Case Study 1: An Evidence-Based Practice Review Report

Theme: School/setting Based Interventions for Learning.

How effective are computer or tablet-based mathematical interventions in improving numeracy skills in young children?

Summary

The importance of numeracy has been widely acknowledged as the foundation for lifelong learning, which must be harnessed from an early age to support young people's success in the wider curriculum and other activities beyond the classroom (Ofsted, 2018). One tool to encourage numeracy skill development in young children is the use of computer or tablet-based mathematical interventions, using app technology to support teaching staff in creating an accessible and active learning environment. The current systematic literature review explored how effective computer or tablet-based mathematical interventions are in improving numeracy skills in young children. Five papers were selected for meeting the inclusion criteria and were evaluated against Gersten et al.'s (2005) coding protocol and Gough's (2007) weight of evidence framework to facilitate the critical review of the evidence. The studies focused on numeracy skill outcomes following the use of a computer or tablet-based mathematical intervention implemented in the classroom. Analysis of each paper revealed small and large effect sizes, pooled to give an overall small, not significant effect size in a meta-analysis. Conclusions about the effectiveness of computer or tablet-based mathematical interventions are limited, which is discussed alongside suggested areas for future research.

Introduction

Numeracy skill development

Numeracy is an important area of learning for young people, not only as a fundamental element of mathematics but also as a life skill that gives students a foundation to succeed in learning and access the wider curriculum (Education Scotland, 2019). Ofsted (2018) have emphasised the need to focus on numeracy skill development in young children, given the impact of early mathematical attainment on future achievement in school. This corroborates with previous research whereby early mathematical knowledge is a strong predictor of later school success (Claessens & Engel, 2013). Encompassing key education bodies and current research highlights the need to develop a strong foundation in maths as early as possible, to support children in reaching their full potential as they progress through their education.

Computer and tablet-based interventions

The use of technology and game-based learning in education has seen a large increase in recent years. Since 2017, there has been a total of $2.25 billion invested by game-based learning companies, more than double the year before (Adkins, 2018). The exponential growth in using such technologies for education suggests a need to explore the extent of its effectiveness in improving numeracy skills in the classroom.

Computer or tablet-based mathematical interventions are grounded in interactive applications that can be used to teach and develop numeracy skills in children. These applications, downloadable from the app store, are selected according to the specific skill taught, such as counting, number knowledge or transformation, delivered via tablets or computers, according to the technology available in the classroom. Within Early Years education, these may be integrated into the existing play-based environment. Of particular note is the research into tablets, highlighting the unique nature of the touchscreen technology that allows young children to practice skills in maths in a practical and engaging manner (Bray & Tangney, 2016).

Given that there are multiple applications available that provide individualised and creative ways of delivering mathematical knowledge and activities, there is diversity in its presentation. Each application is typically learner-centred, in that individuals can work through the app at their own pace which are designed as intuitively motivating for children. This includes the use of colourful pictures, videos and animations, consistent across applications, thus encouraging pupils to work independently on their activities.

Psychological Underpinnings

The Direct Instruction model (Engelmann, Becker, & Gersten, 2014), whereby children's behaviours evolve according to their environment, underpins the use of computer or tablet-based technologies in education. Through feedback, repetition and rewards from applications, children will continue to repeat behaviours that are followed by positive consequences, such as stimulating audio, reward charts or visuals in the app. This behaviourist model highlights how changing behaviour through reinforcement from app stimuli can encourage children to use educational technologies. An important element of Direct Instruction is that children are actively accessing the information (Magliaro, Lockee, & Burton, 2005), a key part of effective learning and education.

The use of computer or tablet-based mathematical interventions embody the principles of motivation theory. This argues for the importance of creating a learning environment that is stimulating and engaging for children so they can perform well and thrive in the classroom (Kaplan, Katz, & Flum, 2011). Within a technology-based learning environment, the self-determination theory of motivation allows learners to be creative and enjoy the activities, giving them autonomy and helping them feel connected to the applications (Ryan, Rigby, & Przybylski, 2006). Moreover, the tactile features of using computers or tablets can increase motivation, as learners can manipulate the size and rotation of shapes and use digital screens to move objects (Tran, Chen, William, Conley, & Dede, 2012).

Rationale and Relevance

Children may be vulnerable to underachievement throughout their education if they have not developed early mastery in mathematics (Jordan, Kaplan, Ramineni, & Locuniak, 2009). If the tools to support mathematical development are harnessed in young children, especially in their early years where they have shown the fastest rate of mathematics development (Hill, Bloom, Black, & Lipsey, 2008), this will support their numeracy skills in the future.

It has been reported that the use of interactive maths applications could be more valuable than non-interactive technology (National Association for Education of Young Children, 2012) as it incorporates modern advancements with education, creating effective and motivating learning aids in the classroom. Research emphasises the value of these multi-touch technology applications whereby young children demonstrate high performance in mathematics with limited exposure time (Moyer-Packenham& Bolyard, 2016). Combining research on children’s motivation and their high performance emphasises the need to evaluate the extent that technologies can transform numeracy skill teaching and learning in the classroom.

One of the key roles of Educational Psychologists (EPs) is to explore how children can learn best, using evidence-based practice to support teachers and adults working with children to improve their future outcomes. This is further reinforced by the HCPC (2016) guidelines that EPs are pivotal in critically appraising the research base to make it easily accessible to other adults working around the child. By supporting school staff with current interventions, EPs are well placed to support the creation of effective learning environments to meet the needs of young people. The use of computer and tablet-based interventions therefore have the potential to change the techniques for supporting numeracy skill development from a very young age.

Despite the growth of touch-screen tablet technologies in schools (Clarke & Svanaes, 2014), evidence arguing the concerns of technology based screen time on early childhood education is also emerging (Madigan, Browne, Racine, Mori, & Tough, 2019). In order to evaluate the impact of technologies in educational settings, a critical review of the research is needed to see the extent of its effectiveness in children's learning and attainment.

Review Question

'How effective are computer or tablet-based mathematical interventions in improving numeracy skills in young children?'

This review focuses on studies evaluating interventions led by teaching staff in the classroom, of high research methodology. This next section will detail the process of selecting appropriate studies, the critical review of the literature and analyse the effect sizes of the studies, including a meta-analysis. The research findings are discussed, and recommendations about the use of the intervention in practice and future research needs are drawn.

Critical Review of the Evidence

Search Procedure

A systematic search of the literature on Web of Science, PsychINFO and Education Resources Information Centre (ERIC) was conducted on 2nd January 2020. Web of Science was chosen as it covers a wide range of multidisciplinary scholarly databases, ERIC for its focus on Education and PsychINFO for its focus on psychology. The key concepts identified for searching are outlined in Table 1, in line with the review question focusing on exploring the effectiveness of computer or tablet based interventions on improving numeracy skills with children in a school environment. By combining the key concepts together, a comprehensive literature search was credible.

