Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Learning

How effective are computer-based interventions in schools for improving mathematical skills in children with dyscalculia?

Summary

This systematic literature review examined computer-based interventions that have been implemented in schools to evaluate how effective they are for improving the mathematical skills of children with dyscalculia and similar specific mathematical learning difficulties.

The review found that there are a number of computer-based interventions that have been shown to be effective for children with dyscalculia and that Educational Psychologists (EPs) can use and support the use of such interventions in school as part of their strategy to help such children improve their mathematical skills. Computer-based tasks and computer-based games designed to improve mathematical skills in children in school were reviewed. Computer games were identified as more effective than computer-based tasks, which may be due to features that make them more engaging and motivating.

Introduction

Mathematical skills are important for children and adults both in school and in everyday life (Parsons & Bynner, 2005). Some children and young people struggle with developing their mathematical skills and knowledge. There can be many reasons for this, such as limited educational opportunities, learning difficulties or social and communication difficulties. In order to assess and support a child who is
having difficulty learning or attaining in maths, an EP, teacher or other professional needs to investigate the key factors.

Developmental dyscalculia (here called dyscalculia) is the name given to a specific learning difficulty that affects the development of mathematical skills and knowledge. International diagnostic criteria for developmental dyscalculia define it as a domain-specific learning disorder that emerges at an early stage of development and cannot be explained by inappropriate schooling or deficient learning opportunities (Kucian & von Aster, 2015; Soares & Patel, 2015) and it is a neurodevelopmental condition (Szűcs & Goswami, 2013). The estimated prevalence rate of dyscalculia is between 3% and 6% (Kaufmann & von Aster, 2012; Kucian & von Aster, 2015). There is some debate about the utility of the concept of dyscalculia in Educational Psychology as there are concerns about the effect of labelling (or not labelling) a person and the possibility of missing children with mathematical difficulties because they do not meet diagnostic criteria (e.g. Gross, 2007). However, increasing understanding of the needs of people with dyscalculia can inform knowledge and practice. This can help assessment and intervention planning for children and young people identified as having mathematical difficulties that are not the result of comorbid difficulties such as dyslexia or social and communication needs.

Research into mathematical difficulties and interventions has increased in recent years, but is still a relatively young field compared to the fields of reading, writing and language skills. For example, despite having similar serious negative implications to its literacy counterpart, dyslexia (Beddington et al., 2008), dyscalculia has been researched far less (Bishop, 2010; Murphy, Mazzocco, Hanich, & Early, 2007). Reviews have evaluated interventions for mathematics from different perspectives and using different techniques such as meta-analyses, focusing on particular age
ranges and focusing on children with special educational needs or other difficulties such as emotional and behavioural disorders (Chodura, Kuhn, & Holling, 2015; Kroesbergen & Van Luit, 2003; Kunsch, Jitendra, & Sood, 2007; Maccini, Mulcahy, & Wilson, 2007; Myers, Wang, Brownell, & Gagnon, 2015; Templeton, Neel, & Blood, 2008). Other reviews were not peer-reviewed and have had time pass since their writing during which further interventions have been developed (e.g. Dowker, 2009).

The area of computer-based educational tools and interventions is also a growing field. Technology such as computers and tablets and apps and games are a fact of modern life and there are ways that they can be used effectively to learn (Bavelier, Green, & Dye, 2010). Computer-based training has been shown to be effective for areas such as language skills, auditory attention, working memory and reading skills (Loosli, Buschkuehl, Perrig, & Jaeggi, 2012; Stevens, Fanning, Coch, Sanders, & Neville, 2008). The use of computer technology has been shown to be not only effective in education in general, but in particular effective in mathematics, along with being effective for motivating students to engage in learning (Hasselbring, 1986; Li & Ma, 2010). An emerging area of psychological research is the impact of computer games on learning and cognitive processes, with research indicating that computer games of varying forms have a positive effective on cognitive processes and learning (Eichenbaum, Bavelier, & Green, 2014; Green & Bavelier, 2015; Kühn, Gleich, Lorenz, Lindenberger, & Gallinat, 2014), including, for example, helping children with dyslexia (Bavelier, Green, & Seidenberg, 2013).

Our understanding of how to use technology to develop skills can be combined with our psychological and neuroscientific understanding of conditions such as dyscalculia (e.g. there is a core deficit in numerosity processing) to develop effective pedagogical and educational interventions (Butterworth & Laurillard, 2010).
Computer-based methods may be particularly effective for mathematical learning because concepts can be demonstrated through this medium effectively, such as interactive physical motion, they provide the opportunity for lots of practice, can be adapted and can be more motivating (Butterworth & Laurillard, 2010; Kadosh, Dowker, Heine, Kaufmann, & Kucian, 2013). Computer games can motivate children to participate and learn, including in mathematical learning (Bottino, Ferlino, Ott, & Tavella, 2007; Ke, 2008; Papastergiou, 2009), which may help address the anxiety that children with difficulties such as dyscalculia can develop (Rubinsten & Tannock, 2010). Focusing on one area of difficulty, such as dyscalculia, may be advantageous because research has shown that in evaluating the effectiveness of computerised training interventions the factor of comorbidity of other conditions is important for the outcomes (Söderqvist, Nutley, Ottersen, Grill, & Klingberg, 2012).

