How effective is Guided Reading in improving reading ability of primary-aged children?

1. Summary

The aim of this literature review is to evaluate the effectiveness of Guided Reading in improving the reading ability of primary-aged children.

Guided Reading is characterised by teachers supporting readers in their development of specific strategies aimed at enhancing reading independence. Such strategies include inference and evaluative reading (Fisher, 2008). One of the key features of Guided Reading is the division of sessions into 3 clear stages: before reading, during reading and after reading (Fountas & Pinnell, 1996).

A systematic review of the literature was conducted using 3 online databases (ERIC, PsychINFO and Web of Science) as well as the UCL online library resource. An ancestral search was also carried out following the identification of the first 4 studies for review. In total, 5 studies met the inclusion criteria for review and each of these was evaluated according to Gough’s (2007) weight of evidence framework. 3 of the studies had an experimental design, 1 had a quasi-experimental design and 1 had a single-subject design. All studies included outcome measures pre- and post-intervention.

The studies showed that the effect of Guided Reading on reading ability ranged from not practically significant to medium, using Pretest-Posttest Control Group standardized mean difference (PPC SMD) effect sizes and Cohen’s (1992) effect categories. The largest effect (medium) was seen on the reading comprehension of
English as an additional language learners. Implications of these findings are discussed along with recommendations for future research.

2. Introduction

2.1 Guided Reading

According to Fountas and Pinnell (1996, p. 25), Guided Reading is “an instructional context for supporting each reader’s development of effective strategies for processing novel texts at increasingly challenging levels of difficulty”. Lanning and LaMere (2000) stated that the purpose of Guided Reading is to assist students in the acquisition of both the behaviours and the strategies of effective and autonomous readers. Examples of some of the strategies of effective and autonomous readers include making predictions, sequencing events and comparing and contrasting (Foresman, 2004).

Guided Reading involves small group work in which students are matched according to reading level. The role of the teacher in Guided Reading is critical as they provide differentiated instruction to suit each small group’s reading level. Part of their role is to assist readers in their development of strategies which aim to enhance reading independence (as described above). Teachers support students’ increasing capability and encourage the undertaking of increasingly challenging texts as they make progress (Fountas & Pinnell, 2001).

According to Fountas and Pinnell (2012), Guided Reading enables students to decode while actively attending to the meaning of the text. In addition, the small group nature of this intervention and the matching of students’ levels, means that the social interaction achieved through conversation within groups can further enhance understanding.
One of the key features of Guided Reading is the division of sessions into 3 clear stages: before reading, during reading and after reading (Fountas & Pinnell, 1996). These 3 stages are broken down into 6 steps of instruction, with an optional seventh step (Fountas & Pinnell, 2012). The steps of instruction are as follows:

1. Selection of text (at an instructional level to support the group’s learning)
2. Introduction to text (in order to scaffold the reading)
3. Reading the text (silently or orally; if orally, teacher prompts and reinforces strategy)
4. Discussion of text (teacher-guided discussion aimed at increasing comprehension)
5. Teaching points (grounded in text and directed towards expanding strategies)
6. Word work (explicit teaching to support students solve words)
7. Extending understanding (optional additional written or creative drawing work to extend understanding of the text if required)


2.2 Psychological basis for Guided Reading

A psychological basis for Guided Reading can be seen in Vygotsky’s (1978) Zone of Proximal Development (ZPD) which represents the difference between what a learner can achieve alone, and what he or she can do with guidance and support. In the case of Guided Reading, it is the teacher (as well as a student’s peers) who provides that guidance (via the introduction of strategies and use of prompting to support the expansion of reading skills), or as Fountas and Pinnell (2012) illuminate:
“bringing readers from where they are to as far as the teaching can take them”. This theoretical underpinning is further supported by Wood, Bruner and Ross (1976) who highlighted the importance of provision of scaffolding experiences in learning.

In addition, Fisher (2008) approaches Guided Reading from a social constructivist perspective in which children’s construction of meaning can be achieved through talking, thinking and reading; in other words, meaning is created through social interaction and language. Guided reading supports this in an educational context.

Guided Reading could also be considered in the context of Self-Determination Theory (Deci & Ryan, 1985) in which people’s natural developmental tendencies towards mastering challenges are said to rely, at least in part, upon social support. By providing students with a specific set of strategies aimed at supporting the development of reading skills as well as enhancing independence in reading, teachers are fostering a sense of competence and autonomy in their students. The interactive element of Guided Reading may also facilitate students’ sense of relatedness which, according to Deci and Ryan (1985), is another critical factor which has an impact on a person’s motivation and quality of performance.

2.3 Rationale for Review

The most recent Special Educational Needs and Disability Code of Practice (DfE and DoH, 2015) places emphasis on high aspirations and challenging targets for learners, as well as the importance of early intervention. By focusing on primary-aged students, literacy interventions that target reading such as Guided Reading, could be deemed early intervention. Reading difficulties left unaddressed at the primary stages of education could potentially lead to further difficulty as a child moves through education and beyond. For example, in a study by Rucklidge, Mclean
and Bateup (2013), it was found that there was a high prevalence of learning difficulties among imprisoned youth offenders; in particular, the researchers found that poor reading comprehension was a predictor of future re-offending.

Guided Reading was recommended as part of the National Literacy Strategy (DfES, 2006), and has been implemented in “thousands of classrooms around the world” (Fountas & Pinnell, 2012), however, Foster (2008) argues that the supporting evidence base is lacking.

Knowledge about the efficacy of literacy interventions is an important consideration for Educational Psychology practice. It will assist Educational Psychologists to make a judgement about whether or not a particular intervention (such as Guided Reading) is appropriate for an individual or group of students within a school. Educational Psychologists can then support schools to implement the intervention and monitor and evaluate progress in literacy development. If Educational Psychologists are able to assist schools in employing effective literacy interventions, this has the potential to improve reading attainment across entire classrooms and even schools, reducing the likelihood of longer term literacy difficulties. Therefore, If Educational Psychologists are to recommend Guided Reading specifically as an intervention for use in primary schools, an evaluation of the effectiveness of Guided Reading is clearly required. This will involve a systematic review of the existing literature.