Table 1

Search terms for key concepts on Web of Science, PsychINFO and ERIC databases

Computer	Children	Maths	Intervention	School
computer OR internet OR online OR web OR website OR "web based" OR "computer based" OR "web-based" OR "computer-based" OR virtual OR technology	child OR children OR juveniles OR minors OR "school aged" OR "school-aged" OR young or youth OR infant OR "primary age"	math* OR arithmetic* OR number* OR numeracy	intervention* or program* or treatment* or therap* or strateg*	learn* OR school* OR educat* OR classroom* OR teach*

Note: the "OR" is used to combine terms to find papers that have any of the words e.g. "computer OR internet" finds all items mentioning either computer or internet. The quotation marks are used to search for an exact phrase .g. “computer based” to find variations of this phrase. The use of truncation (an asterisk) is used to search for differnet word endings, including plurals e.g. “math*” to find “math, maths, mathematics”. These key concepts were then combined using “AND” to find results which had all of these concepts e.g. “computer AND children AND maths”.

Inclusion and Exclusion Criteria

In conjunction with database searching, an ancestral search was conducted to ensure all relevant studies were found. Figure 1 illustrates a flow diagram of the study selection process. Two hundred and eighteen records, excluding duplicates, were first screened using titles and abstracts according to various inclusion and exclusion criteria (see Table 2). Following this, 31 full-text articles were assessed for their eligibility in the present study using the inclusion and exclusion criteria outlined in Table 2, determined by the review question set out for this review. Of these 31 studies, 26 studies did not meet the criteria for various reasons, detailed in Appendix A.

Figure 1

PRISMA Flow Diagram Illustrating Search Strategy

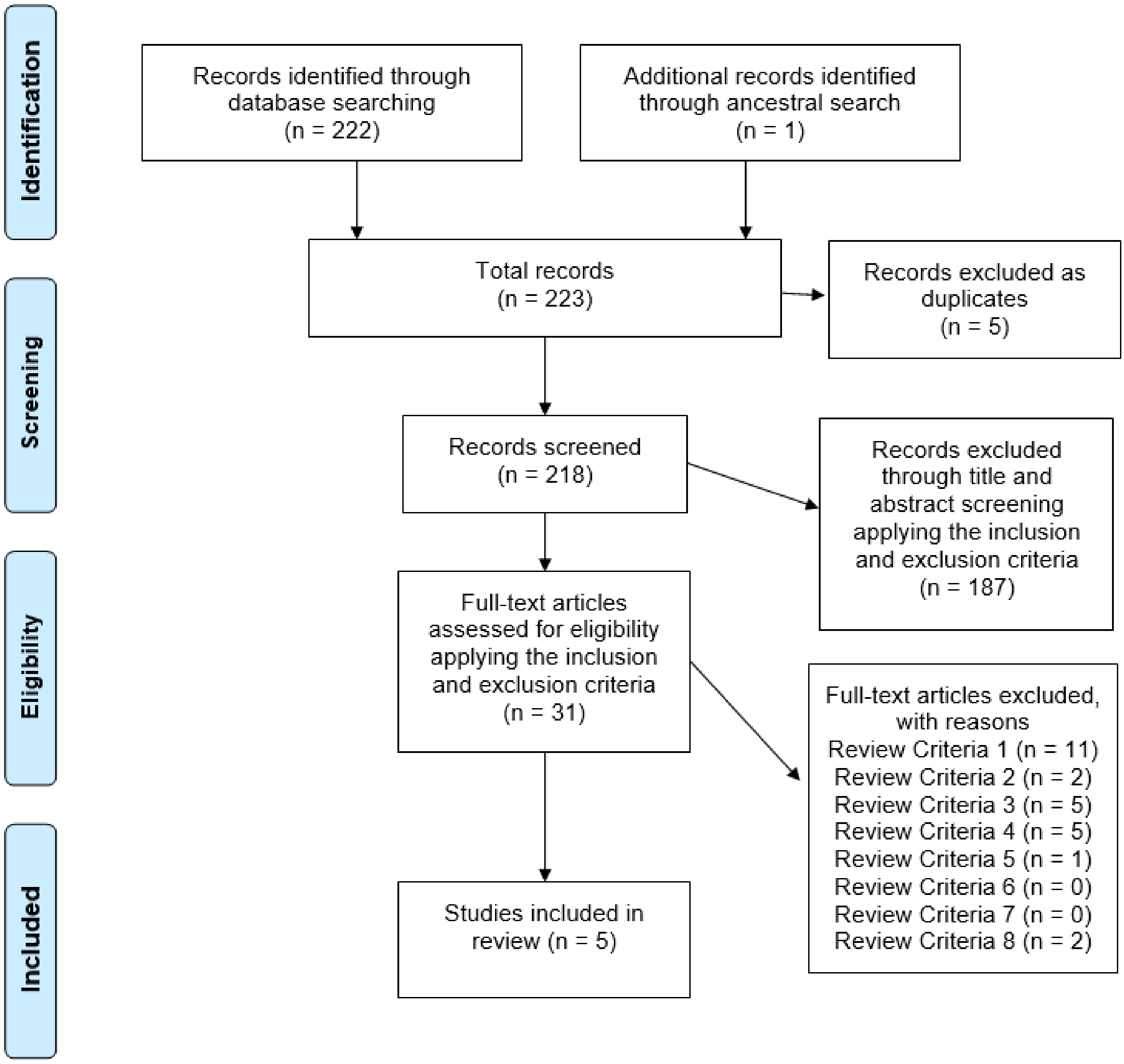


Table 2

Inclusion and Exclusion Criteria for Current Review

		Inclusion Criteria	Exclusion Criteria	Rationale
1.	Population	Typically developing children of up to 7 years	Children greater than 7 years, or children with special educational needs, learning disabilities or identified by teachers as low performing or at risk children	Evidence suggests that between 0 and up to 7 years is typically the age group between informal learning in preschool and formal learning in primary school and thus interventions targeting this population could increase teachers' awareness of how to support young children in this crucial time (Baroody, Eiland & Thompson, 2009)
2.	Setting	The review will include interventions conducted in nursery, pre-school or primary school settings	This review will not include interventions that are community-based	Settings should be typical of regular school settings
3.	Intervention	Technology-based academic game must be completed using a computer or tablet during school hours	All other interventions using other mediums or games for social pleasure	Technology-based programme is specific to this review question.
		Intervention must be facilitated by a class teacher or teaching assistant	Any other intervention facilitated by other adults	This review question is focusing on interventions led by teachers during school hours

		Inclusion Criteria	Exclusion Criteria	Rationale
4.	Outcome	Measure of mathematical attainment	All other outcome measures	The review question is related to mathematical attainment
5.	Study design	Studies with a group intervention design	All other study designs	Methods in which there is a pre-measure, post-measure and a control group are highly valid research methods and allow the impact of the intervention
6.	Published after	Studies published after 2017	Published before 2020	Game based learning companies saw a large increase in investment since 2017, with a total of $2.25bn globally, more than double the year before (Adkins, 2018)
7.	Reported language	Reported in the English language	Reported in a language other than English	Author's primary language is English; author has limited access to translation services
8.	Publication article	Studies were published in a peer-reviewed journal	The studies were not published in a peer reviewed journal	To ensure high quality research

Sourced studies

The remaining five studies that met the inclusion and exclusion criteria were selected for analysis (see Table 3 for full references).

