Case Study 1: An Evidence-Based Practice Review Report

**How effective are one-to-one therapy dog interventions on reading ability in primary school children?**

1. **Summary**

Primary-aged children’s reading ability is a key concern for government and current policy (Department for Education, 2015) with a large number of these children not meeting the expected levels of reading (Department for Education, 2019). Therapy dogs are currently growing in popularity in many countries, including the UK, with many schools already having a therapy dog. There are a number of key differences in the relational aspects of interventions where the primary recipient for the child is an animal rather than a human. Therapy dog interventions in schools can be different in nature, in this case the interventions involve a child reading to a dog without judgement or correction from another person, and in this review this was on a one-to-one basis. This systematic literature review examined five studies which met the inclusion criteria and these were critically appraised for their methodological rigour and relevance. Effect sizes were examined to answer the review question.

There were mixed findings with a number of large, medium and small effect sizes, as well as some studies with no effects observed. There were a number of methodological flaws across the studies, in addition to great variability in the study designs and few participants overall per study. To summarise, caution should be given to concluding that there was an overall positive impact of therapy dog interventions on reading ability in primary schools, and future research is needed to support or refute this notion.
2. Introduction

2.1 What is a therapy dog?

A therapy dog is a trained dog used alongside a trained adult to support a person psychologically, socially, emotionally or physiologically (Friesen, 2010). Therapy dogs have been used with both adults and children in a range of settings such as hospitals and schools (Geist, 2011). Animal-assisted therapy (AAT) as it is known, has been used for many decades, with Levinson being one of the first to coincidentally notice the positive impact of his dog on a number of children’s behaviour (Reichert, 1998).

“The distinguishing features of Animal-Assisted Therapy (AAT) are characterized by the supplemental inclusion of a trained therapy dog in reaching an intervention goal in therapeutic environments, and as a supplement to an educational objective in school contexts” (Friesen, 2010, pg. 261).

Evidence supports the idea that children find animals comforting (Parshall, 2003), and Geist (2011) suggested that therapy dogs provide an emotional aid to humans during a range of interventions. Therapy dogs have been reported to have a positive effect on physiological stress and anxiety across a range of situations (Friedmann, Thomas & Eddy, 2000). For example, reading to dogs has been shown to reduce children’s stress as measured by heart rate and blood pressure (Friedmann, Katcher, Thomas, Lynch & Messent, 1983). Whilst AAT have been reported with a wide range of animals, research suggests that therapy dogs are convenient and cost effective (Friesen, 2010).

Reading to a therapy dog has been documented mostly in schools and libraries where children read to these dogs either on a one-to-one or group basis (Hughes, 2002; Le Roux, Swartz & Swart, 2014). With a growing interest in the intervention, programmes can vary enormously in terms of the structure of the sessions, the session duration
and the number of weeks for the intervention. For example, some sessions allow the child or children to pet the dog, whilst others do not allow any touching of the dogs. Similarly, some interventions do not occur on a regular basis or for a set amount of time, for example this is more likely to be the case in a library group where different people attend on different occasions and it may be less controlled. Another area of difference is that interventions can be used with whole classes, small groups or individuals (Connell, Tepper, Landry & Bennett, 2019; Kirnan, Siminerio & Wong, 2016).

2.2 Psychological Theory

To understand the effectiveness of therapy dogs, principles of social and emotional psychology are fundamental. Many authors describe the relationship between the dog and child during the reading process as unique when compared to a child’s relationship with another human (peer or adult) when reading (Friesen, 2010). When reading to a person there are social expectations and judgements made by the child and the adult, as well as the perceived expectations the child thinks adult has (Wohlfarth, Mutschler, Beetz & Schleider, 2014). It is believed that these automatic thoughts that a child processes about the reading interaction that is forthcoming, have an impact on how the child performs during reading, indicated through stress and other physiological measures (Schretzmayer, Kotrschal & Beetz, 2017). Furthermore, whilst a child reads, they are in continual fear of the judgements made by adults and their peers (Wohlfarth et al., 2014). For some children who do not find reading difficult this thought process may be very different and not problematic. However, for children who have additional needs, find reading difficult and struggle with their confidence and self-esteem this reading interaction is not a positive experience. To add to this, Wohlfarth et al. (2014) describe self-worth theory where they state that a child is likely to stop
trying if they expect negative feedback from another person. “Most frequently such negative feedback occurs after an experience of failure. Failure threatens self-estimates of ability and creates uncertainty about an individual’s capability to perform well thereafter. If the subsequent performance turns out to be poor, then doubts concerning one’s own ability are confirmed” (Wohlfarth et al., 2014, pg. 62). A key component of interventions with therapy dogs is that there is no feedback provided to the child whilst they are reading, including any corrections on errors made. This makes the reading experience less critical, with fewer social confinements and overall a more positive experience for the child.