Understanding computer-based interventions is important for EPs because there are many interventions on offer to schools and professionals at a time when schools have to make decisions about how to spend limited resources. An EP working in a school may want to evaluate whether the resources a school already has would be appropriate for children that they are assessing or may be involved in the process of selecting a new intervention. Evidence such as that found in this review can inform this.

The question that this systematic review of the literature is aimed to address is:

*How effective are computer-based interventions in schools for improving mathematical skills in children with dyscalculia?*

This review evaluates the effectiveness of computer-based interventions that have been implemented in schools and are aimed at improving the mathematical ability
and attainment of children with specific mathematical learning difficulties, i.e. dyscalculia. Only evidence that has been demonstrated to work in a school setting will be evaluated. In addition, this review focuses on interventions that have been implemented with children with dyscalculia or similar specific mathematical difficulties, rather than typically developing children or those with comorbid conditions such as dyslexia or autistic spectrum conditions, in order to identify interventions that are likely to be addressing the mathematical learning needs of the children involved.

**Critical Review of the Evidence Base**

**Systematic Literature Search**

A systematic search of the literature was conducted. Figure 1 shows the search and exclusion process and Table 1 describes the inclusion and exclusion criteria used to identify the studies for the review. The PsycINFO and the Education Resource Information Center (ERIC) databases were searched using OVID and EBSCOHost in February 2016. Search terms were inputted and combined as shown in Figure 1. As part of the preliminary stage of the exclusion process, database tools were used to limit and filter the search results to peer reviewed journals and English language studies only. The results from the two searches were combined and duplicates removed resulting in 239 articles in total. At stage one of the exclusion process, the articles were screened by title and abstract according to the inclusion and exclusion criteria, resulting in the exclusion of 164 studies. At stage two of the exclusion process the full texts of the remaining 75 studies were examined in accordance with the inclusion and exclusion criteria. Seventy studies were excluded at this stage, with five studies remaining as the focus of this review (see Table 2). Appendix A lists the excluded studies.
Figure 1. Flow diagram of the search and exclusion process
<table>
<thead>
<tr>
<th>Study feature</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Type of publication</td>
<td>Study published in a peer-reviewed journal</td>
<td>Study comes from any source that is not a peer-reviewed journal</td>
<td>Research published in peer-reviewed journals has been evaluated by expert reviewers and has met quality standards</td>
</tr>
<tr>
<td>2 Language of study</td>
<td>Entire study report is available in the English language</td>
<td>All or part of the study report is not available in the English language</td>
<td>So that all of the study can be evaluated as translation services are unavailable</td>
</tr>
<tr>
<td>3 Participants / population</td>
<td>a) Study includes children aged between 4 – 16 years as participants</td>
<td>a) All participants are adults or children aged below 4 years or over 16 years, or a combination of these</td>
<td>a) This review is examining interventions for school-aged children</td>
</tr>
<tr>
<td></td>
<td>b) Participants have specific mathematical difficulties in line with dyscalculia and are not identified as having other comorbid difficulties</td>
<td>b) Participants are described as having conditions or difficulties other than or in addition to mathematical difficulties</td>
<td>b) This review is examining interventions for children with dyscalculia</td>
</tr>
<tr>
<td>4 Intervention</td>
<td>The study includes an intervention that: a) is being used with the aim of improving mathematical outcomes</td>
<td>The study does not include an intervention that: a) is being used with the aim of improving mathematical outcomes</td>
<td>This review is examining interventions that: a) are designed to improve mathematical outcomes</td>
</tr>
</tbody>
</table>
## Study feature

<table>
<thead>
<tr>
<th>Study feature</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>b) has a computer-based focus</td>
<td>c) is implemented for at least 3 weeks</td>
<td>b) has little or no computer-based focus (e.g. tutoring, small group work)</td>
<td>b) are computer-based</td>
</tr>
<tr>
<td>b) has little or no computer-based focus (e.g. tutoring, small group work)</td>
<td>c) is implemented for less than 3 weeks</td>
<td>c) would be implemented for a moderate amount of time</td>
<td></td>
</tr>
</tbody>
</table>

### Outcomes

5 | Outcomes | The study measures a mathematical achievement, attainment or skill outcome as a dependent variable | The study does not measure any mathematical achievement, attainment or skill outcomes as a dependent variable | This review is looking at interventions that affect mathematical skills |

6 | Study setting | At least part of the study was conducted in a school | No part of the study was conducted in a school | This review is evaluating interventions that have been shown to be effective in schools |

7 | Type of study | The study contains primary empirical data | The study does not contain primary empirical data (e.g. reviews, commentaries) | This review is examining primary evidence, not evidence that has been reviewed by others or is not based on research |
Table 2
The five studies included in this review