Since reading fluency has been identified as an indicator of overall reading ability (Fuchs, Fuchs & Hosp, 2001) and with comprehension one of the major focuses of Guided Reading (Fountas & Pinnell, 1996; Fountas & Pinnell, 2012), these will be considered as primary indicators of reading ability for the purpose of this review.
2.4 Review Question

How effective is Guided Reading in improving reading ability of primary-aged children?

3. Critical Review of the Evidence Base

3.1 Literature Search

A comprehensive literature search was undertaken using the UCL online library search resource and the online databases ERIC (Education Resources Information Center), PsychINFO and Web of Science between 16th January 2016 and 7th February 2016. In order to identify all studies related to Guided Reading, the reviewer used the following search terms (see Table 1):

<table>
<thead>
<tr>
<th>Table 1. Search terms used with online databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Databases: ERIC, PsychINFO, Web of Science, UCL online library search</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The reviewer also applied filters to the online databases before running the searches in order to exclude any studies that were published before 2006 (see Table 2 for rationale). These searches generated a total of 313 studies from the databases and UCL library search. After duplicates were removed (14), the remaining 299 studies were screened by title and abstract using the inclusion and exclusion criteria (Table 2) leaving a total of 5 studies. 1 of these studies was excluded after full text screening (see Appendix A), leaving a total of 4 studies (see Table 3). A subsequent ancestral search of the reference lists from included studies generated a further 8 studies for title and abstract screening. Of these 8 studies, 7 were excluded using
the inclusion and exclusion criteria and the remaining 1 was included for the review.

Figure 1 represents the literature search and selection process.

*Table 2 (p. 9-10). Inclusion and exclusion criteria*

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Type of publication</strong></td>
<td>The study must appear in a peer reviewed journal OR can be a doctoral thesis which has been approved by an accredited institution.</td>
<td>Research that has been reviewed by experts will be of a higher standard.</td>
</tr>
<tr>
<td><strong>2. Language</strong></td>
<td>The study must be published in English.</td>
<td>To ensure reviewer understanding of the study.</td>
</tr>
<tr>
<td><strong>3. Year of publication</strong></td>
<td>The study was published on or after 1st January 2006.</td>
<td>The reviewer is only concerned with studies conducted in the last 10 years. The Primary Framework for literacy and mathematics, in which guided reading is recommended, was published in 2006.</td>
</tr>
<tr>
<td><strong>4. Research design</strong></td>
<td>a) The study must include primary empirical data.</td>
<td>a) This means that the data are original (i.e. not meta-analyses) and allows for the calculation of effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>5. Intervention</strong></td>
<td>The study must involve the delivery of a Guided Reading intervention in school.</td>
<td>The study does not include the delivery of a Guided Reading intervention in school.</td>
</tr>
<tr>
<td></td>
<td>In order to critically evaluate the effect of Guided Reading on reading ability.</td>
<td></td>
</tr>
<tr>
<td><strong>6. Analysis</strong></td>
<td>The study reports means, standard deviations, test statistics and/or data that allows for these measures to be calculated.</td>
<td>The study does not report means, standard deviations, test statistics and/or data that allows for these measures to be calculated.</td>
</tr>
<tr>
<td></td>
<td>These measures are required in order for effect size to be calculated.</td>
<td></td>
</tr>
<tr>
<td><strong>7. Population</strong></td>
<td>The participants must be of primary school age (5 years to 11 years) and attend a mainstream setting.</td>
<td>Participants fall outside of the age bracket specified. Participants do not attend a mainstream setting.</td>
</tr>
<tr>
<td></td>
<td>This is to ensure that the effect of the intervention on primary aged children attending mainstream settings is reviewed.</td>
<td></td>
</tr>
<tr>
<td><strong>8. Sample</strong></td>
<td>The study has a sample size of 1 or more.</td>
<td>The study has a sample size of 1 (a single case study).</td>
</tr>
<tr>
<td></td>
<td>A higher number of participants increases the reliability of any findings.</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Flow chart of literature search and selection process.

- **PsychINFO**: 23 papers
- **ERIC**: 19 papers
- **Web of Science**: 190 papers
- **UCL online library search**: 81 papers

**Excluded:**
- Did not meet inclusion criteria: 294 papers
- Duplicates: 14

**Full article screening**: 5 papers

**Excluded:**
- 1 study (Montoya, 2008): exclusion criteria 7

**Included:**
- 4 studies

**Ancestral search**: 8 studies

**Excluded:**
- 4 studies (Anderson et al., 2002; Bruce et al., 1999; Chevalier et al., 2002; Gabl et al., 2007): exclusion criteria 1
- 3 studies (Bruce et al., 2002; Klingbeil, 2003; Walizer, 2004): exclusion criteria 3

**Included:**
- 1 study

**Final literature review**: 5 studies
Included studies


Table 3. Included Studies

3.2 Weight of Evidence (WoE)

The 5 studies selected for the review were evaluated according to Gough’s (2007) weight of evidence (or WoE) framework. This framework enables the reviewer to judge studies according to 3 key areas: methodological quality (WoE A), methodological relevance (WoE B) and relevance to the review question (WoE C). WoE B and C are review-specific judgements which examine the extent to which a study is ‘fit for purpose’ as a piece of evidence addressing the review question (Gough, 2007). After these 3 key areas have been evaluated, the reviewer can then judge the overall value of each study as evidence for answering the review question (WoE D). Table 4 provides an overview of the weight of evidence for each of the 5 studies included in this review. Further information regarding each individual weight of evidence can be found in Appendix C and example coding protocols can be found in Appendix D.

Table 4. Weight of Evidence judgements

<table>
<thead>
<tr>
<th>Study</th>
<th>Methodological quality (WoE A)</th>
<th>Methodological relevance (WoE B)</th>
<th>Relevance to review question (WoE C)</th>
<th>Overall weighting (WoE D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobin &amp; Calhoon (2009)</td>
<td>Low (1)</td>
<td>Medium (2)</td>
<td>Medium (2.33)</td>
<td>Low (1.78)</td>
</tr>
<tr>
<td>Nayak &amp; Sylva (2013)</td>
<td>Medium (2)</td>
<td>High (3)</td>
<td>Medium (2.33)</td>
<td>Medium (2.44)</td>
</tr>
<tr>
<td>Oostam,</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
</tr>
</tbody>
</table>

3.3 Participants

A total of 656 primary-aged students were included in this review. Most participants (312) were attending schools in the USA; the remaining 2 studies sampled students in Hong Kong and The Netherlands. None of the studies reviewed recruited participants from UK schools, and so this has implications for the generalisability of any findings to UK student populations attending UK educational settings.