This evidence suggested that providing children with the opportunity to participate in interventions with therapy dogs may give them the opportunity to practise in an environment where they have less anxiety and fear of judgement and negative feedback from humans. As a result, therapy dog interventions may facilitate better reading abilities in children.

2.3 Rationale

Reading is considered a fundamental part of development for children (Department of Children, Schools & Families, 2003). Specifically, reading is a key communicative tool that is central to social, emotional and cognitive abilities (Rose Report, 2006). A report by Save the Children (2015) highlights that in children aged eleven on free school meals, only four out of ten are competent at reading. In the UK almost 30% of children are not achieving the expected levels of reading at Key Stage 2 (Department for Education, 2019). School success is known to impact upon later life outcomes (Johnson & Kossykh, 2008), specifically children who fall behind at school are more likely to experience negative life outcomes, for example relating to health, social welfare and income, as well as engage in crime (Allen, 2011).
As stated in a number of government documents, such as the SEND Code of Practice (DfE & DoH, 2015) and the Every Child Matters agenda (Department of Children, Schools & Families, 2003) ensuring the best outcomes for children and young people is imperative. Educational Psychologists have the opportunity with direct working with children, schools and their families to promote better outcomes for reading through intervention programmes. Additionally, Educational Psychologists in the UK are often involved with suggesting evidence-based reading interventions at the request of schools.

There has been popularity and an increase in the number of dogs that are being used in schools in the UK, as well as across other countries (Jalongo, Astorino & Bombo, 2004). For example, in the UK there are programmes supported by the Bark and Read Foundation (Bark & Read, n.d) where volunteers take dogs into school. Furthermore, the former UK Education Secretary Damian Hinds suggested that he believed dogs in schools were having a positive impact and encouraged more schools having a dog for intervention purposes. With this in mind, a better understanding of what the evidence is to support the use of therapy dog interventions on reading is needed.

2.4 Review Question

How effective are one-to-one therapy dog interventions on reading ability in primary school children?
3. Critical Review of the Evidence

3.1 Literature Search

A systematic literature search was conducted in December 2019 for electronic journals using three databases, PsycInfo, ERIC and Web of Science. PsycInfo was selected for its focus on psychological research, ERIC for its educational focus and Web of Science for a broader range of literature. In addition an ancestral search was performed but no further results were found.

The following search terms were used:

**Intervention Type:** (dog therapy OR therapy dog OR puppy therapy OR therapy puppy OR therapy canine OR canine therapy OR school dog OR animal-assisted OR animal-assisted reading OR canine-assisted OR canine-assisted reading OR classroom dog OR classroom puppy OR classroom canine)

AND

**Population:** (primary aged OR primary OR primary-aged OR key stage 1 OR key stage 2 OR key stage one OR key stage two OR elementary OR four year old OR 4 year old OR five year old OR 5 year old OR six year old OR 6 year old OR seven year old OR 7 year old OR eight year old OR 8 year old OR nine year old OR 9 year old OR ten year old OR 10 year old or eleven year old OR 11 year old OR year 1 OR year one OR year 2 OR year two OR year 3 or year three OR year 4 OR year four OR year 5 OR year five OR year 6 OR year six OR kindergarten OR first grade OR second grade OR third grade OR fourth grade OR fifth grade OR sixth grade)

AND

**Outcome Measure:** (reading OR literacy)
3.2. Inclusion and Exclusion Criteria

The search returned 60 results across the three databases. Screening for duplicates removed nine studies, followed by title and abstract screening against the inclusion and exclusion criteria leaving 11 studies. These 11 studies were screened at full-text level allowing further details to be considered against the inclusion and exclusion criteria and this process resulted in five studies remaining (see Figure 1). A detailed description of the inclusion and exclusion criteria for this can be seen in Table 1. For a full list of the studies included in this review see Table 2. A list of the studies excluded at full text screening can be seen in Appendix A. For a table describing the key features of each study see the Mapping the Field table in Appendix B.
Table 1

