CENTRE FOR EDUCATION POLICY & EQUALISING OPPORTUNITIES

## Working Paper No. 23-02

# Can Maths Apps Add Value to Learning? A Systematic Review


Laura A. Outhwaite
University College London

Erin Early
Queen's University Belfast

Christothea Herodotou
Open University

Jo Van Herwegen
University College London



Educational maths applications (apps) are an emerging trend in young children's learning environments aiming to raise attainment. The current systematic review aimed to thematically synthesise quantitative, qualitative, and mixed methods studies examining the impact of maths app interventions used at home or at school for young children in the first three years of compulsory education. The thematic narrative synthesis included 50 studies evaluating 77 maths apps with 23,981 children across 18 countries. Experimental methods were the most common designs, with 20 randomised control trials and 13 quasi-experimental designs. Most studies focused on mathematical learning outcomes with typically developing children and were conducted in the classroom, where practitioners implemented the app-based interventions. Studies predominately reported greater learning outcomes for young children using the evaluated maths apps compared to a range of control conditions. This provides promising evidence that maths apps can support young children's learning. However, usage and mathematical outcomes before and after the intervention were not consistently or reliably reported across studies, which should be addressed in future research. Based on the current evidence, eight directions for future research are also outlined to enhance the evidence base in this field and raise attainment in mathematics for young children.



VERSION: January 2023

Suggested citation: Outhwaite, L.A., Early, E., Herodotou, C., & Van Herwegen, J. (2023). *Can Maths Apps Add Value to Learning? A Systematic Review* (CEPEO Working Paper No. 23-02). Centre for Education Policy and Equalising Opportunities, UCL. https://EconPapers.repec.org/RePEc:ucl:cepeow:23-02

Disclaimer

Any opinions expressed here are those of the author(s) and not those of the UCL Institute of Education. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

CEPEO Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## Highlights

- Educational maths applications (apps) are an emerging trend aiming to raise attainment and address inequalities with personalised learning opportunities.
- This systematic review synthesised 50 studies evaluating the use of educational maths apps for young children in the first three years of compulsory school (e.g., ages 4-7 years in England).
- 92% of studies found the evaluated maths apps had some positive benefits on children's learning outcomes.
- But few studies have evaluated the use of maths apps in the home learning environment and with children identified as underachieving in maths, including those with special educational needs and disabilities.
- Eight directions for future research are identified to advance the field and ensure that innovations in app-based learning address, rather than exacerbate, inequalities in learning.

## Why does this matter?

There is promising evidence that educational maths apps can add value to learning for young children. Further research is needed to advance understandings about for whom are they best suited and under what circumstances, and inform evidence-based recommendations for teachers, parents, and policy makers.

## Introduction

Approximately 55% (383 million) of the school-age population worldwide are not proficient in basic mathematical skills needed for everyday life and education (UNESCO, 2017). Even in high-income countries mathematical underachievement is a significant issue. In the United Kingdom (UK), 20% of 5-year-old children do not have the mathematical skills expected for their age (Department for Education [DfE], 2017). Children who start formal education with low mathematical skills are significantly more likely to remain low achievers throughout school, compared to their higher attaining peers (Aubrey et al., 2006). Mathematical underachievement has a significant impact on children's later educational, economic, employment, mental and physical health outcomes (Davis-Kean et al., 2021; Reyna et al., 2009; Watts et al., 2017). To support children to develop a strong foundation in mathematical skills, evidence-based interventions are needed to engage children from a young age (Jordan & Levine, 2009).

Since touchscreen tablet devices were first introduced in 2010, educational maths applications (apps) have become increasingly common in young children's early learning experiences. Surveys show 94% of children in the UK (3-11 years) and 96% of children in the United States of America (USA; 3-4 years) own or have access to touchscreen tablet devices (Kabali et al., 2015; Marsh et al., 2020). Parents of young children in the UK are also reported to download educational apps more often than parents of older children (Chaudron, 2015) and 41% of primary school teachers in the USA report using maths apps as supplementary teaching tools in their classroom (Vega & Robb, 2019). In low- and middle-income countries, such as South Africa, children (3-11 years) are more likely to have access to a tablet device than a laptop or television (Marsh et al., 2020). Policy makers around the world are also increasingly advocating for and investing in the use of educational apps to support early learning in school and at home (DfE, 2019), including in low- and middle-income countries for achieving the education Sustainable Development Goals (Tamim et al., 2015; UNESCO, 2020).