Table 3

Full references of the final six studies included in the systematic review

Gecu-Parmaksiz, Z., & Delialioglu, O. (2019). Augmented reality-based virtual manipulatives versus physical manipulatives for teaching geometric shapes to preschool children. *British Journal of Educational Technology*, *50*(6), 3376-3390.
Miller, T. (2018). Developing numeracy skills using interactive technology in a play-based learning environment. *International journal of STEM education*, *5*(1), 39.
Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2017). Closing the gap: efficacy of a tablet intervention to support the development of early mathematical skills in UK primary school children. *Computers & Education*, *108*, 43-58.
Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J. (2019). Raising early achievement in math with interactive applications: A randomized control trial. *Journal of educational psychology*, *111*(2), 284.
Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2018). The effectiveness of computer and tablet assisted intervention in early childhood students' understanding of numbers. An empirical study conducted in Greece. *Education and Information Technologies*, *23*(5), 1849-1871.

Weight of Evidence

To evaluate the extent to which each piece of the evidence contributes to answering the review question, Gough's (2007) Weight of Evidence Framework was employed. It is a well-recognised framework with three dimensions for weighting studies.

Weight of Evidence A (WoE A) focuses on the methodological quality of execution of the study in relation to quality standards for studies of that type (Gough, 2007). In this review, a modified version of Gersten et al.'s (2005) quality indicators for group experimental and quasi-experimental research in special education was employed. This indicator was amended for the specific purpose of this review which is detailed and justified in Appendix B (see Appendix C for example completed coding protocols one paper in this review). Following this, criteria were defined to evaluate WoE A and each study was assigned a WoE A rating (see Appendix D).

Weight of Evidence B (WoE B) consists of a review-specific judgement about the appropriateness of the type of evidence for answering the review question (Gough, 2007). Criteria were created based on the recommendations of Petticrew and Roberts (2003) for research on the effectiveness of interventions (see Appendix E for criteria ratings and WoE B ratings for each study).

Weight of Evidence C (WoE C) is a review-specific judgement about the relevance of the focus of the evidence to the review question (Gough, 2007). Studies were assessed according to various criteria and given a WoE C weighted rating (see Appendix F).

To explain the relative contribution of each WoE to the review, WoE A, B and C were combined to provide an overall judgement, Weight of Evidence D (WoE D; Gough, 2007). Each dimension was treated as having equal weight. The overall WoE A, B and C ratings were added together and divided by three to produce WoE D (see table 4). Each weighting score ranged from 1-3 (see Appendix G for the WoE D average score bands).

Table 4

Overall Weight of Evidence

Study	WoE A	WoE B	WoE C	WoE D[a]
Gecu-Parmaksiz & Delialioglu (2019)	1	2	1.5	1.4 (Low)
Miller (2018)	1	2	2	1.6 (Low)
Outhwaite et al. (2017)	1	1	2.75	4.75 (High)
Outhwaite et al. (2019)	3	2	2.5	2.5 (Medium)
Papadakis et al. (2018)	2	3	2.5	2.5 (Medium)

[a]A score of <2 is marked as low, 2-3 as medium and >3 as high

Participants

The five studies in this review included 1938 participants from the UK, Canada, Turkey and Greece. As part of the inclusion criteria, all children ranged from age four to age seven. Additionally, all participants were recruited through their schools, given that the focus of this review question was assessing the impact of the intervention in improving numeracy skills in a learning environment. Each study varied in the number of participants in the intervention group, ranging from 7 (Miller, 2018) to 183 (Papadakis et al.,2018). All studies randomly assigned participants at an individual level to groups to prevent bias in the results, with a comparative control group, as supported by WoE A ratings. The exception to this was the study by Outhwaite et al. (2017) who followed a within-subject design so that all participants in the Early Years received the intervention and were compared to a control group in Key Stage 1 that was not age-matched, resulting in a low WoE A rating.

Additionally, four studies reported information on the socio-economic status of participants, of which two were representing participants from low socio-economic status (SES; Gecu-Parmaksiz & Delialioglu, 2019; Outhwaite et al., 2017). This suggests samples may be limited in their generalisability to wider populations from different SES. Papadakis et al. (2018) report a homogenous sample with regard to SES; however, the direction was not specified.

Research design

Two studies selected were Randomised Control Trials, a study design recognised with the highest methodological rigour (Petticrew & Roberts, 2003). Other research designs included mixed methods, quasi-experimental and group experimental. All studies met criteria for pre-test and post-test measures, as supported by WoE A and B ratings. Gecu-Parmaksiz and Delialioglu (2019) reported implementation of a pilot study to address concerns over validity and evaluate the assessment procedures and implementation of intervention (Leon, Davis, & Kraemer, 2012), contributing to this study attaining the highest WoE B rating. Outhwaite et al. (2017) was the only study to carry out a 5-month follow up, contributing to its WoE A rating, however this was not weighted strongly enough to increase its overall rating.

Intervention

Computer or tablet-based mathematical interventions were the focus of this review, however the interventions were all delivered in a different format across the studies. Four studies introduced the concept of applications to support numeracy skills in number counting and recognition, which could be easily downloaded from application stores (Gecu-Parmaksiz & Delialioglu., 2019; Miller, 2018; Outhwaite et al., 2019; Outhwaite et al., 2017) while the remaining study used educational software (Papadakis et al., 2018; see Appendix H for mapping of intervention details).

The presentation of the applications varied across the studies. Gecu-Parmaksiz and Delialioglu (2019) used augmented reality-based virtual manipulatives, focusing on children's' understanding of geometric shapes through applications that allowed them to manipulate shapes. Three studies used child-centred instruction to present topics on core mathematical concepts using a virtual teacher or game characters to structure their learning (Outhwaite et al., 2019; Outhwaite et al., 2017; Papadakis et al., 2018). However, the application used in Miller's (2018) study was limited in reporting the detail, accounted for in its low WoE A score.

Intervention fidelity varied across the studies. This was measured by who implemented the intervention and to what extent they received initial training or on-going supervision. Three papers included an element of pre-training or guidance prior to the intervention, through face-to-face meetings or access to teacher manuals, or on-going supervision by the researcher during implementation (Miller, 2018; Outhwaite et al., 2017; Papadakis et al., 2018). This allowed the intervention to be uniform in delivery across participants, thus improving the construct validity of the findings. The extent of treatment fidelity in each paper was further scrutinised in the WoE A ratings. Miller (2018), Outhwaite et al. (2019) and Outhwaite et al. (2017) also scored highly on location of intervention as they were based in the UK or OECD countries, contributing to their high overall WoE C ratings.

Measures

All studies varied in the tools used to measure numeracy skills in the pupils. These were either generalised measures of numeracy (Outhwaite et al., 2019; Outhwaite et al., 2017; Papadakis et al., 2018) or specific measures aligned with the intervention (Gecu-Parmaksiz & Delialioglu, 2019; Miller, 2018). One study was notorious for using both generalised and specific measures of numeracy skills (Outhwaite et al., 2017), reflected in its high WoE C rating.