List of Inclusion and Exclusion Descriptors by Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Inclusion</th>
<th>Exclusion</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Type of publication</td>
<td>Published in a peer-reviewed journal</td>
<td>Not published in peer-reviewed journal</td>
<td>To ensure methodological quality and rigour of the research</td>
</tr>
<tr>
<td>2. Language</td>
<td>Published or translated into English</td>
<td>Not published or translated into English</td>
<td>To ensure researcher understanding</td>
</tr>
<tr>
<td>3. Intervention</td>
<td>a) A therapy dog reading intervention</td>
<td>a) Interventions without a therapy dog reading component</td>
<td>The review focuses on the impact of a therapy dog reading intervention on a one-to-one basis</td>
</tr>
<tr>
<td></td>
<td>b) Delivered on a one-to-one basis</td>
<td>b) A therapy dog reading intervention in groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c) Intervention was more than one session</td>
<td>c) No intervention, therapy dog reading to assess casual impact (one session)</td>
<td></td>
</tr>
<tr>
<td>4. Participants</td>
<td>Attending a primary school</td>
<td>Not attending a primary school</td>
<td>The review focuses on primary children due to children learning to</td>
</tr>
<tr>
<td>Criteria</td>
<td>Inclusion</td>
<td>Exclusion</td>
<td>Rationale</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>a) Quasi-experimental</td>
<td>a) Non-experimental design</td>
<td>read and their reading developing fluency at this age</td>
</tr>
<tr>
<td>5. Design</td>
<td>b) Randomised Control Trial</td>
<td>b) Secondary data</td>
<td>The review aims to assess casual impact</td>
</tr>
<tr>
<td></td>
<td>c) Experimental design (pre and post test data)</td>
<td>c) Qualitative data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d) Primary data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>e) Quantitative data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Date of publication</td>
<td>2006 – 2019</td>
<td>Pre 2006</td>
<td>Considers literature since the implementation of the phonics review</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Rose Report, 2006) as a strategy for teaching reading</td>
</tr>
</tbody>
</table>
Figure 1

Flow Diagram of Selection Process

Articles identified
PsycInfo ($n = 16$)
ERIC ($n = 14$)
Web of Science ($n = 30$)
Total ($n = 60$)

Duplicates removed
$n = 9$

Articles screened by title and abstract
$n = 51$

Articles screened by title and abstract
Did not meet inclusion criteria ($n = 40$)

Articles screened by full text
$n = 11$

Articles screened by full text
Did not meet inclusion criteria ($n = 6$)

Articles identified for review
$n = 5$
### Table 2

**Full List of Studies in this Review**

<table>
<thead>
<tr>
<th></th>
<th>Full Study Reference</th>
<th></th>
</tr>
</thead>
</table>
3.3 Weight of Evidence

Using the Gough (2007) Weight of Evidence (WoE) Framework the five studies in this review were evaluated for their methodological quality and relevance, and moreover to consider the extent to which they answer the review question. Harden and Gough (2012) describe three key areas of weighing the quality and relevance of studies; WoE A methodological quality, WoE B methodological relevance, WoE C topic relevance to the review question. These WoE contribute to an overall WoE D (see Table 3 below).

WoE A was assessed using the Gersten, Fuchs, Compton, Coyne, Greenwood & Innocenti (2005) coding protocol, designed for use with group experimental studies (see Appendix C). This protocol was adapted as appropriate to critically appraise the research within this review (see Appendix F for details). For WoE B the Petticrew and Roberts (2003) typology was employed due to the nature of the review question (see Appendix D), and for WoE C a user-defined criteria was produced (see Appendix E) to examine the topic relevance to the review question.
Table 3

Weight of Evidence for Studies

<table>
<thead>
<tr>
<th>Studies</th>
<th>WoE A</th>
<th>WoE B</th>
<th>WoE C</th>
<th>WoE D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenihan et al. (2016)</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Medium)</td>
</tr>
<tr>
<td>Linder et al. (2018)</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Medium)</td>
</tr>
<tr>
<td>Levinson et al. (2017)</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(High)</td>
</tr>
<tr>
<td>Uccheddu et al. (2019)</td>
<td>2</td>
<td>3</td>
<td>2.25</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Medium)</td>
</tr>
<tr>
<td>Le Roux et al. (2014)</td>
<td>3</td>
<td>3</td>
<td>2.5</td>
<td>2.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(High)</td>
</tr>
</tbody>
</table>