### Previous Reviews

Research evaluating educational maths apps with young children at school and at home is emerging and is vital for supporting evidence-based decisions in their use and scaling for stakeholders in education, including teachers, parents, and policy makers. Two recent systematic reviews have synthesised experimental evaluations (randomised control trials [RCTs] and quasi-experimental designs [QEDs]) of educational apps (not just maths) with 0-6-year-old children. Herodotou (2018) included 19 studies examining the use of apps with children aged 2-5 years, of which five studies focused on mathematics and found positive learning benefits for 4-year-old children compared to younger 3-year-old children. Griffith et al. (2020) identified 15 studies with children aged 0-6 years and found interactive apps supported mathematical learning outcomes and called for more research focused on large-scale evaluations of educational apps. Similarly, Simms et al. (2019) identified 11 studies that used computerised instructional programmes with 4-11-year-old children including large-scale evaluations, but with mixed results. Although Simms et al. (2019) examined a broader age range compared to other systematic reviews (Griffith et al., 2020; Herodotou, 2018), a focus explicitly on educational apps was not included.

The focus on experimental studies in previous reviews give useful insights into whether educational apps can support learning. But it does not capture how these interventions are implemented and under what conditions they may be effective. This holistic understanding can be gained through a range of methodologies and are vital to scaling successful interventions (Pitchford, 2022). For example, other qualitative studies, not included in previous reviews, provide useful insights into how features of the maths app intervention implementation context can impact learning outcomes (e.g., Outhwaite et al., 2019) and how app features can support or hinder children's engagement and learning (e.g., Moyer-Packenham et al., 2016). Likewise, feasibility studies (e.g., Outhwaite et al., 2017), when successful, can support the development of further studies focusing on understanding the mechanisms and efficacy of the maths app interventions, and are a vital component of the scaling process (Green et al., 2019). Existing reviews have also not considered the impact of educational maths apps on non-attainment outcomes, such as enjoyment, preference modality, intrinsic motivation, and flow experience. These factors can be encapsulated as *'willingness to perform'* (Blumberg & Pringle, 1982) and are important to consider in the evaluation of such interventions as they underpin successful task performance and learning (Van Yperen et al., 2015).

Overall, a synthesis of the current quantitative, qualitative, and mixed methods research evidence examining whether app-based learning is an effective form of instruction for mathematical and non-attainment outcomes with young children at the start of formal education is needed to support evidence-based decisions by educational stakeholders and provide directions for future research.

**Current Review**

To address this need, the current review comprehensively synthesised all studies on the impact of educational maths apps for young children, since 2010. In doing so, the current review aimed to capture existing knowledge from the staged development of maths app evaluation studies (Green et al., 2019) and identify important gaps on whether maths apps can support learning, and under what circumstances apps are best implemented, as well as for whom are they most suited.

To achieve this aim, this systematic review completed a thematic narrative synthesis of the quantitative, qualitative, and mixed methods research literature that has examined the impact of educational maths apps used at home or at school for children in the first three years of compulsory school (e.g., in Reception- Year 2 in England children will be aged 4-7 years). This specific age group was chosen because of the popularity and prevalence of educational maths apps in young children's early learning experiences (Chaudron, 2015; Vega & Robb, 2019). The following four Research Questions (RQs) were addressed:

RQ1: What type, frequency, and quality of evaluation studies have been conducted with educational maths apps?

RQ2: What lessons can be learnt from previous research examining educational maths apps, relating to methods, populations, and outcomes?

RQ3: What is the external validity of the current evidence?

RQ4: What gaps can be identified based on the current evidence?

## Method

The systematic literature search was conducted using detailed and pre-defined search terms, as well as inclusion and exclusion criteria developed using the PICO (Population, Intervention, Comparison, and Outcome) method to select relevant studies for the thematic narrative synthesis. The systematic review protocol was reviewed by the advisory group and pre-registered on the Open Science Framework (https://osf.io/pzkmh/). Ethical approval for the current review was granted by IOE ethics committee (REC 1376).

### Search Strategy

A broad search of the literature was conducted using search terms focused on the population, intervention, and outcome components of the PICO statement. The following search string was used: ("early years" OR preschool* OR kindergart* OR "primary school" OR "elementary school" OR "young children" OR "young pupils" OR "young students" OR child* OR pupils OR students) AND ("educational app" OR "interactive app" OR tablet OR "tablet technology" OR "iPad app" OR "android app" OR "math* app") AND (math* OR number* OR "number sense" OR arithmetic* OR measurement OR geomet* OR shape). Preliminary scoping searches were conducted in January 2021 to optimise the search string (see Appendix). As touch-screen tablets were first introduced in 2010, the search focused on studies published between 2010-2021.

The search strategy was completed in January 2021 and included published and grey literature to reduce the risk of publication and selection bias. When conducting the search within online sources, the all-text option was selected on the databases to include studies that contained the search terms anywhere in the source record (i.e., title, abstract, keywords, or in-text), unless otherwise stated. Electronic web searches were completed on eleven scholarly databases, which covered education specific (ERIC, British Education Index, and Australian Education Index), social sciences (Social Science Citation Index), psychology (PsycINFO) and broad academic databases (Medline, Scopus [title, abstract, and keywords only], Science Citation Index-Expanded, Arts & Humanities Citation Index, Emerging Sources Citation Index, and PubMed). Five databases were used to search for grey literature, including ProQuest Dissertations and Theses (abstract only), Open Science Framework Preprints, PsyArXiv Preprints, Nuffield Foundation Research Reports (search string limited to "math*"), and Education Endowment Foundation Completed Projects (search string limited to "math*").

In addition, a backwards citation search of three recent and relevant systematic reviews (Griffith et al., 2020; Simms et al., 2019; Herodotou, 2018) was conducted. These existing reviews synthesise evidence on mathematical interventions and educational apps in general (i.e., not specific to maths), and thus were selected based on the similarities with the current review. A prospective forward citation search of included studies (n = 45) was also conducted in June 2021.

**Inclusion and Exclusion Criteria**

Inclusion and exclusion criteria were developed based on the PICO framework and pre-registered on the Open Science Framework (https://osf.io/pzkmh/).

***Population***

Studies needed to include children in the first three years of compulsory school. This age group was selected based on the emerging trend in the use of educational apps at the start of formal education (DfE, 2019) and evidence suggesting apps are more beneficial for children over 4 years (Herodotou, 2018).

As different countries have different age ranges for the first three years of compulsory education (e.g., children start preschool age 6 in Finland) or they may follow an ability-based system, rather than an age-based system (e.g., Malawi), the ages of children varied across the included studies. In England, the first three years of compulsory school refers to Reception to Year 2, where children are aged 4-7 years. In the USA, this refers to Kindergarten to Grade 2, where children are aged 5-8 years. In some European countries, such as Finland, this refers to preschool to Grade 2, where children are aged 6-9 years. Whereas, in Malawi, this refers to Standard 1- Standard 3, where children typically start school aged 6 years. However, while primary education is free in Malawi, it is not compulsory, and many children start school at different ages. The repetition of school years is also common. This means a child may be in an academic year that is not typically aligned with their chronological age.

If studies included children within the specified school years, as well as older children, only data relating to the year groups of interest were extracted. Corresponding authors were contacted where necessary. Studies that focused on typically developing children, children underachieving in mathematics and children with special educational needs were eligible for inclusion.

***Intervention***

Included studies needed to evaluate a named downloadable maths app(s), not just a particular feature of an app. Apps needed to be commercially available, or researcher developed. An educational app was defined as interactive software that is primarily designed to be used on a hand-held touch-screen tablet or smartphone device. It did not include computer-assisted instruction software e.g., web-based programmes that can only be used on computers but included apps that were connected to external/physical manipulatives.

### *Comparison*

To reflect the staged development of maths app evaluations, the review did not place any restrictions on the study design used for the thematic narrative synthesis. Included studies may or may not have included a control group comparison. No restriction was placed on the *type of* control group (e.g., active control or business-as-usual) or number of intervention groups. This is because different types of control group play an important role in understanding the impact and mechanisms of an intervention (Green et al., 2019).

### *Outcomes*

The primary unit of analysis was mathematical learning outcomes in response to a specific maths app(s) used at school or at home immediately following the intervention period. Other outcomes were also considered including cognitive development, enjoyment, and motivation (secondary outcomes). No restriction was placed on whether outcome measures were researcher developed or standardised.

### *Other Criteria*

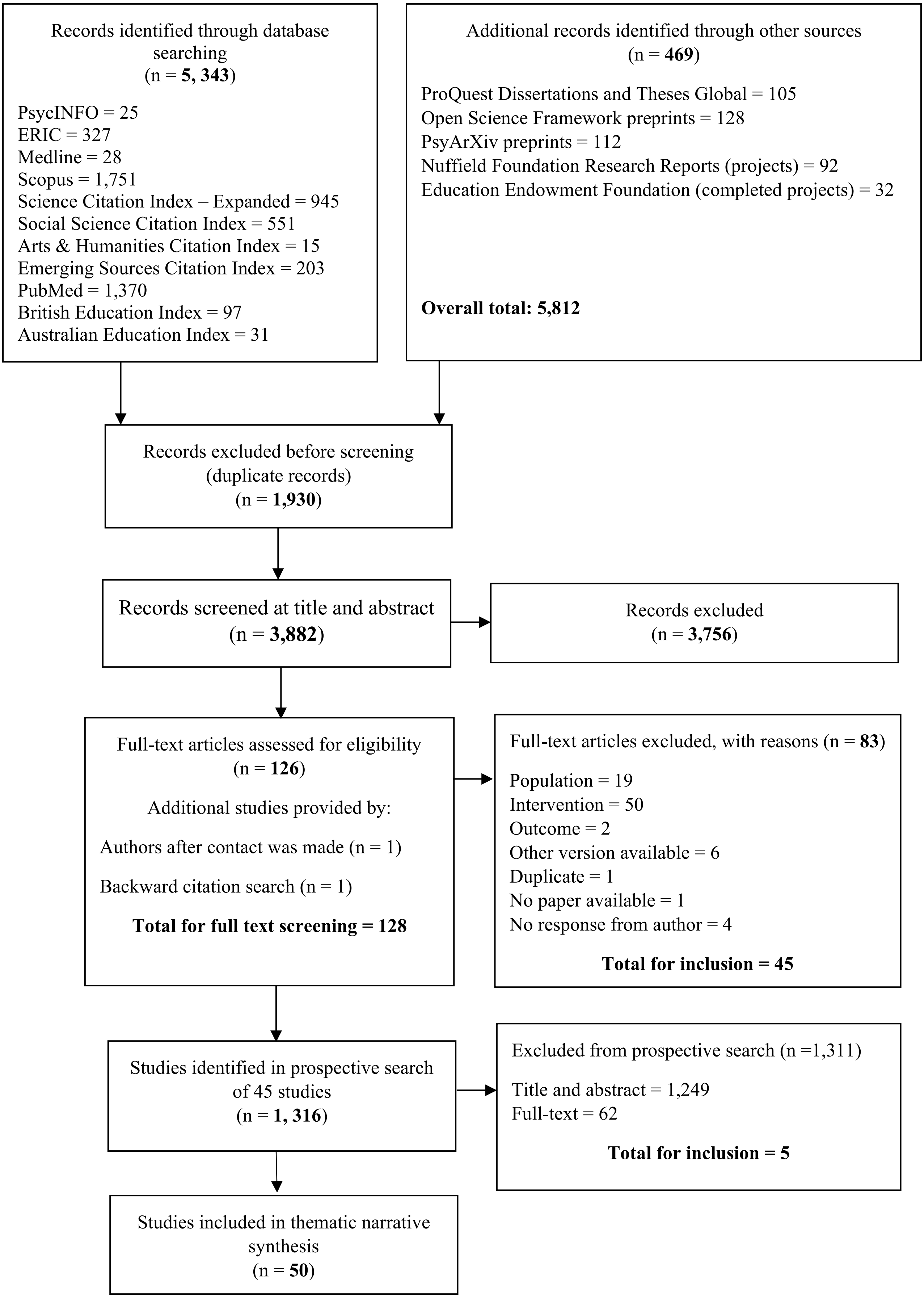
Studies also needed to be published between 2010-2021 and available in English. No restriction was placed on geographical location or the language of assessment. Studies were excluded if they reported on other systematic reviews of the literature, or did not report original data, such as commentary articles (e.g., Hubber et al., 2016).

## Record Screening

As outlined in the PRISMA Flow Diagram (Page et al., 2021; see Figure 1), the database searches returned 5,812 records. The titles and abstracts of the records returned in the database searches were imported to the web-based software EPPI-Reviewer where 1,930 duplicate records were identified and removed, resulting in 3,882 unique records. Unique records were screened based on the inclusion and exclusion criteria at two levels: title and abstract, followed by full-text. In total, 3,756 records were excluded at title and abstract level screening.

The remaining 126 studies were eligible for full-text screening. An additional two studies were included within the full-text screening, following contact with an author (n=1) and the backward citation search of existing systematic reviews (n=1). In line with the pre-registered protocol, full-text screening was also supplemented through contact with authors when more information to determine study eligibility was necessary (n = 22). Following this procedure, 83 records were excluded at full-text screening. Reasons for exclusion are provided in Figure 1. In total, 45 studies were identified for inclusion.

***Figure 1 PRISMA Flow Diagram of studies through the systematic review***



The prospective forward citation search of these 45 included studies returned 1,316 records. 1,249 records were excluded at title and abstract screening. The remaining 67 studies were assessed at full-text screening. An additional five studies were identified, resulting in a total of 50 studies included in the thematic narrative synthesis.

One reviewer (second author) was responsible for screening all records at both levels, with a second reviewer (first author) validating a random 20% sample (at both levels) to reduce potential error and bias in the screening outcomes. When conflicting decisions were apparent, discussions were held between the two reviewers until a consensus was reached. Inter-rater reliability between the two reviewers was calculated and reflected excellent agreement ($\kappa = 0.83$).

### Quality Assessment

The quality of each included study was formally assessed using the study quality checklists developed by Kmet et al. (2004). Quantitative studies were scored (0 for no, 1 for partial, and 2 for yes) on 14 criteria, such as sampling strategy, appropriateness of study design, and group allocation procedures. Individual studies were then given an overall quality score (total score divided by the highest possible score of 28). Following Kmet et al.'s (2004) guidelines, studies were rated as low (0 – 0.44), moderate (0.45 – 0.69), and high (0.70 – 1.00) quality.

The same procedure was implemented for qualitative studies using 10 criteria, including systematic data analysis, reflexivity, and connection to a theoretical framework. If studies included both quantitative and qualitative components, both scoring checklists were used, and an overall quality score was derived from the mean average of the two (quantitative and qualitative) scores. Study quality scores did not affect inclusion in this review; all studies that met the inclusion criteria were subject to data extraction and thematic narrative synthesis.

Attrition rates and the reliable and consistent reporting of data within included studies were also used as indicators of study quality.

## Results and Discussion

### Overview of Studies

Overall, the 50 studies evaluated 77 maths apps with 23,981 children (based on the final analysed sample; min = 1, max = 17,648 children per study) across 18 countries. Just over half of the studies (n = 26) were conducted in the USA. Experimental methods were the most common designs, with 20 RCTs and 13 QEDs. Most studies focused on mathematical learning outcomes (n = 47) and were conducted with typically developing children (n = 43) in the classroom (n = 46). Table 1 summarises the 50 included studies.

*Table 1 Summary of the 50 included studies*

| Study | Country, Setting (Implementer) | Final Sample (*n*, age) | App(s) | Method | Controls | Primary Maths Outcome Measure | Time on Task (minutes) | Overall results | Primary Effect Size* | Quality Score | Attrition Rate % |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ahmad et al. (2014) | Malaysia, School (NR) | SEND (5, 9 years) | *MathDS* | Qualitative | No control | Researcher developed | NR | Mixed | NR | .55 | 0 |
| Berggren & Hedler (2014) | Sweden, School (Practitioner) | TD (30, 4-5 years) | *CamQuest* | Qualitative | No control | NA- enjoyment only | 180 | Positive towards app | NA | .45 | 0 |
| Berkowitz et al. (2015) | USA, Home (Parent) | TD (278, 6-7 years) | *BedTime Math* | RCT | Active-other educational apps | Standardised-Woodcock-Johnson-III | NR | Positive towards app | .82 | .89 | 52.6 |
| Broda et al. (2019) | USA, School (Practitioner and researcher) | TD (18, 4-5 years) | *Fingu* | Single case design | No control | In-app data | 400 | Positive towards app | NA | .64 | 0 |
| Bullock et al. (2017) | USA, School (Researcher) | TD (19, 4-5 years) | 3 apps | Mixed methods | No control | Researcher developed | 30-40 | Mixed | NA | .67 | 0 |
| Cary et al. (2020) | USA, School (Practitioner) | TD/ LA (114, 5-6 years) | *KinderTek* | QED | Multiple | Standardised-ASPENS | 298 | Positive towards app | .88 | .68 | 11.6 |
| Cornu et al. (2019) | Luxembourg School (Researcher) | TD (125, 5-6 years) | *MaGrid* | RCT | BAU | Researcher developed | 400 | Mixed | .71 | .75 | 4.6 |
| Ginsburg et al. (2019) | USA, School (Researcher) | TD (1, 4 years) | *MathemAntics* | Qualitative | No control | Researcher developed | 24 | Positive towards app | NA | .60 | 0 |

| Study | Country, Setting (Implementer) | Final Sample (*n*, age) | App(s) | Method | Controls | Primary Maths Outcome Measure | Time on Task (minutes) | Overall results | Primary Effect Size* | Quality Score | Attrition Rate % |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Griffith et al. (2019) | USA, Home (Parent) | TD (22, 4-5 years) | 7 apps | RCT | Active-placebo | Standardised-TEMA-3 | NR | Positive towards app | .87 (.84) | .79 | 0 |
| Grimes et al. (2020) | USA, School (Practitioner) | TD (46, 5-6 years) | *Native Numbers* | RCT | Active – non-digital | Standardised-Number Sense Screener | 660 | Mixed | 1.10 (1.08) | .79 | 0 |
| Hasanah et al. (2017) | Japan, School (Practitioner) | TD (39, 6-7 years) | *Monsakun* | Single case study | No control | In-app data | 70 | Positive towards app | NA | .46 | 0 |
| Hassler Hallstedt et al. (2018) | Sweden, School (Practitioner) | LA (281, 8-9 years) | *Chasing Planets* | RCT | Multiple | Standardised-Math Battery | 1,122-1,698 | Mixed | 1.19 | .93 | .7 |
| Hieftje et al. (2017) | USA, School (NR) | TD/LA (133, 5-6 years) | *Knowledge Battle* | QED | Active-placebo | Standardised-KeyMath-3 | 480-720 | Mixed | NR | .61 | .7 |
| Hung et al. (2015) | Taiwan, School (Practitioner) | TD (43, 7-8 years) | *Motion Math: Hungry Fish* | QED | Active-other educational apps | Researcher developed | NR | Positive towards app | NR | .50 | 17.3 |
| Judd & Klingberg (2021) | Sweden, School (Practitioner) | TD (17,648, 6-8 years) | *Vektor* | RCT | Active-other educational apps | Researcher developed/ in-app data | 720-1,188 | Positive towards app | NR | .68 | 59.0 |
| Kalmpourtzis (2014) | Greece, School (Practitioner) | TD (17, 4-5 years) | *LadyBug Box* | Qualitative | No control | Researcher developed | NR | Positive towards app | NR | .60 | 0 |

| Study | Country, Setting (Implementer) | Final Sample (*n*, age) | App(s) | Method | Controls | Primary Maths Outcome Measure | Time on Task (minutes) | Overall results | Primary Effect Size* | Quality Score | Attrition Rate % |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Kosko & Ferdig (2016) | USA, Home (Parent) | TD (50, 4-5 years) | *Zorbit's Math Adventure for Preschool* | RCT | BAU | Researcher developed | NR | Positive towards app | 1.45 (1.43) | .61 | 0 |
| Kromminga & Codding (2020) | USA, School (Researcher) | LA (4, 7-8 years) | *Quizlet Plus* | Adapted alternating treatment design | Multiple | Researcher developed | 405- 510 | Mixed | NR | .61 | 0 |
| Lee & Choi (2020) | Tanzania, School (Researcher) | TD (61, 6-10 years) | *KitKit School* | RCT | BAU | Researcher developed | NR | Mixed | .60 | .68 | 50.0 |
| Litster et al. (2019) | USA, School (Researcher) | TD (65, 5-8 years) | *Montessori Numbers for Kids Base-10 Blocks* | Qualitative | No control | NA-engagement only | 30-40 | Mixed | NA | .90 | 0 |
| Mattoon et al. (2015) | USA, School (Practitioner) | TD (24, 4-5 years) | 5 apps | QED | Active – non-digital | Standardised-TEMA-3 | 180 | No difference | 1.09 (1.05) | .68 | 0 |
| Miller (2018) | Canada, School (Researcher) | TD (13, 4-5 years) | 15 apps | RCT + IPE | Active – non-digital | Researcher developed | 200 | No difference | .05 (.05) | .72 | 0 |
| Moyer-Packenham et al. (2016) | USA, School (Researcher) | see Litster et al. (2019) | 11 apps | Mixed methods | No control | Researcher developed | see Litster et al. (2019) | Mixed | NA | .90 | NA |
| Nunes et al. (2019) | England, School (Practitioner) | LA (1,089, 5-6 years) | *Onebillion Maths 3-5 and Maths 4-6* | RCT + IPE | BAU | Standardised-PTM5/6 | 1,440 | Positive towards app | NR | .87 | 3.1 |
| Outhwaite et al. (2017) Study 1 | England, School (Practitioner) | TD (26, 4-5 years) | *Onebillion Maths 3-5 and Maths 4-6. Count to 10* | QED | No control | Researcher developed | 900 | Positive towards app | 1.01 (.98) | .71 | 0 |

| Study | Country, Setting (Implementer) | Final Sample (*n*, age) | App(s) | Method | Controls | Primary Maths Outcome Measure | Time on Task (minutes) | Overall results | Primary Effect Size* | Quality Score | Attrition Rate % |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | *and Count to 20* | | | | | | | | |
| Outhwaite et al. (2017) Study 2 | England, School (Practitioner) | TD (18, 4-5 years) | *Onebillion Maths 3-5 and Maths 4-6* | QED | No control | Researcher developed | 1,950 | Positive towards app | 1.32 (1.26) | See Study 1 | 0 |
| Outhwaite et al. (2017) Study 3 | England, School (Practitioner) | LA (27, 5-7 years) | *Onebillion Maths 3-5 and Maths 4-6* | QED | No control | Researcher developed | 1,950 | Positive towards app | 1.81 (1.76) | See Study 1 | 0 |
| Outhwaite et al. (2017) Study 4 | England, School (Practitioner) | TA/LA (27, 4-5 years) | *Onebillion Maths 3-5 and Maths 4-6* | QED | BAU | Researcher developed | 1,200 | Positive towards app | 3.34 (3.24) | See Study 1 | 0 |
| Outhwaite et al. (2018) | England, School (Practitioner) | TD/LA (389, 4-5 years) | *Onebillion Maths 3-5 and Maths 4-6* | RCT | Multiple | Standardised-PTM5 | 1,800 | Positive towards app | .78 | .86 | 15.6 |
| Outhwaite et al. (2019) | England, School (Practitioner) | see Outhwaite et al. (2018) | *Onebillion Maths 3-5 and Maths 4-6* | IPE | Outhwaite et al. (2018) | NA-implementation only | NA | NA | NA | .85 | NA |
| Outhwaite et al. (2020) | Brazil, School (Practitioner) | TD (61, 5-6 years) | *Onebillion Maths 3-5 and Maths 4-6* | QED | Multiple | Standardised-EGMA | 800 | Positive towards app | 1.46 | .71 | 1.6 |
| Parks & Tortorelli (2020) | USA, School (Practitioner) | TD (298, 5-6 years) | 9 apps | RCT + IPE | Active-other educational apps | Standardised-AIMSweb test of Early Numeracy | 600-1,350 | No difference | NR | .75 | 43.3 |
| Pecora (2015) | USA, School (Practitioner) | SEND (6, 5-6 years) | *GoMath!* | Mixed methods | No control | Researcher developed | NR | Mixed | NR | .70 | 0 |

| Study | Country, Setting (Implementer) | Final Sample (*n*, age) | App(s) | Method | Controls | Primary Maths Outcome Measure | Time on Task (minutes) | Overall results | Primary Effect Size* | Quality Score | Attrition Rate % |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pires et al. (2019) | Uruguay, School (Researcher) | TD (60, 6-7 years) | *BrUNO* | QED | Multiple | Standardised-TEMA-3 | 260 | Mixed | .40 | .64 | 6.3 |
| Pitchford (2015) | Malawi, School (Practitioner) | TD (283, 6-13 years) | *Onebillion Maths 3-5 and Maths 4-6. Count to 10 and Count to 20* | RCT | Multiple | Researcher developed | 600-1,200 | Positive towards app | 1.80 | .71 | 11.0 |
| Pitchford et al. (2018) | Malawi, School (Practitioner) | SEND (32, 7-11 years) | *Onebillion Maths 3-5 and Maths 4-6* | Mixed methods | No control | In-app data | NR | Positive towards app | NA | .90 | 3.0 |
| Pitchford et al. (2019) | Malawi, School (Practitioner) | TD (256, 5-11 years) | *Onebillion Maths 3-5 and Maths 4-6* | QED | BAU | Standardised-EGMA | 540 | Positive towards app | NR | .89 | 79.0 |
| Ramani et al. (2019) | USA, School (NR) | TD (148, 5-6 years) | *The Great Race* | RCT | Multiple | Researcher developed | 100-150 | Positive towards app | .14 | .71 | 9.2 |
| Schacter et al. (2016) | USA, School (Researcher) | TD (86, 4-5 years) | *Math Shelf* | RCT | Active-other educational apps | Researcher developed | 180 | Positive towards app | .90 | .64 | 14.0 |
| Schacter & Jo (2016) | USA, School (Practitioner) | TD (162, 4-5 years) | *Math Shelf* | QED | BAU | Researcher developed | 300 | Positive towards app | 1.22 | .75 | 28.6 |
| Schacter & Jo (2017) | USA, School (Practitioner) | TD (378, 4-5 years) | *Math Shelf* | RCT | Active – non-digital | Researcher developed | 440 | Positive towards app | .20 | .79 | 12.7 |

| Study | Country, Setting (Implementer) | Final Sample (*n*, age) | App(s) | Method | Controls | Primary Maths Outcome Measure | Time on Task (minutes) | Overall results | Primary Effect Size* | Quality Score | Attrition Rate % |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Schaeffer et al. (2018) | USA, Home (Parent) | TD (195, 6-9 years) | *BedTime Math* | RCT | Active-other educational apps | Standardised-Woodcock-Johnson-III | NR | Positive towards app | NR | .82 | 66.8 |
| Schenke et al. (2020) | USA, School (Researcher) | TD (99, 4-5 years) | *MeasureUp!* | RCT | Multiple | Researcher developed | 240- 360 | Positive towards app | .65 | .68 | 2.0 |
| Spencer (2013) | UAE, School (NR) | TD (114, 4-5 years) | *Know Number Lite* | QED | BAU | Researcher developed | 100 | Positive towards app | .53 | .54 | 28.8 |
| