, Spring, 10-15.
- Lindsley, O.R. (1991). Precision Teaching's Unique Legacy from B.F. Skinner. *Journal of Behavioral Education*, 1(2), 253-266.

- Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25*, 51–57.
- Mannion, L., & Griffin, C. (2018). Precision teaching through Irish: effects on isolated sight word reading fluency and contextualised reading fluency. *Irish Educational Studies, 37*(3), 391-410.
- Organisation for Economic Co-operation and Development. (2013). *England & Northern Ireland (UK): Country Note--Survey of Adult Skills First Results*. OECD Publications.
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal Epidemiol Community Health, 57*, 527-529
- Radford, J., Bosanquet, P., Webster, R., Blatchford, P., & Rubie-Davies, C. (2014). Fostering learner independence through heuristic scaffolding: A valuable role for teaching assistants. *Journal of Educational Research, 63*(C), 116-126.
- Roberts, W., & Norwich, B. (2010). Using precision teaching to enhance the word reading skills and academic self-concept of secondary students: a role for professional educational psychologists. *Educational Psychology in Practice, 26*(3), 279-298.
- Rose, J. (2006). *Independent Review of the Teaching of Early Reading*. Nottingham: DfES Publications.
- Skinner, B. F. (1977). The experimental analysis of operant behaviour. *Annals of the New York Academy of Sciences, 291*(1), 374-385.
- Snowling, M. J., & Hulme, C. (2010). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology, 81*, 1–23.
- West, R. P., Young, K. R., & Spooner, F. (1990). Precision teaching, an introduction. *Teaching Exceptional Children, Spring*, 4-9.

Appendices

Appendix A

Excluded Studies from Systematic Review with Reasons

Reference	Reason (exclusion criteria)
Bennett, K., Brooks Newsome, K., & Newsome, D. (2015). Assessing Functional Relations: The Utility of the Standard Celeration Chart. <i>Behavioral Development Bulletin</i> , 20(2), 163-167.	1
Brosnan, J., Moeyaert, M., Brooks Newsome, K., Healy, O., Heyvaert, M., k Onghena, P., & Noortgate, W. (2018). Multilevel Analysis of Multiple-baseline Data Evaluating Precision Teaching as an Intervention for Improving Fluency in Foundational Reading Skills for at Risk Readers, <i>Exceptionality</i> , 26(3), 137-161.	2 and 6
Kubina, R. M. (2005). Developing reading fluency through a systematic practice procedure. <i>Reading & Writing Quarterly</i> , 21, 185-192.	2 and 4
Fitzpatrick, E., McLaughlin, T. F., & Weber, K. P. (2004). The effects of a first day and second day reads on reading accuracy with reading mastery III textbook B for a fifth grade student with learning disabilities. <i>International Journal of Special Education</i> , 19(1), 56-63.	4
Hixson, M. D., Jacob, S., & Morgan, S. (2007). A Comparison of Error Correction Procedures on Word Reading. <i>Journal of precision teaching and celebration</i> , 23, 2-13.	1
Lamport Commons, M., & Goodman, S. A. (2008). Project Giant Step: Ocean Hill -Brownsville and the New York City School Strike. <i>Journal of precision teaching and celebration</i> , 24, 38-42.	2
Selfridge, K. A., & Kostewicz. (2011). Reading Interventions for Four Students With Learning Disabilities. <i>Journal of precision teaching and celebration</i> , 27, 19-24.	2
Hayes, B., Heather, A., Jones, D., & Clarke, C. (2018). Overcoming barriers to using precision teaching with a web-based programme, <i>Educational Psychology in Practice</i> , 34(2), 166-174.	1

Appendix B

Mapping the Field

No	Author	Country	Participants	Study Design	Intervention Details	Measures	Primary Outcomes
1	Downer, (2007)	England	37 participants were selected across seven schools. 1 infant (N=13) 4 junior (N =19) 2 secondary (N =15) Boys outnumbered girls. Participants were selected by class teachers due to their slow progress in reading in comparison to age related peers.	Quasi-experimental design	The intervention lasted 2-26 weeks depending on the needs of the child. 1:1 4 minutes per day following same routine 1) Teach/revise the 5 target words 2) Use probe to test words read per minute. 3) Chart number of words for that day on SCC. 4) Record together on graph.	Infant and junior participants- the number of new words the pupil could read. Secondary- The Salford Reading Test measured changed in reading age. Qualitative data was also gathered on the teaching assistant's views on the intervention.	Infant- significant increase in reading score Junior- Overall significant increase in reading fluency. However the results were fluctuating according to year group. Secondary Overall significant increase in reading age.