Participants in 2 of the studies were classified as being ‘at-risk’ for reading difficulties (Bruce, 2010; Denton et al., 2014); 1 study classified its participants as being ‘poor readers’ (Oostdam, Blok & Boendermaker, 2015). Classification criteria differed across these studies; Bruce’s (2010), participants were classified as being ‘at-risk’ if they scored below grade level on their school reading assessment (STAR reading assessment); participants in Denton et al.’s (2014) study were regarded as ‘at-risk’ for reading difficulties if they scored below the 30th percentile on either the Test of Word Reading Efficiency (TOWRE) or the Woodcock Johnson III Tests of Achievement (WJ III). In Oostdam, Blok and Boendermaker’s (2015) study, students were identified as ‘poor readers’ if their grade-normed scores on a test of word reading fluency (FWR 345678) fell within the lowest quintile.

Participants in Tobin and Calhoon’s (2009) study were considered eligible if they attended 1 of 2 schools participating in the research; in this case, the intervention
was delivered as a class-wide reading intervention for first-grade students. Participants in Nayak and Sylva’s (2013) study were Cantonese speakers (L1) learning English as an additional language (L2). This has implications for the generalisability of their findings to groups other than EAL learners.

Sampling method was generally not explicitly stated, however, it appears that researchers acquired their samples mostly through convenience sampling (Bruce, 2010; Denton et al., 2014; Oostam, Blok & Boendermaker, 2015; Tobin & Calhoon, 2009). This method of non-random sampling, although common in Psychological research, is often biased and unrepresentative of the target population (Barker, Pistrang & Elliot, 2002). Nayak and Sylva (2013) did make use of random sampling, which means it is with greater confidence that inferences about the target population (EAL learners) can be made from their findings.

Power analyses were conducted for each study to determine appropriate sample sizes as recommended by Cohen (1992). As there were no effect sizes available in the previous literature, the reviewer conducted the analyses based on a medium effect size at power 0.8 and an alpha level of 0.5. Three of the studies were sufficiently powered (Denton et al., 2014; Nayak & Sylva, 2013 and Oostam, Blok & Boendermaker, 2015) and two (Bruce, 2010 and Tobin & Calhoon, 2009) had insufficient sample sizes. The insufficient power of these two studies limits the confidence with which any findings or effect sizes can be considered.

3.4 Research design

Experimental, quasi-experimental and single-subject designs were included in this review. Denton et al. (2014), Nayak and Sylva (2013) and Oostam, Blok and Boendermaker (2015) used experimental designs with random assignment of
participants to intervention group, no-intervention control and, where relevant, alternative-intervention control (Denton et al., 2014 and Nayak & Sylva, 2013). In both Nayak and Sylva (2013) and Oostam, Blok and Boendermaker (2015), researchers used blocking / stratification to randomise participants into each group. Nayak and Sylva (2013) stratified according to gender and reading proficiency, and Oostam, Blok and Boendermaker (2015) stratified according to reading proficiency (‘poor readers’ were matched before assignment to intervention or no-intervention control groups). This means that experimental groups were balanced on variables of individual difference, reducing the likelihood of bias in these studies (Barker, Pistrang & Elliott, 2002). All three of these studies subsequently received ‘High’ ratings on Methodological Relevance, or Weight of Evidence B (WoE B – see Table 4).

Tobin and Calhoon (2009) used a quasi-experimental design in which there was no randomisation of participants to the intervention or alternative-intervention control groups. This study subsequently received a lower rating on WoE B due to the uncontrolled selection threatening the internal validity; participants in the 2 groups may have differed systematically in ways that were unaccounted for by solely the presence or absence of the Guided Reading intervention (Barker, Pistrang & Elliott, 2002).

The remaining study (Bruce, 2010), used a single-subject design in which the ‘subject’ is represented by a single group; this one group involved in the research received the Guided Reading intervention. One of the issues with this type of design is the chance that another event external to the intervention may occur at the same time as the intervention, threatening the internal validity and potentially impacting upon the outcome variable (reading ability) (Barker, Pistrang & Elliott, 2002). According to Horner et al. (2005), measurement of the dependent variable (reading
ability) during a baseline should occur at multiple points (typically five or more) in order to be considered adequate; this did not occur in Bruce’s (2010) study and it therefore received Medium and Low ratings on WoE A and WoE B, respectively.

3.5 Intervention

One of the criteria for receiving a rating of ‘High’ on WoE C was that the Guided Reading intervention implementation needed to be fully described with an example lesson plan provided. This was to ensure that researchers were explicit about the nature of the intervention provided and so that it would be clear to the reviewer how the researchers had interpreted ‘Guided Reading’. Three of the studies reviewed did this (Bruce, 2010; Denton et al., 2014; and Nayak & Sylva, 2013). Tobin and Calhoon (2009) provided a description of the Guided Reading ‘approach’, however, there was no sample lesson plan provided, and so in order for their intervention to be replicated, further information would be required. This influenced the WoE A, WoE B and WoE C ratings and the ultimate WoE D rating, which was ‘Low’. The reviewer regarded the description of the Guided Reading intervention provided by Oostam, Blok and Boendermaker (2015) as adequate, but it did not include a sample lesson plan.

Fidelity of implementation was also regarded as a key factor indicative of the relevance of a study to the review question, and so this was included in the criteria for WoE C. This is because, in order to judge the effectiveness of the intervention, the reviewer needs to be certain that the intervention was delivered as specified by the researchers. Otherwise, any effect or outcome measure may not be attributable solely to the intervention. Three of the studies reviewed reported high levels of fidelity (Denton et al., 2014; Oostam, Blok & Boendermaker, 2015; and Tobin &
Calhoon, 2009). Tobin and Calhoon (2009) reported the use of a ‘Qualitative Monitoring of Instruction’ survey in which the principal of each school observed each teacher delivering the intervention on 2 occasions. Although they report a high level of fidelity, this result should be taken with caution due to the fact that it was the principal observing and not an impartial observer; this could lead to potential biases in the monitoring of fidelity. Oostam, Blok and Boendermaker (2015) measured fidelity through logs kept by the teaching assistants delivering the intervention (teaching assistants were instructed to keep a log of any deviation from the intervention protocol). Denton et al. (2014) monitored fidelity through researcher observations using protocols which reflected key features of the intervention; these observations were carried out on 3 occasions throughout the research process.

Intervention-specific training was another key indicator of relevance to the review question (WoE C). This is because intervention-specific training would mean the likelihood that it would be delivered as anticipated by the researchers would be higher. This may have a subsequent impact on fidelity of implementation. Three of the studies (Denton et al., 2014; Nayak & Sylva, 2013; and Tobin & Calhoon, 2009) involved delivery of the intervention by teachers or researchers who had received training specific to Guided Reading. Oostdam, Blok and Boendermaker (2015) used teaching assistants to implement the intervention and subsequently received a lower rating for this criteria due to the lower level of qualification required from these members of school staff. Similarly, Bruce (2010), although stating that qualified teachers delivered the intervention, did not specify whether or not they received Guided-Reading-specific training beforehand and so, it was not possible to rate highly for this criteria.
3.6 Measures

Reliability of measures, along with the use of multiple measures which are associated with the intervention and generalised performance are both considered to be essential quality indicators according to Gersten et al. (2005). Similarly, Horner et al. (2005) highlight the importance of precise, reliable, replicable and valid measurement of the dependent variable. These factors were therefore considered when examining WoE A and are reflected in reviewer judgements (see Table 4 and Appendix C and D).

There was a diverse range of measures used in the 5 studies reviewed. Two of the studies (Bruce, 2010; Tobin & Calhoon, 2009) used computer-based assessment systems to measure participant reading ability. Tobin and Calhoon (2009) used alternate forms of a computerised oral reading fluency (ORF) assessment in order to measure participants’ reading fluency pre- and post- intervention. They reported equivalent forms reliability coefficients of this measure as ranging from .89 to .94 which, according to Barker, Pistrang and Elliot’s (2002) suggested reliability standards is ‘Good’. Bruce (2010) reported reliability coefficients for the STAR reading assessment (STAR Reading, 2006) ranging from .89-.92 (for “generic” reliability); .89-.93 (split-half reliability); .94 (test-retest reliability) and .95 (equivalent forms reliability). Bruce (2010) also reported an average validity coefficient for this measure of .72.

Nayak and Sylva (2013) used the Neale Analysis of Reading Ability (NARA II; Neale, 1989) to assess both reading comprehension and reading accuracy pre- and post-intervention. Interestingly, despite having a relatively stringent design and method
otherwise (WoE A = Medium; WoE B = High), Nayak and Sylva did not report any reliability or validity coefficients for their chosen measure.

Oostam, Blok and Boendermaker (2015) used the AVI (Jongen & Krom, 2009) which they report is a standardised, grade-normed assessment of fluency of text reading. The reported reliability of this measure is .96. The Begrijpend Lezen (Krom, Van Berkel, & Jongen 2006–10) was used to assess reading comprehension and reliability for this measure was reported as ranging from .76-.89.

Denton et al. (2014) used the Test of Word Reading Efficiency (TOWRE) to assess reading fluency and reported Torgesen, Wagner, and Rashotte’s (1999) findings of an alternate forms reliability as exceeding .90 and test-retest reliabilities that ranged from .83 to .96. Reading comprehension was assessed using the WJ III (Woodcock, McGrew & Mather, 2001)) and Gates-MacGinitie (MacGinitie et al., 2000) Reading Comprehension subtest; researchers quoted a test-retest reliability of the WJ III as .86 (McGrew & Woodcock, 2001) and a reliability coefficient for the Gates-MacGinitie Reading Comprehension subtest as ranging from .92 to .92 (MacGinitie et al., 2000).

3.7 Outcomes and Effect Sizes

The focus of this review is to evaluate the effect of Guided Reading on reading ability. Outcome measures which give an indication of reading ability will therefore be the focal point. Reading fluency has been identified as an indicator of overall reading ability (Fuchs, Fuchs & Hosp, 2001) and with reading comprehension being one of the major focuses of Guided Reading (Fountas & Pinnell, 1996; Fountas & Pinnell, 2012), this outcome will also be included for the purpose of this review, along with grade equivalent where available (Bruce, 2010).
Every study measured outcomes pre-intervention and post-intervention. For studies with group-based designs, these pre- and post- outcomes were also reported for the no-intervention control group / alternative intervention group. The Pretest-Posttest Control Group standardized mean difference (PPC SMD, Morris, 2008) was therefore used to calculate effect sizes for all 4 studies with a control group (Denton et al., 2014; Nayak & Sylva, 2013; Oostam, Blok & Boendermaker, 2015; Tobin & Calhoon, 2009). Where a study included both a no-intervention control group and an alternative intervention control group, the no-intervention control group was used. One study (Bruce, 2010) had a single-subject design so the standardized mean difference, or SMD, (Becker, 1988) showing within-person change over time (pre- and post-intervention) was used to calculate the effect sizes for this particular study. All effect sizes can be found in Table 5.

Table 5. Effect sizes (PPC SMD and SMD) for outcomes.

<table>
<thead>
<tr>
<th>Study</th>
<th>Measure</th>
<th>Outcome</th>
<th>Effect size with descriptor (Cohen, 1992)</th>
<th>Effect size with descriptor (Cohen, 1992)</th>
<th>WoE D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>PPC SMD</td>
<td>SMD</td>
<td></td>
</tr>
<tr>
<td>Tobin &amp; Calhoon (2009)</td>
<td>AIMSweb</td>
<td>Oral Reading Fluency (ORF)</td>
<td>-.53 Medium</td>
<td>.86 Large</td>
<td>2.22</td>
</tr>
<tr>
<td>Nayak &amp; Sylva (2013)</td>
<td>NARA</td>
<td>Comprehension Accuracy</td>
<td>.55 Medium</td>
<td>.82 Large</td>
<td>2.44</td>
</tr>
<tr>
<td>Oostam, Blok &amp; AVI</td>
<td>AVI</td>
<td>Fluency of text reading</td>
<td>.28 Small</td>
<td>.45 Small</td>
<td>2.67</td>
</tr>
<tr>
<td>Study</td>
<td>Measure</td>
<td>Effect Size</td>
<td>Significance</td>
<td>Size</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Boendermaker (2015)</td>
<td>Begrijpend Lezen</td>
<td>-0.06</td>
<td>Not practically</td>
<td>Small</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reading comprehension</td>
<td></td>
<td>significant</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.43</td>
<td>.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denton et al. (2014)</td>
<td>WJ III Comprehension</td>
<td>-0.06</td>
<td>Not practically</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.15</td>
<td>significant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gates MacGinitie</td>
<td>Comprehension</td>
<td>.16</td>
<td>Not practically</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>TOWRE</td>
<td>Word reading fluency</td>
<td>.20</td>
<td>Small</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.40</td>
<td>Small</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bruce (2010)</td>
<td>STAR Reading</td>
<td>-</td>
<td>Not practically</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructional reading</td>
<td>.03</td>
<td>significant</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grade equivalent</td>
<td>-</td>
<td>Not practically</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.04</td>
<td>significant</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effect size descriptors were interpreted using Cohen’s (1992) effect categories (small, medium or large). Any effect sizes that were less than .2 (small) have been regarded as not practically significant.

As can be seen in Table 5, for all studies where both the SMD and PPC SMD were calculated, the PPC SMD effect sizes were smaller in comparison to the SMD. This was expected due to the fact that the PPC SMD considers the size of an effect in comparison to no treatment at all (i.e is the intervention better than no treatment at all?), whereas the SMD looks solely at the pre-post change in relation to the intervention. SMD Effect sizes ranged from not practically significant to large and PPC SMD effect sizes ranged from not practically significant to medium. The largest PPC SMD effect size can be seen for Nayak and Sylva’s (2013) data. The outcomes
considered here were reading comprehension (.55, medium) and reading accuracy (.28, small). This study received an overall weight of evidence (WoE D) judgement of 2.44 – it was considered to be of Medium weighting due to the research design used (e.g. randomization of participants). Participants in this study were learning English as an additional language. It could be that Guided Reading is a relatively effective intervention to use with this particular group of learners (although the effect sizes were still only small-medium) as opposed to learners who are considered to be ‘at-risk’ for reading difficulties for example. In particular, it appears that Guided Reading may be an effective intervention for targeting the reading comprehension of this group of learners.

Two of the studies reviewed (Oostam, Blok & Boendermaker, 2015; and Denton et al., 2014) both produced effect sizes for reading comprehension that were not practically significant. This raises the question of whether or not Guided Reading is an appropriate intervention for those readers who are considered ‘at-risk’ or ‘poor readers’, particularly in terms of targeting reading comprehension. Similarly, Bruce’s (2010) study produced effect sizes (SMD) that were not practically significant. The target population for this research was also students who were ‘at-risk’. This adds to the evidence that Guided Reading is not the most effective intervention for this particular group of students.
4. Conclusion and Recommendations

The aim of this systematic literature review was to evaluate the effectiveness of Guided Reading in improving the reading ability of primary-aged children. Following a literature search, a total of 5 studies were identified as meeting the inclusion criteria for the review. Subsequent evaluation using Gough’s (2007) weight of evidence framework, along with calculation and scrutiny of effect sizes has led to the conclusion that overall, these studies provide a limited body of evidence for the effectiveness of Guided Reading at improving reading ability in primary-aged children. Although some of the studies received overall weight of evidence ratings (WoE D) that were medium (Nayak & Sylva, 2013) or high (Denton et al., 2014; and Oostam, Blok & Boendermaker, 2015), the effect sizes (PPC SMD) were generally either not practically significant or small. Interestingly, a medium effect of the intervention on reading comprehension scores (PPC SMD = .55) in Nayak and Sylva’s (2013) study was found. As discussed in the previous section, participants in this study were learning English as an additional language; in the context of this review, Guided Reading could therefore be considered a relatively effective intervention for this particular group of learners. However, this is clearly tentative, and these findings would need to be replicated before generalisation could be considered. Hence, one of the recommendations as a result of this review is that further research into the effectiveness of Guided Reading on the reading ability (and specifically, the reading comprehension) of English as an additional language learners should be carried out. This would enable future reviewers to make a more informed judgement on whether or not Guided Reading is particularly suited to English as an additional language learners.
Given that none of the studies included in this review were conducted in the UK, it is difficult to consider these findings in the context of UK primary schools. More up to date research is required in the UK in order to evaluate Guided Reading in this context before it can be considered for recommendation by professionals such as Educational Psychologists. Any recommendation of Guided Reading currently would have to be imparted with caution, given the limited research evidence base for its effectiveness.

Another recommendation on the basis of this review would be the need for research with more rigorous research designs i.e. randomised controlled trials. This would enhance the methodological quality of the research and any findings could be taken with greater confidence; any change in reading ability could be more confidently attributed to the intervention itself and not systematic differences between groups for example.

Before more research is carried out, it is clear that the evidence base at present is not sufficient to support the effectiveness of Guided Reading at improving the reading ability of primary-aged children.
5. References:


## Appendix A: Excluded Studies

Studies excluded at full-text screening with rationale

<table>
<thead>
<tr>
<th>Study reference</th>
<th>Reason for exclusion</th>
</tr>
</thead>
</table>

Studies excluded at ancestral search with rationale (p. 28-29).

<table>
<thead>
<tr>
<th>Study reference</th>
<th>Reason for exclusion</th>
</tr>
</thead>
</table>


Criteria 3. Published prior to 1st January 2006.
### Appendix B (p. 30-36): Summary of Included Studies (Mapping the Field)

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Sample</th>
<th>Design</th>
<th>Intervention</th>
<th>Relevant Measures</th>
<th>Primary Outcomes Relating to Relevant Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobin &amp; Calhoon (2009)</td>
<td>USA</td>
<td>107 first-grade students from 2 elementary schools in a north eastern US state.</td>
<td>Intervention group (IG): 60 Alternative Intervention group (AIG): 47</td>
<td><strong>Guided reading (IG):</strong> Instruction is divided into blocks. In the first block the teacher leads a class discussion to establish background knowledge of the story to be read. This is followed by a guided reading block in which students are divided into groups; a book will be read and discussed. During the discussion comprehension strategies are taught.</td>
<td><strong>Oral Reading Fluency (ORF)</strong> Assessed via ‘AIMSweb’ Computer-based assessment system. 1 minute oral reading test giving the number of correct words per minute. Uses first-grade neutral reading passages. Students given 3 passages in total, researchers used score from median passage.</td>
<td><strong>Effect of reading programme (intervention and alternative) on Oral Reading Fluency</strong> Researchers ran a repeated measures ANOVA to compare time (Winter [pre] &amp; Spring [post]) and Condition (Horizons Fast Track A-B &amp; Guided Reading) Researchers report a significant effect for Time for both conditions ($F(1, 105) = 22.85, p = .000$). Report large effect size (.92) favouring students in Horizons group (AIG).</td>
</tr>
</tbody>
</table>
Nayak & Hong Kong (2013) 205 Year 4 students from 6 schools. Eligibility of students to participate depended on their language learning backgrounds (Cantonese L1, English L2) – this was established via parent self-report questionnaire.