<table>
<thead>
<tr>
<th>Full reference of study</th>
</tr>
</thead>
</table>

**Mapping the Field**

The five studies that were identified through the systematic literature search described computer-based interventions that had been implemented in schools with children who had mathematical difficulties. Table 3 shows the key features of each study.
<table>
<thead>
<tr>
<th>Country</th>
<th>Age (yrs)</th>
<th>Identification of mathematical difficulties</th>
<th>Intervention name</th>
<th>Intervention description</th>
<th>Comparison condition</th>
<th>Intervention intensity and duration</th>
<th>Outcome measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns,</td>
<td>USA</td>
<td>8–10</td>
<td>Low score on</td>
<td>Math Facts in a Flash</td>
<td>Children who used the intervention for less than once per week for 8–15 weeks</td>
<td>5–15 minutes, three times per week for 8–15 weeks</td>
<td>Star Math assessment (Renaissance Learning; a computer-adapted assessment system)</td>
</tr>
<tr>
<td>Kanive &amp;</td>
<td></td>
<td></td>
<td>assessment</td>
<td>(Renaissance Learning)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeGrande</td>
<td></td>
<td></td>
<td></td>
<td>computer-based</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2012)</td>
<td></td>
<td></td>
<td></td>
<td>mathematical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>fluency intervention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Castro,</td>
<td>Brazil</td>
<td>7-10</td>
<td>Met criteria for</td>
<td>Tom’s Rescue</td>
<td>Reinforcement using traditional teaching techniques covering the same skills as the intervention</td>
<td>60 mins, 2 times per week for 5 weeks</td>
<td>Arithmetic test within the Scholastic Performance Test (SPT), a standardised test used in Brazilian schools</td>
</tr>
<tr>
<td>Bissaco,</td>
<td></td>
<td></td>
<td>dyscalculia</td>
<td>Virtual environment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pancccioni,</td>
<td></td>
<td></td>
<td></td>
<td>incorporating 18 educational computer games</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodrigues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&amp; Domingue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leh &amp;</td>
<td>USA</td>
<td>8–9</td>
<td>Low score on</td>
<td>GO Solve Word Problems</td>
<td>Teacher-mediated instruction that was similar to the intervention for instructional practice, problem types and</td>
<td>50 mins, 2-3 times per week for 6 weeks (15 sessions total)</td>
<td>Researcher-developed word problem-solving test and state-level administered school assessment of mathematics</td>
</tr>
<tr>
<td>Jitendra</td>
<td></td>
<td></td>
<td>standardised</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2013)</td>
<td></td>
<td></td>
<td>school mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Age (yrs)</td>
<td>Identification of mathematical difficulties</td>
<td>Intervention name</td>
<td>Intervention description</td>
<td>Comparison condition</td>
<td>Intervention intensity and duration</td>
<td>Outcome measures</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>-----------------------------------------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>---------------------</td>
<td>-------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Finland</td>
<td>6–7</td>
<td>Scored below the tenth percentile for verbal counting</td>
<td>GraphoGame Math; Number Race</td>
<td>Educational computer games</td>
<td>N / a</td>
<td>10–15 minute sessions, 12-15 occasions in a 3 week period</td>
<td>Cognitive skills measures (corti block task, nonword repetition task) and early number skills measures (verbal counting, object counting, basic arithmetic)</td>
</tr>
<tr>
<td>Salminen, Koponen, Räsänen &amp; Aro (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>7–9</td>
<td>Teacher report and WISC-III assessment</td>
<td>The Number Race</td>
<td>Adaptive computer game for remediation of dyscalculia</td>
<td>N / a</td>
<td>30 minutes, 4 days per week for 4 weeks</td>
<td>Computerised testing battery (enumeration, symbolic and non-symbolic numerical comparison, addition, subtraction) and three subtests from the TEDI-MATH battery (counting, transcoding, base-10 comprehension)</td>
</tr>
<tr>
<td>Wilson, Revkin, Cohen, Cohen &amp; Dehaene (2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Critical Evaluation

The evidence that was identified through the systematic literature review was evaluated and compared with regard to multiple factors using a weight of evidence approach (Gough, 2007).

The studies were assigned ratings that were calculated using sets of criteria that were weighted with regard to their relevance to the review question. Three sets of ratings were generated: weight of evidence A (WoE A), weight of evidence B (WoE B) and weight of evidence C (WoE C). These were combined to create the superordinate evaluation rating of weight of evidence D (WoE D).

WoE A evaluated the methodological quality of the studies. This was judged using published coding protocols for research. Four studies were evaluated using the Gersten et al. (2005) coding protocol because they had group experimental or quasi-experimental designs, and one study was evaluated using the Horner et al. (2005) coding protocol because the research design was single-group. Appendix B specifies how the coding protocol criteria were converted into WoE A ratings and Appendix C contains examples of the two coding protocols used.

WoE B evaluated how appropriate the forms of evidence from the studies were for answering the research question in this review, for example by considering the research design. To evaluate and quantify how well-suited the type of evidence was for this, a typology of evidence criteria was applied (Petticrew & Roberts, 2003; details in Appendix B).

WoE C examined how relevant the focus of the studies were for answering the question of the review (criteria details are in Appendix B). All the weightings ranged
from 1-3. For the purposes of this review, ratings of 1-1.6 are considered “low”, 1.7-2.3 “medium” and ratings of 2.4-3 are considered “high”.

The three weight of evidence ratings were combined with equal weighting to create the superordinate rating of weight of evidence D (WoE D). Table 4 contains these ratings.