No	Author	Country	Participants	Study Design	Intervention Details	Measures	Primary Outcomes
2	Roberts & Norwich (2010)	England	Secondary students across 5 secondary schools N = 77. 3 mainstream and 2 specialist schools. Ages 11:4 to 16:2. 53 boys and 24 girls. Participants selected by school Senco's as those in need of additional support for reading. Participants randomly allocated to either intervention group or delayed control.	Quasi-experimental Randomized allocation. Study run twice (Cohort 1 and 2)	Cohort 1 (Groups A and B) lasted 6 weeks and then swapped. Then cohort 2 (groups C and D) repeated the study over 5 week period. Daily intervention following teach-test-chart-review structure	Word reading skills measured using the 332 commonly occurring sight words from NLS. Participants tested before and after intervention. Each participant completed Burden's Myself as a Learner scale test (pre-mid and post measures taken.)	In cohort 1 no statistical difference in scores. Both made continuing progress in their word reading skills. Cohort 2: after intervention statistically increased score compared to control. Follow up assessments (19 and 13 weeks following) indicate both C and D continued to progress.
3	Griffin & Murtagh (2015)	Ireland	40 primary school children across 3 primary schools (7:4 years to 12:2)	Quasi-experimental design	10 minutes 1:1 daily for 8 weeks	Pupils level of sight vocabulary (wordlists from SNIP PT-Pack)	Statistically significant difference in reading fluency

No	Author	Country	Participants	Study Design	Intervention Details	Measures	Primary Outcomes
			Experimental group- 20 children already attending learning support for reading. Control: pupils not receiving any specialist reading support at that time.		Teach-test-chart-review structure	Gray Oral Reading Test measured overall reading ability score. Qualitative data also gathered from teachers and experimental group.	and reading comprehension for pre to post test Statistically significant increase in overall reading ability from pre to post test.
4	Lambe et al. (2015)	Ireland	7 participants from a primary school. (7:2 to 8:4 years) No participants identified as having any reading difficulties. 4 girls and 3 boys.	Single case experimental design	Over 6 weeks (20 minutes 3 days a week) 2 phases 1) Say all fast a minute every day shuffled 2) Dolch story	Baseline SAFMEDS cards read-score kept	Reading fluency increased.
5	Hughes et al. (2007)	Wales	7 pupils (11-12 years) from a secondary school. 5 were randomly assigned to the PT	Single case experimental design	20 minute 1:1 session 3-4 times a week for 10 Weeks.	Group Reading test 2 (GRT 2) and the Vocabulary Scale of the Middle Years	All PT children increased their reading word frequency significantly

No	Author	Country	Participants	Study Design	Intervention Details	Measures	Primary Outcomes
6	Mannion & Griffin (2018)	Ireland	<p>36 primary school students. Participants randomly assigned to experimental and control N= 18 per group</p> <p>14 males and 22 females.10-12 years. All participants were required to be learning Irish as a second language in school. No participants came from a background where Irish was spoken at home.</p>	Quasi-experimental design	<p>TAU group were supported by a TA for the same amount of time.</p> <p>Children in pairs 10 minutes per day implemented by teacher and using a fluency probe on Excel. Children recorded on probe themselves. Over 3 weeks.</p> <p>Control engaged in typical teaching of Irish with their class teacher.</p>	<p>Information System (MidYIS) year 7.</p> <p>1) Read list of isolated Irish vocabulary 2) Read a piece of contextualised Irish text taken from Sin E 6.</p>	<p>TAU children showed little effect on their reading frequency. Significant difference between the experimental and control groups for increase in scores for isolated sight word reading fluency and contextualised reading fluency in post intervention test scores.</p>

Appendix C - Weight of Evidence

Weight of Evidence A – Methodological quality of the research

Two different coding protocols were used to review the studies within this review. The four quasi experimental studies were evaluated using the Gersten et al. (2005) protocol in full. The total score was calculated by adding together the scores from the essential criteria (more than 9 = 1, less than 9 = 0) and desirable criteria (4 or more = 2, less than 4 = 1, 0 = 0). For example the study by Downer (2007) had 4 essential criteria scoring 0 and 3 desirable criteria scoring 1, resulting in an overall WoE A score of 1. Table C 1 displays the ratings found after reviewing.

Table C 1

Judgement Ratings for Quasi-Experimental Design Studies

Study	Essential Criteria	Desirable Criteria	WoE A
Downer (2007)	4	3	1 (low)
Roberts & Norwich (2010)	5	4	2 (medium)
Griffin & Murtagh (2015)	8	4	2 (medium)
Mannion & Griffin (2018)	7	2	1 (low)

The coding protocol from Horner et al. (2005) was used to review the two single case experimental design studies. Both of these papers indicate quality ratings for different areas of methodological design used in a study. Table C3 displays the ratings found after reviewing.

Table C 2

Judgement Ratings for Single Case Experimental Designs

Study	WoE A	Rationale
Lambe et al. (2015)	2 (medium)	Based on judgment criteria from Horner et al. (2005).
Hughes et al. (2007)	1 (low)	

Weight of Evidence B: Appropriateness of Design

The weight of evidence B was evaluated using the hierarchy from Petticrew and Roberts (2003) who stated that in order to answer a question of ‘effectiveness’ the studies most appropriate to use should follow this system of weighting. Table C 3 displays WoE B criteria: (systematic literature reviews have been removed from the list). Table C 4 displays the judgments made for the studies included in this review.