Intervention Group (IG): 70 Alternative Intervention Group (AIG): 67 Control Group (CG): 68

Each ‘treatment’ was delivered for 35 minute weekly sessions over an 8 week period along with the normal classroom English instruction.

Guided Reading (IG): Each session split into 3 stages: before reading (teacher introduction, key words and themes introduced); during reading (teacher/researcher read story aloud, key vocabulary & story development highlighted, predictive questions; each child read a section of the story aloud followed by feedback on accuracy); and after reading (teacher/researcher initiated discussion on comprehension strategies, negotiation of meaning).

E-book reading (AIG): Children were assigned a computer with headphones. Children went through all pages of Neale Analysis of Reading Ability (NARA II) Produces scores for reading accuracy (number of words read aloud correctly) and reading comprehension (number of questions about the meaning of passages answered correctly). Consists of 2 practice passages and 6 test passages of graded narrative text with comprehension questions at the end of

Effect of intervention on reading comprehension

Repeated measures ANOVA, one within-subjects variable (time, 2 levels) and one between-subjects variable (group, 3 levels).

Researchers reported a main effect for time ($F(1, 202) = 89.31, p < .05$) but not for group ($F(1, 202) = .78, p = ns$). Report significant group x time interaction ($F(2, 202) = 6.38, p < .05$). Post hoc tests (2 x 2 repeated-measures ANOVAs) showed significant time x group interaction between guided reading vs. control reading.

Effect of intervention on reading accuracy

Repeated measures ANOVA. Main effect for time ($F(1, 202) = 85.98, p < .05$) but no main effect for group ($F(1, 202) = .47, p = ns$). Significant group x time interaction ($F(1, 202) = 5.40, p < .05$). Post hoc tests
Oostm, Blok & Boendermaker (2015) The Netherlands 139 students from 8 primary schools across Grades 2, 3, and 4. Pre-test, 2 mid-tests, and post-test were conducted. Students had been identified as "poor readers" according to national norms; fall within the lowest quintile on the FWR Grade 3 test.

UK standardised and used to assess EAL learners: AVI A standardised test of fluency of text reading which consists of 11 short passages of increasing difficulty. Number of passages read within time and accuracy limits defines a student's score (0-11).

Effect of intervention on fluency of text reading
Researchers used multilevel modelling and found students in the Group-Based Guided Oral Reading group improved in text reading fluency compared to the control group.

Effect of intervention on comprehension
There was no significant effect of intervention on comprehension in Guided Reading group.

The e-book (with option to listen to computerized voice reading aloud with animations and sound effects) and then completed word, sentence and text level activities with automated feedback provided.

Available in parallel forms (Form A and Form B) which were used pre- and post-intervention.

Researchers used multilevel modelling and found students in the Group-Based Guided Oral Reading group improved in text reading fluency compared to the control group.

There was no significant effect of intervention on comprehension in Guided Reading group.

Effect of intervention on fluency of text reading
 Researchers used multilevel modelling and found students in the Group-Based Guided Oral Reading group improved in text reading fluency compared to the control group.

Effect of intervention on comprehension
There was no significant effect of intervention on comprehension in Guided Reading group.
which measures fluency of word reading) matched (blocked) into pairs according to pre-test scores on fluency of word reading. Each pair were then randomly allocated to either treatment or control group.

| Denton et al. (2014) | USA | 162 Grade 1 & 2 students from 9 schools across 2 school districts in south western USA. Eligibility of students to participate was dependent on their | Experimental design. Pre-/post-test design. Students randomly assigned to IG, AIG or CG. Students in the IG and AIG had sessions for 45 minutes, 4 days per week, for 23 to 25 weeks along with their normal classroom instruction. **Guided Reading (IG):** Researchers developed a manual (based on the Guided Reading literature) which included all lesson content. | **Test of Word Reading Efficiency (TOWRE)** Sight Word Efficiency & Phonemic Decoding Efficiency subtests were used to assess word reading. **Effect of intervention on reading fluency** Comparisons of model-adjusted means from ANCOVA revealed non-significant treatment effect on reading fluency. Researchers reported 31% additional gains for Guided Reading over typical instruction (CG). |

**Intervention Group (IG):** 84

**Control Group (CG):** 55

Begrijpend Lezen

Standardised grade-normed test of reading comprehension. Students read text passages silently and answer questions. Grade-specific text booklets contain 5-10 passages and 25 multiple choice questions.

Denton et al. (2014) USA 162 Grade 1 & 2 students from 9 schools across 2 school districts in south western USA. Eligibility of students to participate was dependent on their

Experimental design. Pre-/post-test design. Students randomly assigned to IG, AIG or CG. Students in the IG and AIG had sessions for 45 minutes, 4 days per week, for 23 to 25 weeks along with their normal classroom instruction. **Guided Reading (IG):** Researchers developed a manual (based on the Guided Reading literature) which included all lesson content. **Test of Word Reading Efficiency (TOWRE)** Sight Word Efficiency & Phonemic Decoding Efficiency subtests were used to assess word reading. **Effect of intervention on reading fluency** Comparisons of model-adjusted means from ANCOVA revealed non-significant treatment effect on reading fluency. Researchers reported 31% additional gains for Guided Reading over typical instruction (CG).
performance on the Test of Word Reading Efficiency (TOWRE) or the Basic Skills Cluster on the Woodcock Johnson III Tests of Achievement (WJ III); students had to score below the 30th percentile to be considered at-risk for reading difficulties.