Table 4

<table>
<thead>
<tr>
<th>Study</th>
<th>WoE A</th>
<th>WoE B</th>
<th>WoE C</th>
<th>WoE D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns, Kanive &amp; DeGrande (2012)</td>
<td>1</td>
<td>1</td>
<td>2.3</td>
<td>1.4</td>
</tr>
<tr>
<td>de Castro, Bissaco, Pancioni, Rodrigues &amp; Domingues (2014)</td>
<td>3</td>
<td>2</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Leh &amp; Jitendra (2013)</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>Salminen, Koponen, Räsänen &amp; Aro (2015)</td>
<td>3</td>
<td>2</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Wilson, Revkin, Cohen, Cohen &amp; Dehaene (2006)</td>
<td>2</td>
<td>2</td>
<td>2.5</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Note: Ratings between 1 – 1.6 are “low”, 1.7 – 2.3 are “medium” and 2.4 – 3 are “high”
<table>
<thead>
<tr>
<th>Study</th>
<th>Comparison relevant to this review</th>
<th>Allocation to groups</th>
<th>Design¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns, Kanive &amp; DeGrande (2012)</td>
<td>Family group 1: Used intervention more than once per week</td>
<td>Non-random</td>
<td>NR O₁ X O₂</td>
</tr>
<tr>
<td></td>
<td>Family group 2: Used intervention less than once per week</td>
<td></td>
<td>NR O₁ O₂</td>
</tr>
<tr>
<td></td>
<td>Family group 2: Traditional teaching of the same skills</td>
<td></td>
<td>R O₁ Y O₂</td>
</tr>
<tr>
<td></td>
<td>Family group 2: Teacher-mediated intervention teaching the same skills</td>
<td></td>
<td>R O₁ Y O₂</td>
</tr>
<tr>
<td>Salminen, Koponen, Räsänen &amp; Aro (2015)</td>
<td>1 GraphoGame Math Intervention (pre intervention score)</td>
<td>Random</td>
<td>O₁ X O₂</td>
</tr>
<tr>
<td></td>
<td>2 Number Race Intervention (pre intervention score)</td>
<td></td>
<td>O₁ X O₂</td>
</tr>
<tr>
<td>Wilson, Revkin, Cohen, Cohen &amp; Dehaene</td>
<td>Family group 1: Number Race Intervention (pre intervention score)</td>
<td>-</td>
<td>O₁ X O₂</td>
</tr>
<tr>
<td>(2006)</td>
<td>Family group 2: Number Race Intervention (post intervention score)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹Based on research designs as describe by Barker, Pistrang and Elliot (2002)
Effect sizes were calculated for the findings of the five studies that were relevant to the review question (Table 6). For the three studies that used group comparison of means (Burns, Kanive, & DeGrande, 2012; de Castro, Bissaco, Panccioni, Rodrigues, & Domingues, 2014; Leh & Jitendra, 2013), Hedge’s $g$ was calculated (Hedges, 1981). This is because all of these studies demonstrated that the pre-intervention tests showed similar scores on the dependent variable between intervention and comparison groups. Hedge’s $g$ was calculated using the following formula:

$$
\text{Hedge’s } g = \frac{\text{mean}_{\text{post}} - \text{mean}_{\text{pre}}}{\text{SD}_{\text{pooled}}}
$$

$$
\text{SD}_{\text{pooled}} = \sqrt{\frac{(N_A - 1) \times SD_A^2 + (N_B - 1) \times SD_B^2}{N_A + N_B - 2}}
$$

As the Wilson et al. (2006) study used a small, one-group pre-test, post-test design, Becker’s standardised mean change was calculated (Becker, 1988) using the following formula:

$$
\text{SMD} = \frac{\text{mean}_{\text{post}} - \text{mean}_{\text{pre}}}{\text{SD}_{\text{pre}}}
$$

This was calculated where there was sufficient information to do so. For some other effects, the Campbell Collaboration Effect Size Calculator was used to calculate Cohen’s d.

Becker’s standardised mean change was also calculated for the findings of the Salminen et al. (2015) study because the findings were based on pre- and post-testing of the two small groups.
Cohen’s interpretation of effect sizes, where 0.2 is small, 0.5 is medium and 0.8 is large can be used to evaluate the effect sizes (Hedge’s $g$, Becker’s standardised mean change and Cohen’s $d$) found here (Cohen, 1988).

The effect sizes were calculated with the means ordered so that in Table 6 a positive effect size indicates an improvement due to the intervention and a negative effect size indicates change in the other direction except for in the case of timed tasks, where a reduction in time is associated with improvement and therefore indicated by a negative number.

Table 7 shows the findings, effect size range and WoE D for each of the five studies included in this review.
### Table 6