Low 1.4 or less, Medium 1.5-2.4, High 2.5 or above

Participants

A total of 199 participants were recruited into the five studies in this review, the studies ranged from nine participants in the smallest study (Uccheddu, Albertini, Pierantoni, Fantino & Pirrone, 2019) to 102 in the largest (Le Roux et al., 2014). These participants ranged in age from six to thirteen years of age, with three studies having a wide range of ages (Le Roux et al., 2014; Levinson, Vogt, Barker, Jalongo & Van Zandt, 2017; Uccheddu et al., 2019), and two studies solely looking at one year group (Lenihan, McCobb, Diurba, Linder & Freeman, 2016; Linder, Mueller, Gibbs, Alper & Freeman, 2018). These studies characterise a good age range found within primary schools and as a result provide a representative sample.
When exploring WoE C, one important criteria for consideration was the location of participants, with the United Kingdom (UK) having the most relevance to the review, followed by OECD countries and finally non-OECD countries. The studies ranged in location from the United States of America (USA), Italy and South Africa. Le Roux et al. (2014) scored lowest for this criterion in WoE C due to being in South Africa, which is the least likely to be comparable to the UK education system. The remaining four studies were from OECD countries (Lenihan et al., 2016; Levinson et al., 2017; Linder et al., 2018; Uccheddu et al., 2019) which are likely to be a closer match to the UK education system due to their social and economic similarities, and therefore they are more likely to be representative of schools in the UK. None of the studies scored the highest rating for location in WoE C, with none of the studies conducted in the UK.

**Design**

All five studies were categorised as randomised control trials (RCTs), resulting in receiving the highest rating for WoE B. RCTs are considered to be the highest quality methodological design, seen as ‘gold standard’ in relation to efficacy research (Petticrew & Roberts, 2003), and RCTs were also the most appropriate for the review question. Despite the random allocation of participants to groups in RCTs, these can vary tremendously in the procedure of the experimental and control groups. For example, the control group serves as a comparison to the experimental condition however if the control do not partake in an additional activity or treatment to usual (no treatment control), it is unclear if the change occurred due to the intervention itself or another factor such as the attention received. In contrast, by participants actively partaking in an alternative activity or treatment (treatment control) the effect of the experimental group (intervention) can be measured more accurately. The experimental groups were similar in content, however the control groups varied greatly
and this was reflected in the WoE C. Linder et al. (2018) scored the lowest for WoE C control group criteria due to the participants in the control group not receiving an alternative treatment but continued their normal school and home-based school activities- in this case reading at home (no treatment control). In comparison, a number of studies used a treatment control, and received a medium rating for WoE C control group criteria, where the participants read to an adult (Lenihan et al., 2016; Uccheddu et al., 2019) or a peer (Levinson et al., 2017). Le Roux et al. (2014) received the highest rating for WoE C control group criteria, having multiple control groups, both encompassing a no treatment control where participants continued with their normal school activities and two treatment controls with participants reading to an adult or teddy bear.

Whilst in most studies participants only took part in the experimental or control condition (Lenihan et al., 2016; Le Roux et al., 2014; Linder et al., 2018; Uccheddu et al., 2019), one of the studies in this review involved participants taking part in both conditions, using a repeated measures design (Levinson et al., 2017). This makes the effect of the intervention harder to interpret, due to half the participants having already experienced reading to a dog for a number of weeks before then taking part in the control condition where they read to a peer.