However, what was not consistent between the studies was their reporting of reliability and validity of the measures provided, whereby either one or the other was reported (Miller, 2018; Papadakis et al., 2018) or neither. Outhwaite et al. (2019) provided information on the high internal consistency and validity of the measure, the Progress Test in Maths 5 (PTM5), contributing to its high WoE A rating.

Findings

Outcomes and effect sizes

Given the focus of this review is on interventions improving numeracy skills in young children, only effect sizes on general measures of numeracy were reported (see table 5). Cohen’s d effect sizes were extracted from two studies (Gecu-Parmaksiz & Delialioglu, 2019; Outhwaite et al., 2019). Papadakis et al. (2018) reported partial eta squared, which was converted to Cohen’s d using the Campbell Collaboration Calculator (see Appendix I for details of the calculator), in order to compare effect sizes across studies. The remaining papers did not report effect sizes; hence, these were calculated using means and standard deviations using post-intervention data.

The study by Gecu-Parmaksiz and Delialioglu (2019) was the sole study to report a large effect size for the intervention, suggesting the use of augmented reality-based virtual manipulatives has a statistically significant impact on children’s typical and atypical shape knowledge. This is in concurrent with previous studies that have found that using virtual manipulatives as a new technological tool can help children understand mathematical concepts, given their visual and tactile teaching components (McNeil & Jarvin, 2007). However, these results must be interpreted with caution, given that they were compared to a control group which had a significantly higher pre-intervention score on the outcome measure, so the construct validity may be limited. Furthermore, Geu-Parmaksiz and Delialioglu’s (2019) study had one of the smallest sample sizes, 72, across all the papers in this review, limiting the reliability and replicability of the study (Slavin & Smith, 2009).

The remaining studies all presented with small effect sizes. In the case of Miller (2018), this suggests that the difference between using mathematical applications compared to traditional play-based learning is trivial when assessing children’s numeracy development.

Outhwaite et al. (2019), Outhwaite et al. (2017) and Papadakis et al. (2018) all used an ANOVA to compare mean group differences to that led to significant improvements in numeracy development because of the intervention. However, these studies did not report pre-exposing children to the tablets and computers, suggesting there may be a novelty effect from playing with new technology. Thus, this may have influenced the improved numeracy skills noted in these interventions.

It is important to note that although Outhwaite et al.'s (2017) study had a small effect size, it had the highest WoE D rating due to the significant improvement in numeracy knowledge sustained at 5 months follow-up.

Table 5

Effect sizes and Descriptors for studies.

Study	Sample size	Outcome measures	Cohen's d[a]	Effect size descriptor	Significance value	95% Confidence interval	Overall WoE D rating
Miller (2018)	13	In-app numeracy test	.126	Small	$p < .001$	-.966 - 1.22	1.6 (Low)
Papadakis et al. (2018)	365	The Test of Early Mathematics Ability – Third Edition (TEMA-3)	-.079	Small	$p < .001$	-.332 - .174	2.5 (Medium)
Outhwaite et al. (2019)	389	PTM5 (Math Assessment Resource Service, 2015)	.21	Small	$p < .042$	-.03 - .46	2.5 (Medium)
Gecu-Parmaksiz & Delialioglu (2019)	72	Geometric Shape Recognition Task (GSRT)	1.17	Large	$p < .001$	0.668 – 1.67	1.4 (Low)
Outhwaite et al. (2017)	133	Measure of maths concepts (similar to Numerical Operations sub-test of the WIAT-II)	0.3	Small	$p < .008$	11.7 – 14.1	4.75 (High)

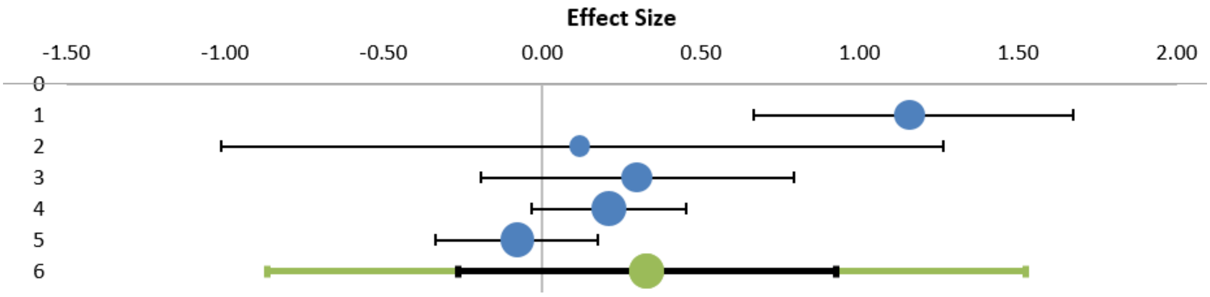
[a] An effect size of less than 0.2 is considered small, 0.5 is medium and 0.8 or above is large (Cohen, 1998)

Meta-Analysis

Given that the studies in this review presented with methodologically sound effect sizes for the specified populations, a meta-analysis was carried out for hypothesis testing using the Meta-Essentials software (see Appendix J for details). All five papers were included in this meta-analysis, applying a random-effects model. Figure 2 illustrates a forest plot to represent that various study effect sizes and confidence intervals.

Figure 2

Forest Plot of Effect Sizes at Post Intervention



Two of the studies showed a significant positive effect while three did not. The overall combined effect size was medium, 0.54 (95% CI -.22, 1.30) which was not statistically significant ($z = 1.96$, $p = 0.05$), for the use of tablet or computer-based interventions and improvement in children's numeracy skills. There is evidence of considerable heterogeneity (I^2=92.49%, p_Q= 0.000), suggesting the variation across studies may be due to true heterogeneity rather than chance.

Conclusions and Recommendations

Conclusions

This review aimed to evaluate the effectiveness of computer or tablet-based mathematical interventions in improving numeracy skills in young children. The methodological quality of the studies must be accounted for when forming conclusions from this review. Two studies received a low overall WoE D rating (Gecu-Parmaksiz & Delialioglu., 2019; Miller, 2019), two received a medium (Outhwaite et al., 2019; Papadakis et al., 2018) and only one study was weighted with a high WoE D (Outhwaite et al., 2017). Majority of these papers reported small effect sizes of the post-intervention outcomes.

The meta-analysis revealed these results were not statistically significant in favour of this technology being effective for improving numeracy skills, suggesting the interaction between tablet or computer-based interventions and numeracy skills may be varied by other factors beyond chance.

Recommendations

This review has contributed to the existing research that it is important to build interventions to support numeracy skill development, given it is a strong predictor of later life success (Claessens & Engel, 2013). Although these studies provide insufficient evidence to suggest tablet or computer based interventions can improve numeracy skills, there is evidence to suggest that these technological manipulatives can be more advantageous than traditional tools. This review highlights the benefits of a multiagency approach to numeracy development where teachers and Educational Psychologists can work together to use new resources in the classroom. Furthermore, it highlights the importance and value of using modern technological advances to improve engagement and success in numeracy.