**Effect sizes of the intervention effects for maths outcomes in the five studies in this review**

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Measure</th>
<th>Hedge’s g</th>
<th>Becker’s SMD</th>
<th>Cohen’s d</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns, Kanive &amp; DeGrande (2012)</td>
<td>442</td>
<td>Star Math assessment score (3rd grade children)</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Star Math assessment score (4th grade children)</td>
<td></td>
<td>0.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Castro, Bissaco, Panccioni, Rodrigues &amp; Domingues (2014)</td>
<td>26</td>
<td>SPT arithmetic test</td>
<td></td>
<td>1.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leh &amp; Jitendra (2013)</td>
<td>25</td>
<td>Word problem solving</td>
<td></td>
<td>-0.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PSSA math assessment score</td>
<td></td>
<td>-0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salminen, Koponen, Räsänen &amp; Aro (2015)</td>
<td>17</td>
<td>GraphoGame Math Verbal counting</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Object counting subitizing (seconds)</td>
<td></td>
<td>-0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Object counting dot counting (seconds)</td>
<td></td>
<td>-1.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Basic arithmetic</td>
<td></td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number Race Verbal counting</td>
<td></td>
<td>1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>Measure</td>
<td>Hedge's $g$</td>
<td>Becker's SMD</td>
<td>Cohen's $d$</td>
<td>Confidence interval</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>-------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>Wilson, Revkin, Cohen, Cohen &amp; Dehaene (2006)</td>
<td>Object counting subitizing (seconds)</td>
<td>-0.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Object counting dot counting (seconds)</td>
<td>-0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic arithmetic</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Addition accuracy (tie)</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Addition accuracy (small)</td>
<td>-0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Addition accuracy (large)</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtraction reaction time (rule)</td>
<td>-0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtraction reaction time (small)</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtraction reaction time (large)</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TEDImath counting</td>
<td>1.49</td>
<td>0.02-2.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TEDImath transcoding</td>
<td>1.61</td>
<td>0.11-3.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Study findings</td>
<td>Effect size range¹</td>
<td>WoE D²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burns, Kanive &amp; DeGrande (2012)</td>
<td>Intervention group scored significantly higher on post test than comparison group</td>
<td>0.36 – 0.49</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Castro, Bissaco, Panccioni, Rodrigues &amp; Domingues (2014)</td>
<td>Intervention group scored significantly higher on post test than comparison group</td>
<td>1.45</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leh &amp; Jitendra (2013)</td>
<td>No significant difference between the post test scores of the experimental and comparison groups</td>
<td>-0.35 – -0.36</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salminen, Koponen, Räsänen &amp; Aro (2015)</td>
<td>Verbal counting and dot counting fluency improved with GraphoGame Math and basic arithmetic improved with Number Race</td>
<td>0.17 – 1.04</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wilson, Revkin, Cohen, Cohen &amp; Dehaene (2006)</td>
<td>Improvement on a number of tasks testing mathematical skills</td>
<td>0.1 – 1.61</td>
<td>2.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ 0.2 = small, 0.5 = medium, 0.8 = large (Cohen, 1988)
² For WoE D, 1 - 1.6 = “low”, 1.7 - 2.3 = “medium”, 2.4 - 3 = “high”

**Interventions**

Three of the studies had interventions that consisted of or had many elements that consisted of computer games (de Castro et al., 2014; Salminen et al., 2015; Wilson,
Revkin, et al., 2006), while the other two interventions involved computer-based tasks (Burns et al., 2012; Leh & Jitendra, 2013). The computer-based task interventions involved working through mathematical problems and tasks on a computer while the computer-based game interventions included activities that were designed to be fun, engaging and entertaining with elements of competition, user choice and increased interactivity. The same game, Number Race, appeared in all three of the studies that involved computer games. This suggests that Number Race can be an effective computer game intervention for children with dyscalculia (Wilson, Revkin, et al., 2006; Wilson, Dehaene, et al., 2006) as all three studies found a positive effect of this intervention on mathematical skill of the participants.

Of the two interventions that involved computer-based tasks, one found no effect as compared to having teacher-mediated instruction (Leh & Jitendra, 2013), suggesting that possibly an important factor in the effectiveness and additional value of computer-based interventions as opposed to teacher-mediated interventions may be whether it is motivating and engaging for the child, which is in line with the concept of games being effective at least partly because they are engaging (Eichenbaum et al., 2014).

**Intervention focus**

There was a difference in the skills targeted by the interventions of the studies evaluated here, with the two computer-based task interventions targeting higher order skills such as word problem-solving and mathematical fact knowledge (Burns et al., 2012; Leh & Jitendra, 2013), and the three computer-based game interventions being aimed at more foundational skills such as counting and understanding numerosity (de Castro et al., 2014; Salminen et al., 2015; Wilson,
Revkin, et al., 2006). These factors could interact with the finding of one study that the intervention was not effective as compared to the comparison group (Leh & Jitendra, 2013) and the methodological weakness of the evidence of the other study (Burns et al., 2012) which had a very low WoE A rating. This makes it difficult to assess whether intervention focus was a possible determining factor in effectiveness of the intervention, but this is a possibility that must be considered.

Sample

The age range of the participants in the studies reviewed here is limited in that it ranges from 6 to 10 years. This means that the findings are not immediately generalizable to children outside of this age range.

The participants all had some mathematical difficulties, but some studies specified the nature of the participants’ difficulties and assessed whether they had other needs in more detail than other studies. For example, de Castro et al. (2014) used dyscalculia criteria, assessment scores, teacher ratings, school records and interviews to identify children that had mathematical learning difficulties and did not have comorbid conditions such as dyslexia or attention deficit hyperactivity disorder (ADHD) or frequent absences. This contributed to the high WoE C rating for this study. In contrast, two studies identified participants only as those who scored below a particular cut-off point in a mathematics assessment (Burns et al., 2012; Leh & Jitendra, 2013). Occupying the middle ground for participants’ needs identification are the studies that used some form of dyscalculia-related screening (Wilson, Revkin, et al., 2006) and psychological theory (verbal counting; Salminen et al., 2015) to identify the participants.
The two studies that demonstrated weaker evidence of intervention efficacy were conducted with slightly older children than the other interventions. It could be that computer-based interventions may be better suited to younger children. However, these interventions were also task- rather than game-focused, which brings about further confounding factors such as the possibility that game-focused interventions may be more engaging and motivating than task-focused interventions.

**Setting**

None of the interventions were evaluated in a UK school setting, however all five of the studies implemented the interventions in school settings. Arguably, the findings of the studies conducted in the USA, Finland and France (Burns et al., 2012; de Castro et al., 2014; Salminen et al., 2015; Wilson, Revkin, et al., 2006) are most generalizable to UK school settings due to the similarities between the respective education systems and countries. The UK, USA, Finland and France are all members of the Organisation for Economic Co-operation and Development (OECD), meaning that they have democratic and economic features in common as compared to Brazil, which is not an OECD country.