**Intervention**

The intervention consisted of the participants reading to a therapy dog, and as a result these did not differ in content due to the nature of the intervention. Where the interventions did vary was in the duration and number of sessions. Overall the total duration of the intervention was different across the studies, with the longest being 5 hours across 10 weeks (Uccheddu et al., 2019), followed by 3 hours reading (6 hours
total with the dog) over 6 weeks (Linder et al., 2018), and a number of the studies were approximately 2.5 hours over 5 weeks (Lenihan et al., 2016; Le Roux et al., 2014; Levinson et al., 2017). As seen above the number of weeks ranged from five to ten and the amount of time and the number of session in the intervention per week was different ranging from 15 minutes per session (Levinson et al., 2017; Le Roux et al. 2014) to 30 minutes per session (Lenihan et al., 2016; Linder et al., 2018; Uccheddu et al., 2019) as well as once (Le Roux et al. 2014; Lenihan et al., 2016; Linder et al., 2018; Uccheddu et al., 2019) or twice (Levinson et al., 2017) per week. Consequently, the intensity of the intervention varied greatly.

Measures

All studies measured reading accuracy, either with curriculum-based (Lenihan et al., 2016; Levinson et al., 2017; Linder et al., 2018) or standardised reading assessments (Le Roux et al., 2014; Uccheddu et al., 2019). Additional measures of reading ability were examined in the studies, including comprehension by some (Le Roux et al., 2014; Linder et al., 2018; Uccheddu et al., 2019) but not all studies, as a result the review focuses on reading accuracy.

Outcome Measures and Effect Sizes

Outcome measures are reported for reading accuracy for all five studies in the review. These were measured consistently despite the use of different instruments. Some additional measures of reading ability were reported for some studies however there was not consistent measures resulting in these not being reported as part of this review. The results presented focus on differences in control and intervention groups on reading ability to assess the impact of therapy dogs (see Table 4). The effect sizes
reported are Cohen’s d, and where effect sizes were not available they were calculated permitting that sufficient information was present to do this.

Lenihan et al. (2016) had a medium effect (as displayed in Table 4) despite a small sample size. The results indicate that there is a strong relationship between the variables, thus given a larger sample size the findings may have had a larger effect size. Le Roux et al. (2014) reported a small effect size, despite the impact of a larger sample size. Le Roux et al. (2014) had the largest sample in the studies in this review, and also received the highest overall rating for methodological rigour and relevance on WoE D. Research by Levinson et al. (2017) found a range of effect sizes across a range of year groups from small effects for the oldest children measured in 5th Grade, medium for both 3rd and 4th Grades and large for the youngest children in 2nd Grade. These samples per year group were comparatively some of the smallest in comparison to other studies in this review. This study demonstrates the impact of therapy dogs specifically with a range of ages, and interestingly the effects on reading ability are larger for the youngest children and the effect sizes get smaller the older the children become. This has implications when considering at what age this intervention would be most effective.

Uccheddu et al. (2019) reported no effect between the variables, however this was the smallest sample with just nine participants and therefore it is likely any real effects would not be detected with such small representation of the population. Similarly, Linder et al. (2018) found no effect of intervention (see Table 4). This study received the lowest overall WoE D, with a low rating overall on WoE A. There was a lack of clarity in the procedures in how the intervention was approached in terms of set up for reading and what the participants were told. In addition the control group
was a non-treatment control with little known about the allocated reading time the children had or what was done during this time.
Table 4

**Effect Sizes and Overall Quality Indicators**

<table>
<thead>
<tr>
<th>Study</th>
<th>Measures of Reading Accuracy</th>
<th>Number of Participants</th>
<th>Sample</th>
<th>Effect Size (Cohen’s d)</th>
<th>Descriptor</th>
<th>WoE D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenihan et al. (2016)</td>
<td>Curriculum-based measures</td>
<td>15</td>
<td>2nd Grade</td>
<td>.79</td>
<td>Medium</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Medium)</td>
<td></td>
</tr>
<tr>
<td>Linder et al. (2018)</td>
<td>DIEBELS (standard class assessment)</td>
<td>28</td>
<td>2nd Grade</td>
<td>-.11</td>
<td>No Effect</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Medium)</td>
<td></td>
</tr>
<tr>
<td>Levinson et al. (2017)</td>
<td>Curriculum-based measures</td>
<td>11</td>
<td>2nd Grade</td>
<td>.93</td>
<td>Large</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>3rd Grade</td>
<td>-.68</td>
<td>Medium</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(High)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>4th Grade</td>
<td>.59</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>5th Grade</td>
<td>.28</td>
<td>Small</td>
<td></td>
</tr>
<tr>
<td>Uccheddu et al. (2019)</td>
<td>Cornoldi Reading Test</td>
<td>9</td>
<td>6-11 years of age</td>
<td>.00</td>
<td>No Effect</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Medium)</td>
<td></td>
</tr>
<tr>
<td>Le Roux et al. (2014)</td>
<td>Neale Analysis of Reading Ability</td>
<td>102</td>
<td>7-13 years of age</td>
<td>.29</td>
<td>Small</td>
<td>2.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(High)</td>
<td></td>
</tr>
</tbody>
</table>