Limitations

Intervention fidelity varied considerably across the studies. Two papers reported the extensive number of years of early years teaching experience that the teachers implementing the intervention had (Miller, 2018; Papadakis et al., 2018) however this was not documented in Outhwaite et al. (2019). Teacher's experience may have been a confounding variable that affected pupils' success in the intervention (Podolsky, Kini & Darling-Hammond, 2019), not accounted for in the studies. Furthermore, two researchers were responsible for carrying out the intervention in their studies (Gecu-Parmaksiz & Delialioglu., 2019; Outhwaite et al., 2019), which may have led to a publication bias and an overestimation of the effectiveness of the intervention.

The meta-analysis indicated substantial heterogeneity across the studies. From a methodological perspective, this may be explained by the outcome measures used whereby two papers did not report any means of standardisation of measures such as the Geometric Shape Recognition Task (Gecu-Parmaksiz & Delialioglu, 2019, 2019) and 'Explain Everything' numeracy test application (Miller, 2019), limiting the reliability and validity of the measures. The factors that varied the outcome may have been explored through a moderator analysis, however this was not the aim of the current review. Substantive factors could have accounted for the heterogeneity such as SES, the number of applications accessed online and the duration of the intervention.

Areas for Future Research

The first area of future research focuses on assessing the effectiveness of technology based mathematical interventions in special education. Given the sensory nature of tablet-based interventions with an appealing tactile element, these technologies may create highly stimulating learning environments for learners who may benefit from high sensory input and a highly sensitive touch-screen (Kagohara et al., 2013).

While EPs have a pivotal role in working with teachers within a school context, their collaborative role in working with families is also a fundamental part of their practice (Todd, 2007). To take these findings further, it would be good practice to identify whether the use of computer or tablet-based mathematical interventions may have an effect on numeracy skills development when delivered at home. The engagement of parents and families alongside schools is vital when considering the child in a system and thus exploring the effectiveness of interventions in multiple contexts would further support this research area.

Given the substantial heterogeneity in the studies critiqued in this review, it is plausible that the differences between the studies may be attributable to moderators that may be explored in future research. For example, the intervention fidelity varied across studies, whereby three studies did not report teachers' being supervised by researchers to ensure validity of the intervention (Gecu-Parmaksiz et al., 2019; Miller, 2018; Outhwaite et al., 2017). Future research may see benefit from providing supervision to teachers when delivering such new technological interventions and their effectiveness.

When considering the use of computer or tablet-based mathematical interventions in young children, this review provides limited evidence in its favour and as a result, its use in mainstream classrooms may be contended. However, in light of its limitations and areas of future research, it is important to explore whether they may be more effective in special educational needs populations and to what extent its' effectiveness depends on the environment it is used in. With technology advancing at exponential rates (Butler, 2016), it is important this research is explored in the future in order to account for new technology being brought into education.

References

Adkins, S. S. (2018). *The 2017 Global Learning Technology Investment Patterns*. 30. Retrieved from http://users.neo.registeredsite.com/9/8/1/17460189/assets/Metaari_s-Analysis-of-the-2017-Global-Learning-Technology-Investment-Pat27238.pdf

Baroody, A. J., Eiland, M., & Thompson, B. (2009). Fostering at-risk preschoolers' number sense. *Early Education and Development*, *20*(1), 80-128.

Bray, A., & Tangney, B. (2016). Enhancing student engagement through the affordances of mobile technology: a 21st century learning perspective on Realistic Mathematics Education. *Mathematics Education Research Journal*, *28*(1), 173-197.

Butler, D. (2016). Tomorrow 's Technological Change Is Accelerating Today At an Unprecedented World. *Nature*, *530*(7591), 398401. Retrieved from https://ezp.waldenulibrary.org/login?url=https://search-proquest-com.ezp.waldenulibrary.org/docview/1768830517?accountid=14872

Claessens, A., & Engel, M. (2013). How Important Is Where You Start? Early Mathematics Knowledge and Later School Success. *Teachers College Record, 115*(6), 1-29

Clarke, B., & Svanaes, S. (2014). An updated literature review on the use of tablets in education. *Tablets for Schools. UK: Family Kids & Youth*.

Education Scotland. (2019). *Numeracy across learning: Principles and practice*. Retrieved from https://education.gov.scot/Documents/numeracy-across-learning-pp.pdf

Engelmann, S., Becker, W. C., & Gersten, R. (2014). The direct instruction follow through model : design and outcomes. *Education and Treatment of Children,11*(4), 303–317.

Gecu-Parmaksiz, Z., & Delialioglu, O. (2019). Augmented reality-based virtual manipulatives versus physical manipulatives for teaching geometric shapes to preschool children. *British Journal of Educational Technology*, *50*(6), 3376–3390.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children*, *71*(2), 149-164.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research papers in education*, *22*(2), 213-228.

Hak, T., Van Rhee, H. J., & Suurmond, R. (2016). How to interpret results of meta-analysis. (Version 1.3) Rotterdam, The Netherlands: Erasmus Rotterdam

Institute of Management. Retrieved from www.erim.eur.nl/research-support/meta-essentials/downloads

Health and Care Professions Council (2016). Standards of Conduct, Performance and Ethics. HCPC: London. Retrieved from https://www.hcpc-uk.org/standards/standards-of-conduct-performance-and-ethics/

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes. *Developmental Psychology*, *45*(3), 850–867.

Kagohara, D. M., van der Meer, L., Ramdoss, S., O'Reilly, M. F., Lancioni, G. E., Davis, T. N., Rispoli, M., Marschik, P. B., Sutherland, D., Green, V. A., & Sigafoos, J. (2013). Using iPods® and iPads® in teaching programs for individuals with developmental disabilities: A systematic review, *34*(1), 147-156.

Kaplan, A., Katz, I., & Flum, H. (2011). Motivation theory in educational practice: Knowledge claims, challenges, and future directions. *APA Educational Psychology Handbook: Individual Differences and Cultural and Contextual Factors*, *2*(1), 165–194.

Leon, A., Davis, L., & Kraemer, H. (2012). The role and interpretation of pilot studies in clinical research. *45*(5), 626–629.

Madigan, S., Browne, D., Racine, N., Mori, C., & Tough, S. (2019). Association between Screen Time and Children's Performance on a Developmental Screening Test. *Journal of American Medical Association Pediatrics*, *173*(3), 244–250.

Magliaro, S. G., Lockee, B. B., & Burton, J. K. (2005). Direct instruction revisited: A key model for instructional technology. *Educational Technology Research and Development*, *53*(4), 41–55.

McNeil, N., & Jarvin, L. (2007). When theories don't add up: disentangling he manipulatives debate. *Theory into practice*, *46*(4), 309-316.

Moyer-Packenham, P. S., & Bolyard, J. J. (2016). Revisiting the definition of a virtual manipulative. In *International perspectives on teaching and learning mathematics with virtual manipulatives*, 3-23.