Table C 3

WoE B Criteria

Weighting	Description
High (3 points)	1. Randomised control trials
Medium (2 points)	2. Cohort studies, quasi-experimental studies, single case experimental designs
Low (1 point)	3. Qualitative research, survey, case control, non-experimental evaluation

Table C 4

WoE B Judgments

Study	Weight of Evidence B
Downer, (2007)	2
Roberts & Norwich (2010)	2
Griffin & Murtagh (2015)	2
Lambe et al. (2015)	2
Hughes et al. (2007)	2
Mannion & Griffin (2018)	2

Weight of Evidence C – Topic relevance

The weight of evidence C questions were developed to review how appropriate the design of the study was to the review question. The scores have been found by finding the average of the judgements from criteria A to D.

Table C 5 displays the criteria used to evaluate the studies. Table C 6 displays the ratings for each study.

Table C 5

WoE C Criteria

Criteria	Ratings	Rationale
A Setting and Interventionist	3. Intervention is carried out outside classroom in 1:1 condition with teacher/TA. 2. Intervention is carried out inside classroom setting with teacher/TA. 1. Intervention is carried out within the classroom in pairs with computer support.	To see if the intervention is replicable within a mainstream school.

Criteria	Ratings	Rationale
B Identification of Participants	<p>3. Participants are screened explicitly before being selected for intervention</p> <p>2. Participants are chosen by school staff based on their appropriateness for the intervention.</p> <p>1. There is no clear rationale for selection of participants.</p>	To see whether the intervention has an effect on participants dependent on their level of difficulties.
C Measurement of reading fluency improvement	<p>3. The effects of the intervention on reading fluency are present at follow up assessment.</p> <p>2. The effects of the intervention are not evident at follow up assessment.</p> <p>1. No follow up assessment is conducted.</p>	To see if the intervention has results that are long lasting with the participants.
D Use of control Group	<p>3. Control group is used and alternative provision is described.</p> <p>2. Control group is used but no explanation of alternative provision/ lack of is described.</p> <p>1. No control group is used.</p>	To understand whether the intervention has had an effect on the children's reading that would not have happened anyway.

Table C 6

Scores for WoE C

Study	Criteria A	Criteria B	Criteria C	Criteria D	WoE C
Downer (2007)	3	2	1	1	1.8
Roberts & Norwich (2010)	3	2	3	2	2.5
Griffin & Murtagh (2015)	3	2	1	2	2
Lambe, Murphy & Kelly (2015)	3	1	1	1	1.5
Hughes, Beverley & Whitehead (2007)	2	2	1	3	2
Mannion & Griffin (2018)	1	2	1	2	1.5

Appendix D – Weight of Evidence A Coding Protocol

Coding protocol: Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C, & Innocenti, M. (2004). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71,149-164.

Study: Downer, C. (2007). The National Literacy Strategy Sight Recognition Programme Implemented by Teaching Assistants: A precision teaching approach. *Educational Psychology in Practice*, 23(2), 129.

Essential and Desirable Quality Indicators for Group Experimental and Quasi-Experimental Research Articles and Reports

Essential Quality Indicators - Quality indicators for describing participants

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

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

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

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

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

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

Essential Quality Indicators - Quality indicators for Implementation of the Intervention and Description of Comparison Conditions

Was the intervention clearly described and specified?

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

Was the fidelity of implementation described and assessed?

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

Was the nature of services provided in comparison conditions described?

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

Essential Quality Indicators – Quality Indicators for Outcome Measures

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

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

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

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

Essential Quality Indicators – Quality Indicators for Data Analysis

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

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

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

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

Essential Quality Indicators Total Score: 4

Desirable Quality Indicators

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Were results presented in a clear, coherent fashion?

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

Desirable Quality Indicators Total Score: 3

	Total	Score
Essential Quality Indicators Total of >9 = Score 1 Total of <9 = Score 0	4	0
Desirable Quality Indicators Total of ≥4 = Score 2 Total of <4 = Score 1 Total 0 = Score 0	3	1
Total Score (3 = High Quality; 2 = Acceptable Quality; <2 = Poor Quality)		1

Study Rating: Poor Quality

Coding Protocol: Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M.(2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2), 165-179

Study: Lambe, D., Murphy, C., & Kelly, M. E. (2015). The impact of a precision teaching intervention on the reading fluency of typically developing children. *Behavioral Interventions*, 30, 364-377.

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
 - N/A
 - Unknown/Unable to Code

- 2) The process for selecting participants is described with replicable 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 Weighting 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
 - No
 - N/A
 - Unknown/Unable to Code

- 3) Measurement of the dependent variable is valid and described with replicable precision.
 - Yes
 - 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 Weighting of Evidence: 3

Section C: Independent Variable

- 1) 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 Weighting of Evidence: 2

Section D: Baseline

- 1) 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 Weighting of Evidence: 2

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 Weighting of Evidence: 2

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 Weighting of Evidence: 3

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

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 Weighting of Evidence: 2

Ratings of Evidence

A: 1

B: 3

C: 2

D: 2

E: 2

F: 3

G: 2

Average WoE A = 15/7

Average WoE A= 2.1

Average WoE A = Sum of X / NX = individual quality rating for each judgement area

N = number of judgement areas