*Reading accuracy was the only common measure of reading ability across studies – the review question focuses on this outcome measure*

*Cohen’s d effect size: small = .20, medium = .50, large = .80*
4. Conclusions and Recommendations

This literature review explored to what extent one-to-one therapy dog interventions are effective at enhancing reading abilities of primary school children. The research findings for this intervention are mixed, with some large (Levinson et al., 2017), medium (Lenihan et al., 2016; Levinson et al., 2017), and small effect sizes (Le Roux et al., 2014) as well as no effects found (Linder et al. 2018; Uccheddu et al., 2019). Whilst there were a number of effects shown, due to the variation in intervention details, small sample sizes and methodological quality of each study it is difficult to make conclusive statements regarding the overall effectiveness. Larger samples would give more representative findings of a whole population.

Of the studies that found no effect, Uccheddu et al. (2019) had the smallest sample size overall and it is likely that the study was underpowered with such a small sample. Linder et al. (2018) also demonstrated no effect, however this study did not have a treatment control group. Instead the control group was described as continuing with everyday lessons and a comparison was made to the after school reading homework the children were set being similar to the additional reading time the intervention group experienced. Despite this, there was no understanding of the reading experience at home, or in fact if the controls felt more comfortable reading at home, or were also reading to a dog. This study received the lowest overall WoE D, with a low rating overall on WoE A. There was a lack of clarity in the procedures in how the intervention was approached in terms of the set up for reading and what the participants were told. This may have also led to less clarity for those running the intervention if these details had not been specified.
The research in this review involved the same intervention, however therapy dog interventions are not prescribed to the finer details. Consequently, while many reading programmes will be defined to run for a set number of weeks and session lengths, the specifics of therapy dog interventions are more indistinct. This has resulted in programmes that varied in the overall duration of intervention, with some studies having double the amount of intervention time than others, as well as for double the number of sessions and overall weeks of intervention. The intensity and the repetition of reading sessions may have had an unknown effect on reading ability. Similarly, most programmes were only five weeks with short durations of sessions. To consider the efficacy of such a programme these factors need to be considered in future research.

Despite post-test follow up in reading abilities, only one study (Le Roux et al., 2014) examined the longer term impact of the intervention on reading ability. With reading development being a gradual process it is possible that a lack of long-term follow up from the majority of the studies could have led to a missed effect of the intervention. Indeed there may have been an effect on reading but it would have taken longer to find observable measureable differences.

Levinson et al. (2017) was the only study to evaluate the impact of a therapy dog on reading abilities in different age ranges. This provided insightful evidence as to when it is most successful to intervene and when the effectiveness of these interventions become less so, although still effective overall. This has implications as to recognising when reading interventions may be most effective and should be considered by future researchers as well as with government initiatives.

Recommendations for future research would be to replicate the studies in this review, giving more standardised instructions to all participants, and using a substantially
larger sample size to ensure better generalisation of the findings. Interestingly, whilst a number of control groups were used, some of which were treatment controls, none of the comparison groups were for alternative programmes. It may be a consideration for future research to see how effective therapy dogs are compared to more accessible and less costly programmes to facilitate reading abilities. Equally, consideration of how long the sessions were and how many weeks would aid understanding of what would make a therapy dog intervention more effective.

At the present time, it would be difficult for Educational Psychologists to recommend the use of a therapy dog to facilitate reading abilities in primary school children. However, there is indicative evidence that there may be positive effects, and Educational Psychologists could be involved in conducting future research in the area. Despite this, therapy dogs are used for a range of purposes in schools, and the impact a therapy dog can have on other areas of development have not been considered in this review.
5. References


https://doi.org/10.1177/001440290507100202


### 6. Appendices

#### Appendix A: List of Excluded Studies

<table>
<thead>
<tr>
<th>Reference</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>Reason for Exclusion</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------</td>
</tr>
</tbody>
</table>
# Appendix B: Mapping the Field Table