Miller, T. (2018). *Developing numeracy skills using interactive technology in a play-based learning environment*.

National Association for the Education of Young Children, & Fred Rogers Center for Early Learning and Children's Media at Saint Vincent College. (2012). Technology and interactive media as tools in early childhood programs serving children from birth through age 8 Retrieved from. http://www.naeyc.org/files/naeyc/file/positions/PS technology WEB2.pdf

Ofsted. (2018)*. Bold beginnings: The Reception curriculum in a sample of good and outstanding primary schools*. London, UK: Ofsted

Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2017). Closing the gap: Efficacy of a tablet intervention to support the development of early mathematical skills in UK primary school children. *Computers and Education*, *108*, 43–58.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology & Community Health*, *57*(7), 527-529.

Podolsky, A., Kini, T., & Darling-Hammond, L. (2019). Does teaching experience increase teacher effectiveness? A review of US research. *Journal of Professional Capital and Community*, 1-72.

Rosenfeld, D., Dominguez, X., Llorente, C., Pasnik, S., Moorthy, S., Hupert, N., … Vidiksis, R. (2019). Early Childhood Research Quarterly A curriculum supplement that integrates transmedia to promote early math learning : A randomized controlled trial of a PBS KIDS intervention. *Early Childhood Research Quarterly*, *49*, 241–253.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 347–363.

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

Todd, L. (2007). *Partnerships for Inclusive Education: A critical approach to collaborative working*. Routledge.

Tran, C., Chen, J., William, C., Conley, A., & Dede, C. (2012). Applying Motivation Theories to the Design of Educational Technology, *48*. Retrieved from http://publish.wm.edu/articles/48

Van Rhee, H.J., Suurmond, R., & Hak, T. (2015). User manual for *Meta-Essentials*: Workbooks for meta-analysis (Version 1.4) Rotterdam, The Netherlands: Erasmus Research Institute of Management. Retrieved from www.erim.eur.nl/research-support/meta-essentials

Appendix A

Table 1

List of excluded studies at full review

Excluded Studies	Exclusion Criterion number
Freiman, V., Polotskaia, E., & Savard, A. (2017). Using a computer-based learning task to promote work on mathematical relationships in the context of word problems in early grades. *ZDM*, *49*(6), 835-849.	1. Study is based on pupils 7-8 years old
Kermani, H. (2017). Computer mathematics games and conditions for enhance young children's learning of number sense. Malaysian Journal of Learning and Instruction (MJLI), 14 (2), 23-57.	1. Study defines population as at risk
Cornu, V., Schiltz, C., Pazouki, T., & Martin, R. (2019). Training early visuo-spatial abilities: A controlled classroom-based intervention study. Applied Developmental Science, 23(1), 1-21.	4. Intervention assesses visuo-motor skills and it's generalisation to mathematical attainment
McCarthy, E., Tiu, M., & Li, L. (2018). Learning Math with Curious George and the Odd Squad: Transmedia in the Classroom. *Technology, Knowledge and Learning*, *23*(2), 223-246.	3. Intervention includes optional activity using technology with parents at home
Taylor, M. S. (2018). Computer programming with Pre-K through first-grade students with intellectual disabilities. *The journal of special education*, *52*(2), 78-88.	1. Population identified to be children with intellectual disabilities
Barreto, D., Vasconcelos, L., & Orey, M. (2017). Motivation and learning engagement through playing math video games. Malaysian Journal of Learning and Instruction (MJLI), 14 (2), 1-21.	1. Study is based on pupils 8-9 years old
Fulton, L., Paek, S., & Taoka, M. (2017). Science notebooks for the 21st century. *Science and Children*, *54*(5), 54.	8. Not published in a peer reviewed journal
Beckman, A., Mason, B. A., Wills, H. P., Garrison-Kane, L., & Huffman, J. (2019). Improving Behavioral and Academic Outcomes for Students with Autism Spectrum Disorder:	1. Population identified to be

Excluded Studies	Exclusion Criterion number
Testing an App-based Self-monitoring Intervention. *Education and Treatment of Children*, *42*(2), 225-244.	children with autism
Coxon, S. V., Dohrman, R. L., & Nadler, D. R. (2018). Children using robotics for engineering, science, technology, and math (CREST-M): The development and evaluation of an engaging math curriculum. *Roeper Review*, *40*(2), 86-96.	3. Intervention carried out during summer programme
Yanikoglu, B., Gogus, A., & Inal, E. (2017). Use of handwriting recognition technologies in tablet-based learning modules for first grade education. *Educational Technology Research and Development*, *65*(5), 1369-1388.	4. Looks at motivation of children using technology rather than attainment
Burte, H., Gardony, A. L., Hutton, A., & Taylor, H. A. (2017). Think3d!: Improving mathematics learning through embodied spatial training. *Cognitive Research: Principles and Implications*, *2*(1), 13.	3. Looks at the effect of spatial thinking skills in paper folding programme on attainment
Techaraungrong, P., Suksakulchai, S., Kaewprapan, W., & Murphy, E. (2017). The design and testing of multimedia for teaching arithmetic to deaf learners. *Education and Information Technologies*, *22*(1), 215-237.	1. Population to be children identified as deaf learners
Liu, Z., Zhi, R., Hicks, A., & Barnes, T. (2017). Understanding problem solving behavior of 6–8 graders in a debugging game. *Computer Science Education*, *27*(1), 1-29.	4. Intervention assesses programming skills in computer science and not mathematics
Falco, L. D. (2019). An intervention to support mathematics self-efficacy in middle school. *Middle School Journal*, *50*(2), 28-44.	4. Intervention assess self-efficacy in mathematics and not attainment
Aragón-Mendizábal, E., Aguilar-Villagrán, M., Navarro-Guzmán, J. I., & Howell, R. (2017). Improving number sense in kindergarten children with low achievement in mathematics. *Anales De Psicología/Annals of Psychology*, *33*(2), 311-318.	1. Study compares performance between low and high-performing maths students
Stacy, S. T., Cartwright, M., Arwood, Z., Canfield, J. P., & Kloos, H. (2017). Addressing the math-practice gap in elementary school: Are tablets a feasible tool for informal math practice?. *Frontiers in psychology*, *8*, 179.	2. Community-based intervention