Intervention group (IG): 50

Alternative Intervention Group (AIG): 59

Control Group (CG): 53

components and instructional activities. Each lesson had 4 main components: introducing the text; supporting effective reading; teaching processing strategies; and discussing and revisiting the text. Intervention was implemented with texts levelled for difficulty.

Explicit Instruction (AIG) comprised 3 main elements: explicit, sequential, phonologically based instruction in word reading, phonemic decoding and spelling; text reading practice to build fluency; and explicit, sequential comprehension instruction.

Woodcock Johnson III Tests of Achievement (WJ III) Passage Comprehension subtest was used to assess reading comprehension. Cloze-based assessment in which students read brief sentences / paragraphs and supply missing words.

Gates-MacGinitie Reading Comprehension subtest. Involves reading passages and answering multiple choice questions.

Effect of intervention of reading comprehension

Comparisons of model-adjusted means from ANCOVA revealed no significant difference between Guided Reading group and CG and AIG on both measures of reading comprehension.
Bruce (2010) USA 43 fourth-grade students from 9 classes in 1 elementary school in the state of Georgia. Eligibility of students to participate was based on their identification of being ‘at-risk’ because they scored below grade level on the STAR reading assessment.


Baseline/pre-/post-test design.

After the first 9 weeks of the study (during which students were receiving whole-group reading instruction), participants were grouped based on their pre-intervention STAR reading grade equivalent scores – students with similar scores were grouped together.

Guided Reading (IG): Small group instruction for students on similar reading levels. Teachers supported students to process increasingly challenging texts with understanding and fluency. Intervention consisted of before, during and after-reading activities.

STAR Reading Computer-adaptive assessment which provides norm-referenced reading scores for students in grades 1-12. It consists of 25 different questions for each test. Provides teacher with a summary report which includes Grade Equivalent, Normal Curve Equivalent, and Instructional Reading Level.

Effect of intervention on Grade Equivalent (reading comprehension)

A repeated measures ANOVA comparing STAR Grade Equivalent gain scores (gains from Time 1 to Time 2 and from Time 2 to Time 3) revealed a non-significant result ($F(1, 42) = 1.97, p = .168$). This indicates that change in grade equivalent scores prior to intervention was not significantly different from the change in grade equivalent scores post-intervention.

Effect of intervention on instructional reading level (reading comprehension)

A repeated measures ANOVA comparing STAR Instructional Reading Level gain scores (gains from Time 1 to Time 2 and from Time 2 to Time 3) revealed a non-significant result ($F(1, 42) = 2.43, p = .126$). This indicates...
that changes in instructional reading levels prior to intervention was not significantly different from the change in instructional reading levels post-intervention.
Appendix C: Weight of Evidence

All of the studies included in this systematic literature review were rated for methodological quality (Weight of Evidence A), methodological relevance (Weight of Evidence B) and relevance to the review question (Weight of Evidence C). An overall rating (Weight of Evidence D) was subsequently calculated as the average of Weight of Evidences A, B and C.

Weight of Evidence A (WoE A): Methodological quality

In order to establish the methodological quality of the studies included in this systematic literature review, the reviewer used the Gersten et al. (2005) coding protocol for studies with group experimental and quasi-experimental designs. The reviewer also used the Horner et al. (2005) coding protocol for studies with single-subject research designs.

For the Gersten et al. (2005) coding protocol, studies were rated for both essential and desirable quality indicators. In order to receive ‘high’, ‘medium’ or ‘low’ quality ratings, studies must have fulfilled at least the number of criteria shown in the table below:

<table>
<thead>
<tr>
<th></th>
<th>High (3)</th>
<th>Medium (2)</th>
<th>Low (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential criteria</td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Desirable criteria</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary of Weight of Evidence A (WoE A) Ratings

Group designs

<table>
<thead>
<tr>
<th>Study</th>
<th>Essential criteria</th>
<th>Desirable criteria</th>
<th>WoE A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobin &amp; Calhoon (2009)</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Nayak &amp; Sylva (2013)</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Oostam, Blok &amp; Boendermaker (2015)</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Denton et al. (2014)</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

For the Horner et al. (2005) coding protocol, scores from each of the 7 sections were averaged in order to give an overall numerical rating for methodological quality.

To receive a ‘High’ rating, a study must receive an average rating of 2.5 or above.

To receive a ‘Medium’ rating, a study must receive an average rating between 1.5 and 2.4.

To receive a ‘Low’ rating, a study must receive an average weighting of 1.4 or below.
Single-subject designs

<table>
<thead>
<tr>
<th>Study</th>
<th>Description of participants &amp; setting</th>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Baseline</th>
<th>Experimental control/internal validity</th>
<th>External validity</th>
<th>Social validity</th>
<th>WoE A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bruce (2010)</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2.43</td>
</tr>
</tbody>
</table>
Weight of Evidence B (WoE B): Methodological relevance

Weighting for methodological relevance takes into account whether or not the research design was suitable for evaluating the impact of a school-based Guided Reading intervention on the reading ability of primary-aged students. Criteria for WoE B are informed by Brannen’s evidence hierarchies (1992). Ratings for high (3), medium (2) and low (1) can be found in the table below.

<table>
<thead>
<tr>
<th>High (3)</th>
<th>Medium (2)</th>
<th>Low (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Random assignment of participants</td>
<td>• Comparison group</td>
<td>• No comparison group</td>
</tr>
<tr>
<td>• No-intervention control or alternative intervention control group</td>
<td>• Non-random assignment of participants</td>
<td>• Non-randomisation</td>
</tr>
<tr>
<td>• Pre- and post-intervention outcome data collected</td>
<td>• Pre- and post-intervention outcome data collected</td>
<td>• Pre- and post-intervention outcome data collected</td>
</tr>
<tr>
<td></td>
<td>• If single-subject design, at least 3 demonstrations of experimental effect at 3 different points in time (this can include multiple baseline)</td>
<td>• If single-subject design, fewer than 3 demonstrations of experimental effect at 3 different points in time</td>
</tr>
</tbody>
</table>

Weight of Evidence C (WoE C): Relevance to review question

<table>
<thead>
<tr>
<th>High (3)</th>
<th>Medium (2)</th>
<th>Low (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Guided Reading intervention is delivered by teachers or psychologists who have received training specific to the intervention</td>
<td>• Guided Reading intervention is delivered by trained teachers</td>
<td>• Guided Reading intervention is delivered by school support staff (e.g. Learning Support Assistants)</td>
</tr>
<tr>
<td>• High fidelity reported</td>
<td>• Level of fidelity reported</td>
<td>• No level of fidelity reported</td>
</tr>
<tr>
<td>• Guided Reading intervention is fully described and includes sample lesson plan</td>
<td>• Guided Reading intervention is described in sufficient detail but includes no sample lesson plan</td>
<td>• Guided Reading intervention is described in insufficient detail and includes no sample lesson plan</td>
</tr>
</tbody>
</table>


In order to calculate total WoE C, scores for Training, Fidelity and Description for each study are averaged.