Four of the studies described intervention implementation in enough detail to ascertain that the interventions took place in the school setting in a way that could be replicable by schools who wish to use the intervention in the future without the requirement of excessive additional resources. Burns et al. (2012) were unable to provide much detail about how the interventions were implemented in schools as the researchers were not able to observe the children and teachers in the school settings, but the study did provide some data about school information and
intervention use time, suggesting that the interventions were implemented in the schools in a typical way.

**Study design**

Table 5 gives an overview of the research designs used in the five studies. Two studies used an experimental design involving a control group and random allocation (de Castro et al., 2014; Leh & Jitendra, 2013). Interestingly, Leh and Jitendra (2013) did not find an effect of their intervention as compared to the comparison group that has been characterised as the control group for the purposes of this review. Of the three studies that specified random allocation of participants to groups, none of them detailed how this was carried out, making it difficult to evaluate how much the research design adhered to an experimental design. In addition, other features were not controlled for that could have impacted the findings of the studies such as adults being aware of intervention conditions and therefore having expectancy effects.

One study in particular has significant issues in its research design. Burns et al. (2012) consistently characterise their research design in terms that suggest it is a quasi-experimental design. However, it could be argued that the research was more correlational in nature because their control group were children that had used the intervention computer program for less than once per week over the intervention period. It seems that they did not allocate the children to groups (intervention and control) but rather assigned group designations to the children that had already either accessed the intervention or had not. This means that there could be a wide range of reasons that the mathematical skills of the children in the control group did
not improve as much as those of the children in the intervention group such as poor school attendance or behavioural problems. This is why this study received a low WoE B rating.

All of the studies used dependent variable measures that are relevant to evaluating whether it seems likely that the intervention will affect mathematical skills outcomes that are of importance to children and educators in schools. Some studies used a wide battery of measures such as Wilson et al. (2006), which is partially explained by the fact that the study was an open trial of a small sample, while other studies limited the dependent variable to school assessment outcomes (e.g. de Castro et al., 2014). Leh and Jitendra (2013) used a researcher-developed measure of word problem-solving ability, which could be considered to be focused on skills too similar to those being targeted by the intervention, however they balanced this with a school-based assessment measure, which supports the methodological rigour and relevance to this review question of the study design.

According to the guidance set out by Cohen (1992), four of the five studies in this review (de Castro et al., 2014; Leh & Jitendra, 2013; Salminen et al., 2015; Wilson et al., 2006) were underpowered to find effects that exist because the sample size was too small. Cohen’s (1992) guidance suggests group sizes of at least 26 to identify a large effect, 64 for a medium effect and 393 for a small effect for designs comparing two means at the .05 significance level and with power of 0.8. This means that for the study that found no significant improvement from the intervention (Leh & Jitendra, 2013), we can conclude only that there was not enough power to find an effect that may have been there rather than confidently concluding that there is no effect to be found (Altman & Bland, 1995). The fact that three of these four underpowered studies found significant improvement with two having a small
(Salminen et al., 2015; Wilson et al., 2006) and one having a large (de Castro et al., 2014) effect is surprising given that they were underpowered. Underpowered studies such as these can lead to inconsistent findings across the literature (Maxwell, 2004). One study in the review (Burns et al., 2012) had a large enough sample to have sufficient power to find an effect.

**Overall findings**

Overall, of the computer-based interventions reviewed here, the computer-based games studies provide stronger evidence of effectiveness for the group of children and intervention setting being evaluated in this review. This may be due to methodological differences, in that, for example, the Burns et al. (2012) study had significant methodological weaknesses (resulting in it being the only study with a low WoE D rating) and the other computer-based task intervention (Leh & Jitendra, 2013) found no significant effect of the intervention (however it was likely underpowered). The studies were able to provide evidence regarding interventions that can be implemented in schools and for children with dyscalculia or similar specific mathematical learning difficulties.

**Conclusion and Recommendations**

The findings of this review are that there are computer-based interventions that have been designed with the aim of improving mathematical skills in children with dyscalculia and specific mathematical learning difficulties. These interventions have been implemented in school settings using research designs that vary in strength and the findings suggest that some of them are effective in improving the mathematical skills of children with dyscalculia when implemented in schools for a
moderate amount of time. The interventions that had the strongest evidence to support the positive effect that they appear to have were Number Race (de Castro et al., 2014; Salminen et al., 2015; Wilson, Revkin, et al., 2006), GraphoGame Math (Salminen et al., 2015) and the virtual environment, Tom’s Rescue, that incorporated multiple games from multiple sources (de Castro et al., 2014). From the evidence evaluated in this review, computer games have a stronger evidence base for being an effective intervention for children with dyscalculia than computer-based interventions that use tasks, but this is because one study found no difference between the intervention and a similar teacher-mediated intervention (Leh & Jitendra, 2013), and the other study found evidence that is weak methodologically making it difficult to draw conclusions from it (Burns et al., 2012).

Recommendations for EPs are that there are effective interventions available for improving mathematical skills of children with dyscalculia and similar specific mathematical learning difficulties. There are a promising number of computer-based game interventions that have been developed, and EPs can advise teachers and parents that may be concerned about the negative reputation of computer games that research shows there are benefits that come from using appropriate computer games.