Excluded Studies	Exclusion Criterion number
Rodríguez, J. L., Morga, G., & Cangas-Moldes, D. (2019). Geometry teaching experience in virtual reality with NeoTrie VR. *Psychology, Society, & Education*, *11*(3), 355-366.	1. Population are children in secondary school
Dowker, A. (2017). Interventions for primary school children with difficulties in mathematics. In *Advances in child development and behavior* (Vol. 53, pp. 255-287). JAI.	8. Not published in a peer reviewed journal
Van de Ven, F., Segers, F., Takashima, A., & Verhoeven, L. (2017). Effects of a tablet game intervention on simple addition and subtraction fluency in first graders computers in human behavior. *Computers in Human Behavior*, *72*, 200-207.	3. Intervention based on teaching Arabic numbers
Constantin, A., Johnson, H., Smith, E., Lengyel, D., & Brosnan, M. (2017). Designing computer-based rewards with and for children with Autism Spectrum Disorder and/or Intellectual Disability. *Computers in Human Behavior*, *75*, 404-414.	1. Population identified to be children with autism
Hassler Hallstedt, M., Klingberg, T., & Ghaderi, A. (2018). Short and long-term effects of a mathematics tablet intervention for low performing second graders. *Journal of Educational Psychology*, *110*(8), 1127.	1. Population of low-performing maths students
Hwa, S. P. (2018). Pedagogical Change in Mathematics Learning: Harnessing the Power of Digital Game-Based Learning. *Journal of Educational Technology & Society*, *21*(4), 259-276.	3. Intervention carried out at home
Sánchez-Pérez, N., Castillo, A., López-López, J. A., Pina, V., Puga, J. L., Campoy, G., & Fuentes, L. J. (2018). Computer-Based Training in Math and Working Memory Improves Cognitive Skills and Academic Achievement in Primary School Children: Behavioral Results. *Frontiers in psychology*, *8*, 2327.	4. Intervention assess working memory on mathematical attainment
Ginsburg, H. (2017). Maya Loves Tubby, and You Should, too, if You Want to Understand and Evaluate Children's Math Learning in the Digital Age. Teachers College Record, 119(6), 1.	5. Single-case design
Nicholas, K., & Fletcher, J. (2017). What is happening in the use of ICT mathematics to support young adolescent	1. Population of

Excluded Studies	Exclusion Criterion number
learners? A New Zealand experience. *Educational review*, *69*(4), 474-489.	students in secondary school
Rosenfeld, D., Dominguez, X., Llorente, C., Pasnik, S., Moorthy, S., Hupert, N., Gerard., S., & Vidiksis, R. (2019). A curriculum supplement that integrates transmedia to promote early math learning: A randomized controlled trial of a PBS KIDS intervention. *Early Childhood Research Quarterly*, *49*, 241-253.	2. Mixture of community and school-based intervention

Appendix B

The coding protocol from Gersten et al. (2005) has been used in this review. Table 1 details the amendments and justifications to this quality indicator.

Table 1

Amendments to the Gersten et al. (2005) protocol

Section Heading	Indicator	Section modified	Rationale
Describing participants	Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?	Section removed	The present review focused on typically developing populations, as part of its inclusion criteria
Describing participants	Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?	Were appropriate procedures used to ensure that participants were comparable across intervention conditions on relevant characteristics?	The present review focused on typically developing populations, hence the characteristics for comparison were not relevant but it was important to compare the populations across the intervention
Describing participants	Was sufficient information given characterising the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?	Were appropriate procedures used to increase the probability that teachers or interventions were comparable across conditions?	The focus of this review was on the effectiveness of interventions, comparing the experimental groups with control groups, hence it was important to ensure procedures were in place to ensure probable comparability

Appendix C

Example Coding Protocol for Weight of Evidence A

Adapted from Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children*, *71*(2), 149-164.

Study: Gecu-Parmaksiz and Delialioglu (2019)

Essential Quality Indicators

Quality Indicators for Describing Participants

1. Were appropriate procedures used to ensure that participants were comparable across intervention conditions on relevant characteristics?

☒Yes
☐No
☐N/A
☐Unknown/unable to code

2. Were appropriate procedures used to increase the probability that teachers or interventions were comparable across conditions?

☒Yes
☐No
☐N/A
☐Unknown/unable to code

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

☒Yes
☐No
☐N/A
☐Unknown/unable to code

Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions

1. Was the intervention clearly described and specified?

☒Yes
☐No
☐N/A
☐Unknown/unable to code

2. Was the fidelity of implementation described and assessed?

☐Yes

☒No
☐N/A
☐Unknown/unable to code

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

☒Yes
☐No
☐N/A
☐Unknown/unable to code

Quality Indicators for Outcome Measures

1. 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

2. Were outcomes for capturing the intervention’s effect measured at the appropriate times?

☒Yes
☐No
☐N/A
☐Unknown/unable to code

Quality Indicators for Data Analysis

1. 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

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

☒Yes
☐No
☐N/A
☐Unknown/unable to code

Desirable Quality Indicators

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

☐Yes
☒No
☐N/A
☐Unknown/unable to code

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

☐Yes
☒No
☐N/A
☐Unknown/unable to code

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

☐Yes
☒No
☐N/A
☐Unknown/unable to code

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

☐Yes
☒No
☐N/A
☐Unknown/unable to code

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

☐Yes
☒No
☐N/A
☐Unknown/unable to code

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

☒Yes
☐No

☐N/A
☐Unknown/unable to code

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

☐Yes
☐No
☒N/A
☐Unknown/unable to code

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

☒Yes
☐No
☐N/A
☐Unknown/unable to code

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

Appendix D

Weight of Evidence A (WoE A): Methodological Quality

Table 1 highlights the criteria set for evaluating WoE A, based on recommendations of Gersten et al's (2005) coding protocol. Table 2 summarises the overall WoE A ratings for the six studies included in the present review.

Table 1

Weight of Evidence A Criteria (WoE A) for RCTs/Quasi-Experimental Studies

WoE A Rating	Criteria
3 (High)	1. Study meets at least 9 essential criteria 2. Study meets four or more desirable criteria
2 (Medium)	1. Study meets at least 9 essential criteria 2. Study meets at least 1 and less than 4 desirable criteria
1 (Low)	1. Study meets less than 9 essential criteria

Table 2

Summary of Weight of Evidence A (WoE A) Ratings for RCTs/Quasi-Experimental Studies

Study	Number of essential criteria	Number of desirable criteria	Overall WoE A
Gecu-Parmaksiz & Delialioglu (2019)	8	2	1
Miller (2018)	8	2	1
Outhwaite et al. (2017)	8	3	1
Outhwaite et al. (2019)	9	5	3
Papadakis et al (2018)	9	5	2

Appendix E

Weight of Evidence B (WoE B): Methodological Relevance

WoE B is a review-specific judgement on the appropriateness of a form of evidence used to answer the review question (Gough, 2007). Typologies are preferred to hierarchies as they assess the appropriateness of the type of study in relation to the review question (Petticrew & Roberts, 2003). Table 1 illustrates the criteria created based on "Typology of evidence" recommendations for research best suited to studying the effectiveness of interventions (Petticrew & Roberts, 2003). Each study must meet every part of the section to receive that rating. Table 2 provides the overall WoE B ratings for each paper.