## Table 6.2

A Summary of Key Features for Studies in Review

<table>
<thead>
<tr>
<th>No</th>
<th>Author</th>
<th>Participants</th>
<th>Country</th>
<th>Study Design</th>
<th>Intervention</th>
<th>Measures</th>
<th>Primary Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lenihan et al.</td>
<td>15 x 2\textsuperscript{nd} Grade (18 at the start)</td>
<td>USA</td>
<td>Independent samples. Random allocation to intervention and treatment control (dog and non-dog) groups</td>
<td>5 weeks once per week for 30 minutes</td>
<td>Curriculum-based measures</td>
<td>CBM reading (3 texts for 1 min) number of correct words per minute.</td>
</tr>
<tr>
<td>2</td>
<td>Linder et al.</td>
<td>28 x 2\textsuperscript{nd} Grade</td>
<td>USA</td>
<td>Independent samples. Random selection for those eligible. Random allocation to intervention and no treatment control (dog and non-dog) groups</td>
<td>6 weeks once per week for 1 hour (30 minutes reading)</td>
<td>DIEBELS (standard class assessment)</td>
<td>Number of words correct per minute, also looks at errors and (1 minute) comprehend a written text</td>
</tr>
<tr>
<td>No</td>
<td>Author</td>
<td>Participants</td>
<td>Country</td>
<td>Study Design</td>
<td>Intervention</td>
<td>Measures</td>
<td>Primary Outcomes</td>
</tr>
<tr>
<td>----</td>
<td>----------------------------</td>
<td>--------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>3</td>
<td>Levinson et al. (2017)</td>
<td>45 x 2nd-5th Grades USA</td>
<td>Repeated measures. Random stratified. Once selected randomly allocated (stratified on sex and grade) to intervention and treatment control (read to peer).</td>
<td>5 weeks (twice per week for 15 minutes)</td>
<td>Curriculum-based measures</td>
<td>Oral fluency (pre) mid (week 5) post (after week 10)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Uccheddu et al. (2019)</td>
<td>9 x 6-11 years (with Autism Italy)</td>
<td>Random allocation based on demographics and diagnosis severity. Intervention and treatment control</td>
<td>10 sessions (once per week for 30 minutes)</td>
<td>Cornoldi reading test, test of reading comprehension, metaphorological competence</td>
<td>Reading fluency (speed and accuracy), reading comprehension, phonological awareness</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Le Roux et al. (2014)</td>
<td>102 x 7-13 years (3rd grade South Africa)</td>
<td>3 intervention groups (dog, adult, teddy bear) and no treatment control.</td>
<td>10 sessions (once per week)</td>
<td>ESSI Reading and spelling test</td>
<td>Reading rate, accuracy and comprehension.</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Author</td>
<td>Participants</td>
<td>Country</td>
<td>Study Design</td>
<td>Intervention</td>
<td>Measures</td>
<td>Primary Outcomes</td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>--------------</td>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>week for 15-20 minutes</td>
<td>Measured 3 times (pre, post and delayed)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix C: WoE A

Using the adapted Gersten et al. (2005) coding protocol the following ratings were found for WoE A (see Table 6.3). This protocol was edited to account for a typical rather than atypical population. Please see Appendix F for the Gersten et al. (2005) coding protocol and for the full details of this adaptation.

Table 6.3

<table>
<thead>
<tr>
<th>Studies</th>
<th>Number of Essential Indicators</th>
<th>Number of Desirable Indicators</th>
<th>Overall WoE A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenihan et al. (2016)</td>
<td>8</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Linder et al. (2018)</td>
<td>8</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Levinson et al. (2017)</td>
<td>10</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Uccheddu et al. (2019)</td>
<td>9</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Le Roux et al. (2014)</td>
<td>10</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Note: Overall WOE A (3 = High Quality; 2 = Acceptable Quality; <2 = Poor Quality)
Appendix D: WoE B

Using the Petticrew and Roberts (2003) guidelines for quality and rigour of research evidence, WoE B was defined as seen in Table 6.4 and coded (see Table 6.5). In this review using Petticrew and Roberts (2003) was considered appropriate to define WoE B, with the most effective way of measuring impact being experimental design and using an RCT being the best method to do this.