There was a difference in the age and focus of the interventions that were shown to be more effective with stronger evidence in this review. The interventions that were less effective and had weaker evidence also targeted more advanced skills in a slightly older age range of children than the other interventions. This suggests that it is possible that the types of interventions being evaluated in this review are more well-suited to developing the foundational skills such as counting and a sense of numerosity as targeted by the computer-based game interventions for younger
children. Therefore in evaluating the interventions here, age and intervention skill focus are potential confounding variables that could be delineated in future research.

Some limitations of this review are that only a small subset of computer-based interventions for mathematics were evaluated because there was a focus on evidence of efficacy in children with dyscalculia. It would be worthwhile to conduct a similar review that considers a wider range of computer-based interventions in a more diverse sample of children, such as including typically developing children, to begin to evaluate whether findings from such studies could be applied to the set of children that were the focus of this review.

In addition, this review is only able to generalise to a small subset of ages based on the evidence identified. More research into the efficacy of interventions for different age ranges would be beneficial. For example, the effects of interventions may be different according to age both because the specific learning needs may be different (such as comparing children under 4 years with adolescents) and also because engagement with and motivation to participate in interventions may be different in different aged children, such as children in secondary school as compared to primary school.

As identified in the introduction of this review, more research is needed to understand mathematical difficulties and develop effective interventions, and this is particularly the case for older children such as those at secondary school. Dyscalculia is a neurodevelopmental disorder (Bishop, 2010; Butterworth & Laurillard, 2010); it will not cease to affect a person as they transition through adolescence. Yet it is more important than ever for adolescents to achieve mathematics outcomes, such as with the UK education system requirement that
young people who do not gain a grade C in their GCSE Mathematics assessments being required to continue studying the subject (Department for Education, 2011), meaning that effective interventions for older children and young people are also needed. There is much scope for further research and reviews in the field of interventions for mathematical learning.
References and Appendices

References


Appendix A

Table A1

Articles excluded at full text screening stage (stage 2) with criteria reasons

<table>
<thead>
<tr>
<th>Article</th>
<th>Exclusion criteria number(s)</th>
</tr>
</thead>
</table>


---

Table A2

*Articles excluded at title / abstract screening stage (stage 1) with criteria reason*

<table>
<thead>
<tr>
<th>Article</th>
<th>Exclusion criteria number(s)</th>
</tr>
</thead>
</table>

*a* indicates that the report was inaccessible.


data. *Erdelyi Pszichologiai Szemle*, 16(1), 57-69.


students. *Journal Of Experimental Education*, 75(3), 221-244.


Addressing teacher shortages in disadvantaged schools: Lessons from two institute of education sciences studies. NCEE Evaluation Brief. NCEE 2013-4018. National Center For Education Evaluation And Regional Assistance,


<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Journal</th>
<th>Volume</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ostad, S. A.</td>
<td>Private speech use in arithmetical calculation: Relationship with phonological memory skills in children with and without mathematical difficulties.</td>
<td><em>Annals Of Dyslexia</em>, 65(2)</td>
<td>103-119.</td>
<td></td>
</tr>
<tr>
<td>Prevatt, F., Proctor, B., Swartz, S. L., &amp; Canto, A. I.</td>
<td>Profiles of academic achievement and cognitive processing in college students with foreign language difficulties.</td>
<td><em>Journal Of Postsecondary Education And Disability</em>, 16(2)</td>
<td>63-77.</td>
<td></td>
</tr>
<tr>
<td>Raches, D. &amp; Mazzocco, M. M. M.</td>
<td>Emergence and nature of mathematical difficulties in young children with Barth syndrome.</td>
<td><em>Journal</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

47
of Developmental and Behavioral Pediatrics, 33, 328-335.


<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Year</th>
<th>Title</th>
<th>Journal</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith, D. N.</td>
<td>1997</td>
<td>Independent mathematical modeling.</td>
<td>Teaching Mathematics And Its Applications, 16(3), 101-06.</td>
<td>4</td>
</tr>
<tr>
<td>Tropper, N., Leiss, D., &amp; Hänze, M.</td>
<td>2015</td>
<td>Teachers’ temporary support and worked-out examples as elements of scaffolding in</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>


## Appendix B

### Table B1

*Weight of evidence A criteria and rationale*

<table>
<thead>
<tr>
<th>WoE A Score</th>
<th>Criteria</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Meets ≥ 9 essential criteria and ≥ 4 desirable criteria</td>
<td>Based on recommendations of Gersten et al. (2005)</td>
</tr>
<tr>
<td>2</td>
<td>Meets ≥ 9 essential criteria and ≥ 2 desirable criteria</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Meets &lt; 9 essential criteria and &lt; 2 desirable criteria</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Average score of 2–3 across the seven judgement areas</td>
<td>Possible scores range from 0–3, this criteria converts them into scores between 1–3.</td>
</tr>
<tr>
<td>2</td>
<td>Average score of 1–1.9 across the seven judgement areas</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Average score of 0–0.9 across the seven judgement areas</td>
<td></td>
</tr>
</tbody>
</table>