Table 1

Weight of Evidence B (WoE B) Criteria

WoE B Rating and Qualitative Descriptor	Criteria
3 (High)	• At least one age-matched control/comparison group • Participants are randomly assigned to the treatment and control groups • Pre-test and post-test measures are reported for the treatment and control groups • Measures used to test the effectiveness are outlined and include reliability and validity information with reference to the use of a pilot
2 (Medium)	• At least one age-matched control/comparison group • Participants are randomly assigned to the treatment and control groups • Pre-test and post-test measures are reported for the treatment and control groups • Measures are used to test effectiveness and include reliability of validity information
1 (Low)	• No age-matched control/comparison group • Participants are not randomly assigned to the treatment or control groups

WoE B Rating and Qualitative Descriptor	Criteria
	• Pre-test and post-test measures are reported for the treatment and control groups • Measures used to test the effectiveness are outlined

Table 2

Weight of Evidence B

Study	Overall WoE B
Gecu-Parmaksiz & Delialioglu (2019)	2
Miller (2018)	2
Outhwaite et al. (2017)	1
Outhwaite et al. (2019)	2
Papadakis et al. (2018)	2

Appendix F

Weight of Evidence C (WoE C): Topic Relevance

WoE C is a review-specific judgement that evaluates a study based on the evidence in the review question (Gough, 2007). Studies were assessed on various criteria and given a WoE C rating (see table 1).

Table 1

WoE C Criteria Rating and Rationale

Criteria	WoE Rating	Descriptor	Rationale
Location of Intervention	3	Schools in the UK	The relevance of the research to the practice of Educational Psychologists in the UK would be highly valued if the interventions took place in the UK education settings, or those most similar, so that findings may be generalised
	2	Schools in OECD countries	
	1	Schools outside OECD countries	
Intervention fidelity	3	Teacher led intervention with researcher support, through initial training or ongoing supervision	Implementation with the support of researchers, will ensure the quality of delivery is maintained but still ensures that the teacher takes the lead
	2	Teacher led	
	1	Child led	
Breadth of outcome measures	3	Outcome measures include intervention software specific measures and general measures of numeracy skills	It would be of interest to assess whether the intervention has an impact on numeracy skills the general classroom context and not limited to the numeracy skills supported through the intervention
	2	Outcomes measures include general measure of numeracy skills	
	1	Outcome measure is specific to the intervention software only	

Criteria	WoE Rating	Descriptor	Rationale
Duration of intervention	3	The intervention is implemented over a period of 12 or more weeks	When evaluating research, evidence suggests that interventions should be implemented for a minimum of 12 weeks in order to determine the full benefits of the intervention (Higgins, Xiao & Katsipataki, 2012)
	2	The intervention was implemented over 6-11 weeks	
	1	The intervention was implemented for under 6 weeks	

Each study was reviewed against these criteria, which were totalled and divided by four to calculate an overall WoE C rating (see table 2 for rating).

Table 2

Summary of Weight of Evidence C Ratings

Study	Location of Intervention	Intervention Fidelity	Breadth of Outcome Measures	Duration of Intervention	Overall WoE C Rating and Qualitative Descriptor
Gecu-Parmaksiz & Delialioglu (2019)	2	2	1	1	1.5 (low)
Miller (2018)	3	3	1	1	2 (medium)
Outhwaite et al. (2017)	3	3	3	2	2.75 (high)
Outhwaite et al. (2019)	3	2	2	3	2.5 (high)
Papadakis et al. (2018)	2	3	2	3	2.5 (high)

Appendix G

Weight of Evidence D (WoE D): Overall WoE

Table 1

WoE D Rating Descriptor

Score	WoE D Descriptor
<2	Low
2-3	Medium
>3	High

Appendix H

Table 1

Mapping the Field

Author and geographical distribution	Sample and age/gender	Study type	Intervention investigated	Context of intervention	Outcome variables measured	Follow up
Miller (2018); Canada	13 (intervention group 7); 4-5 years boys and girls	Mixed-methods	Children introduced to a total of 15 applications to support various numeracy skills including number drawing, counting and recognition	2 week interactive technology in a play-based mathematics classroom implemented by a veteran teacher with over 20 years of early years teaching experience alongside researchers	Pre- and post-tests on children's mathematics ability based on the provincial curriculum document and in consultation with the teacher; children's numeracy test scores using application, 'Explain Everything' Observational findings of children's behaviours and applications most favoured	None

Papadakis, Kalogiannakis and Zaranis (2018); Greece	365 (intervention group 183); mean age 5 years boys and girls	RCT	Educational software on computers and/or tablets in key areas of early mathematics e.g. number word sequence, enumeration, basic addition and subtraction	24 x 40-minute activities carried out over 14 weeks in children's classrooms by teachers and trained undergraduate or graduate students as part of regular mathematics instruction	The Test of Early Mathematics Ability – Third Edition (TEMA-3)	None
Outhwaite et al. (2019); UK	389 (intervention group 126 and time-equivalent intervention group 131); mean age 5.05 years boys and girls)	RCT	"Maths 3-5" and "Maths 4-6" math app intervention on iPads based on core mathematical concepts in Number and Shape, and Space and Measure, covered in the EYFS Profile (DfE, 2013) and UK National Primary Curriculum for Key Stage 1 (DfE, 2014) covering conceptual understanding, factual knowledge and mathematical reasoning skills	30 minutes daily, over 12-week intervention period. Children complete this individually, using headsets at their own pace, or in activities led by the class teacher in small groups or as a whole class	PTM5 (Math Assessment Resource Service, 2015) measure of mathematical ability with items from the EYFS Profile (DfE, 2013)	None

Gecu-Parmaksiz & Delialioglu (2019); Turkey	72 (intervention group 36); 5-6 years boys and girls	Quasi-experimental	Learnt about geometric shapes using virtual manipulatives through an Augmented Reality (AR) application called “Augment” on a tablet computer to categorise 2D and 3D geometric figures	45 minute per class, 4 weeks, with teacher (also researcher) support with terminology	Geometric Shape Recognition Task (GSRT)	None
Outhwaite, Gulliford & Pitchford (2017) UK	133 (intervention group 61); 4-7 years boys and girls	Group experimental	Maths tablet intervention with child-centred tuition focusing on Number, Shape, Space and Measure	Daily 30 minute intervention, every day for 6 weeks. Pupils work individually through app with headphones with a virtual teacher to scaffold learning with clear instructions and demonstrations. Children are in small groups of up to 15, under the supervision of teaching staff	Measure of curriculum knowledge (based on quiz items from the applications). Measure of maths concepts (similar to Numerical Operations sub-test of the WIAT-II)	5-months follow-up

Appendix I

Campbell Collaboration Calculator

A web-based effect-size calculator, provided by the Campbell Collaboration, was used to compute and calculate effect sizes for the analysis, using the input data provided in the studies. The calculator can be accessed from: https://campbellcollaboration.org/research-resources/effect-size-calculator.html

Appendix J

Meta-Essentials Software

In order to conduct the meta-analysis, the Meta-Essentials user manual version 1.4 was consulted to use the software tool. This manual detailed how to insert data, perform a meta-analysis and generate a forest plot, as appropriate for this review. This can be accessed from: Van Rhee, H.J., Suurmond, R., & Hak, T. (2015). User manual for *Meta-Essentials*: Workbooks for meta-analysis (Version 1.4) Rotterdam, The Netherlands: Erasmus Research Institute of Management. Retrieved from www.erim.eur.nl/research-support/meta-essentials.