Table 6.4
WoE B Coding Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>3 – High</th>
<th>2 – Medium</th>
<th>1 – Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Design</td>
<td>Randomised control trials</td>
<td>Quasi-experimental, single-case study design or cohort studies</td>
<td>Non-experimental, case study, qualitative or surveys</td>
</tr>
<tr>
<td>Studies</td>
<td>WoE B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lenihan et al. (2016)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linder et al. (2018)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levinson et al. (2017)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uccheddu et al. (2019)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Le Roux et al. (2014)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix E: WoE C

Criteria were defined to be appropriate for addressing the research question in this review (see Table 6.6). The studies in this review were then weighted for each of the criteria to produce an overall WoE C (Table 6.7).

Table 6.6

WoE C Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>3 – High</th>
<th>2 – Medium</th>
<th>1 – Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Location</td>
<td>UK</td>
<td>Intervention is conducted in a country that is economically similar- OECD</td>
<td>Intervention is conducted in a country that is not economically similar- non-OECD</td>
</tr>
<tr>
<td>B. Control Group</td>
<td>There are two or more control groups – attends session outside the normal classroom routine</td>
<td>There is one control group – attends session outside the normal classroom routine</td>
<td>Do not attend an additional session outside of the normal classroom routine</td>
</tr>
<tr>
<td>C. Depth of outcome measures</td>
<td>Multiple measures relating to reading</td>
<td>Single measure of reading</td>
<td>No measure of reading</td>
</tr>
<tr>
<td>D. Follow Up</td>
<td>Includes a follow up some time after post-test to consider impact of intervention on reading ability over time</td>
<td>Incudes post-test findings to evaluate impact of the intervention</td>
<td>No follow up data</td>
</tr>
</tbody>
</table>
Table 6.7

**WoE C Overall Ratings**

<table>
<thead>
<tr>
<th>Studies</th>
<th>Criteria A</th>
<th>Criteria B</th>
<th>Criteria C</th>
<th>Criteria D</th>
<th>WoE C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenihan et al. (2016)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Linder et al. (2018)</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Levinson et al. (2017)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Uccheddu et al. (2019)</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2.25</td>
</tr>
<tr>
<td>Le Roux et al. (2014)</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Appendix F: Coding Protocol

Gersten et al’s (2005) coding protocol was used and adapted to suit the range of studies in this review. Specifically, the participants were from a typical population in this review rather than from a population with specific disabilities or difficulties. In response to this, Question 1 of the Gersten et al. (2005) protocol was edited to become relevant to the target population (see below).


Study:

Author:

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

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

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

Did the study use appropriate participants for the research question?

☐ Yes
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 a sufficiently large sample size used?
☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Total size of sample (start of study): _________

Were participants randomly assigned to conditions?
☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Did the research include a control group?
☐ Yes
☐ No
☐ N/A
☐ Unknown/Unable to Code

Total size of control group (start of study): _________
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

Did the research specify the length of intervention?

☐ 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
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

Did the research use well-known or standardized measures?

☐ 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:

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
<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Essential Quality Indicators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total of &gt;9 = Score 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total of &lt;9 = Score 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Desirable Quality Indicators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total of ≥4 = Score 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total of &lt;4 = Score 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total 0 = Score 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3 = High Quality; 2 = Acceptable Quality; &lt;2 = Poor Quality)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Study Rating:**

*Essential Quality Indicators - Quality indicators for describing participants*

Was sufficient information provided to determine/confirm participants demographics?

☒ 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
Was the fidelity of implementation described and assessed?

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

- ☒ Yes
- ☐ No – described what they did but no standard way of doing in and not assessed
- ☐ N/A
- ☐ Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?

- ☒ Yes – read to human volunteer
- ☐ 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 – before and after
- ☐ 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: 8

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 – not blind? Does not specify
☐ N/A
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 – justified why used and benefits, measures what intended
☐ 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 effect sizes, S.D
☐ No
☐ N/A
☐ Unknown/Unable to Code

Essential Quality Indicators Total Score: 3

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Essential Quality Indicators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total of &gt;9 = Score 1</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Total of &lt;9 = Score 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Desirable Quality Indicators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total of ≥4 = Score 2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Total of &lt;4 = Score 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total 0 = Score 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3 = High Quality; 2 = Acceptable Quality; &lt;2 = Poor Quality)</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Study Rating: Poor Quality