### Table B2

*Weight of evidence B criteria and rationale*

<table>
<thead>
<tr>
<th>Weighting</th>
<th>Type of study design</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Qualitative research, surveys, case-control studies, non-experimental evaluations</td>
<td>These are the types of evidence, aside from systematic reviews, that are most suited to research questions about “effectiveness” (Petticrew &amp; Roberts, 2003)</td>
</tr>
<tr>
<td>2</td>
<td>Cohort studies, quasi-experimental designs</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Randomised control trial studies</td>
<td></td>
</tr>
<tr>
<td>Criteria</td>
<td>Weightings</td>
<td>Rationale</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>A Sample</strong></td>
<td>3</td>
<td>Specified as only having dyscalculia or mathematical difficulties</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Has dyscalculia or mathematical difficulties and does not specify whether has comorbid conditions</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Has or is likely to have other comorbid conditions in addition to dyscalculia or mathematical difficulties</td>
</tr>
<tr>
<td><strong>B Intervention implementation</strong></td>
<td>3</td>
<td>Intervention implemented by school staff in a school using school resources without additional costs or staff; implementation could be reproduced by school staff in a typical school</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Intervention implemented by specialist staff or using additional research resources, but could be implemented by school staff using school resources</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Intervention requires specialist staff and / or significant specialist resources such as specialist computing equipment or prohibitively expensive software</td>
</tr>
<tr>
<td><strong>C Measurement of mathematical skills improvement</strong></td>
<td>3</td>
<td>School testing outcomes or typical educational measures of numeracy / mathematical ability</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Outcomes that are more relevant to research and are not typically used in school but are reflective of skills and abilities tested in education settings</td>
</tr>
</tbody>
</table>

To identify interventions that help mathematical skills because they target mathematical difficulties.

An intervention is more likely to be used and more feasible if it can be implemented using typical school resources and is not dependent on specialist personnel and equipment.

To identify interventions that could typically be used to help children improve their mathematical skills so that they can succeed in their...
1. Measures that are specific to the study or are not reflective of educational attainment or required skills.

**D** School setting
3. OECD educational system

2. Similar to OECD educational system

1. Very different to OECD educational system

To identify interventions that could be implemented in an OECD educational system such as the UK educational system

**Note** Scores on A, B, C and D are averaged to create WoE C rating

### Table B4

*Weight of evidence C ratings for the five studies included in the review*

<table>
<thead>
<tr>
<th>Study</th>
<th>Criteria score</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A  B  C  D</td>
<td></td>
</tr>
<tr>
<td>Burns, Kanive &amp; DeGrande (2012)</td>
<td>1  3  2  3</td>
<td>2.3</td>
</tr>
<tr>
<td>de Castro, Bissaco, Panccioni, Rodrigues &amp; Domingues (2014)</td>
<td>3  2  3  2</td>
<td>2.5</td>
</tr>
<tr>
<td>Leh &amp; Jitendra (2013)</td>
<td>1  2  2  3</td>
<td>2</td>
</tr>
<tr>
<td>Salminen, Koponen, Räsänen &amp; Aro (2015)</td>
<td>2  3  2  3</td>
<td>2.5</td>
</tr>
<tr>
<td>Wilson, Revkin, Cohen, Cohen &amp; Dehaene (2006)</td>
<td>3  2  2  3</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Appendix C
Coding Protocol Examples

Name of Coder:

Date: February 2016

Full Study Reference:


Intervention Name (description of study): Virtual environment (Tom’s Rescue)

Research design: Group Experimental

Type of Publication: Peer-reviewed journal

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

**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?

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

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

☐ 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?
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:

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
X No
☐ N/A
☐ Unknown/Unable to Code

The process for selecting participants is described with operational precision.

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

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

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

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

Dependent Variable
Dependent variables are described with operational precision.

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

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

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

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

Overall Rating of Evidence: 3 X 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
- No
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence: 3 X 2 1 0
**Baseline**  
The majority of single-subject research studies will include a baseline phase that provides repealed 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
- [x] No
- [ ] N/A
- [ ] Unknown/Unable to Code

Baseline conditions are described with replicable precision.

- [ ] Yes
- [x] No
- [ ] N/A
- [ ] Unknown/Unable to Code

**Overall Rating of Evidence:**

- [ ] 3
- [ ] 2
- [ ] 1
- [x] 0

**Experimental Control/Internal Validity**  
The design provides at least three demonstrations of experimental effect at three different points in time.

- [ ] Yes
- [x] No
- [ ] N/A
- [ ] Unknown/Unable to Code

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

- [x] Yes
- [ ] No
- [ ] N/A
- [ ] Unknown/Unable to Code

The results document a pattern that demonstrates experimental control.

- [x] Yes
- [ ] No
- [ ] N/A
- [ ] Unknown/Unable to Code

**Overall Rating of Evidence:**

- [ ] 3
- [x] 2
- [ ] 1
- [ ] 0

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

- [ ] Yes
X No
□N/A
□Unknown/Unable to Code

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

Social Validity
The dependent variable is socially important.
X Yes
□ No
□N/A
□Unknown/Unable to Code

The magnitude of change in the dependent variable resulting from the intervention is socially important.
X Yes
□ No
□N/A
□Unknown/Unable to Code

Implementation of the independent variable is practical and cost effective
X 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.
X Yes
□ No
□N/A
□Unknown/Unable to Code

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

Average WoE A across the 7 judgement areas:
Sum of X / N = 9 / 7 = 1.3
X = individual quality rating for each judgement area
N = number of judgement areas
Overall Rating of Evidence: □ 3 X 2 □ 1 □ 0