To receive a ‘High’ overall rating, a study must receive an average rating of 2.5 or above.

To receive a ‘Medium’ overall rating, a study must receive an average rating between 1.5 and 2.4.

To receive a ‘Low’ overall rating, a study must receive an average weighting of 1.4 or below.

<table>
<thead>
<tr>
<th>Study</th>
<th>Training</th>
<th>Fidelity</th>
<th>Description</th>
<th>Total WoE C</th>
<th>Overall Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobin &amp; Calhoon (2009)</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2.33</td>
<td>(Medium)</td>
</tr>
<tr>
<td>Nayak &amp; Sylva (2013)</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2.33</td>
<td>(Medium)</td>
</tr>
<tr>
<td>Oostam, Blok &amp; Boendemaker (2015)</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>(Medium)</td>
</tr>
<tr>
<td>Denton et al. (2014)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>(High)</td>
</tr>
<tr>
<td>Bruce (2010)</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>(Medium)</td>
</tr>
</tbody>
</table>

Weight of Evidence D (WoE D): Overall weighting

In order to calculate overall weighting (WoE D), scores for WoE A, B and C for each study are averaged.

To receive a ‘High’ overall rating, a study must receive an average rating of 2.5 or above.

To receive a ‘Medium’ overall rating, a study must receive an average rating between 1.5 and 2.4.

To receive a ‘Low’ overall rating, a study must receive an average weighting of 1.4 or below.
Appendix D: Coding Protocols

Gersten et al., (2005). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education

Date: 11.02.2016

Full Study Reference:


Intervention Name (description of study): The Guided Reading intervention was delivered by teachers specifically trained in Guided Reading Researchers used a computerised assessment programme (AIMSweb) with a measure of Oral Reading Fluency (ORF); reading fluency was the outcome measure.

Research design: Quasi-experimental pre-post design with an intervention group and an alternative intervention control group. Participants in the 1 of participating schools received the intervention and participants in the other school received the alternative intervention (non-random assignment of participants).

Type of Publication: Journal article

Study ID: 1

Essential Quality Indicators

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?
Doctorate in Educational and Child Psychology

Harriet Courtney

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

Implementation of the Intervention and Description of Comparison Conditions

Was the intervention clearly described and specified?
☐ Yes
☒ No  A clear description was provided but this was not sufficient for replication and no sample lesson plan was provided
☐ 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

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

Data Analysis

Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?
Did the research report include not only inferential statistics but also effect size calculations?

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

**Desirable Quality Indicators**

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  *No data available on attrition rates among intervention samples*
☐ 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  *Only alternate form reliability provided. Data collectors were not blind to conditions.*
☐ N/A
☐ Unknown/Unable to Code

Were outcomes for capturing the intervention’s effect measured beyond an immediate posttest?

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

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

☑ Yes  *Criterion-related validity only*
☐ 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  Researchers had the school Principals complete a ‘Qualitative Monitoring of Instruction Survey’ on each teacher on 2 occasions
☐ No
☐ N/A
☐ Unknown/Unable to Code

Was any documentation of the nature of instruction or series provided in comparison conditions? ☒ Yes  Description of Horizons Fast Track A-B detailing lesson structure and with an example of teacher instruction
☐ 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

**Overall Rating of Evidence:** ☒ 3 ☒ 2 ☐ 1 ☐ 0
Horner et al., (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education

Quality Indicators Within Single-Subject Research

Article Reference: Bruce, L. T. (2010). The Effects of Guided Reading Instruction on the Reading Comprehension and Reading Attitudes of Fourth-Grade At-Risk Students.

Type of Publication: Doctoral thesis

Study ID: 5

Description of Participants and Setting
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

The process for selecting participants is described with operational precision.

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

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

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

Overall Rating of Evidence: ☐ 3 ☐ 2 ☐ 1 ☐ 0

Dependent Variable

Dependent variables are described with operational precision.

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

Each dependent variable is measured with a procedure that generates a quantifiable index.
Measurement of the dependent variable is valid and described with replicable precision.
- Yes
- No
- N/A
- Unknown/Unable to Code

Dependent variables are measured repeatedly over time.
- Yes
- No
- N/A
- Unknown/Unable to Code

Data are collected on the reliability or inter-observer agreement associated with each dependent variable, and IOA levels meet minimal standards.
- Yes
- No
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence: □3 □2 □1 □0

**Independent Variable**

Independent variable is described with replicable precision.
- Yes
- No
- N/A
- Unknown/Unable to Code

Independent variable is systematically manipulated and under the control of the experimenter.
- Yes
- No
- N/A
- Unknown/Unable to Code

Overt measurement of the fidelity of implementation for the independent variable is highly desirable.
- Yes
Baseline

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

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

One baseline measurement was retrieved from the school’s archival reading data.

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

Baseline conditions are described with replicable precision.

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

Overall Rating of Evidence: 3 2 1 0

Experimental Control/internal Validity

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

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

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

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

The results document a pattern that demonstrates experimental control.

☐ Yes
☐ No
External Validity

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

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

Overall Rating of Evidence: 3 2 1 0

Social Validity

The dependent variable is socially important.

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

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

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

Implementation of the independent variable is practical and cost effective

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

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

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

Overall Rating of Evidence: 3 2 1 0
Average WoE A across the 7 judgement areas:
Sum of $X / N = 17/7 = 2.43$
$X$ = individual quality rating for each judgement area
$N$ = number of judgement areas
Overall Rating of Evidence: $\square 3 \square 2 \square 1 \square 0$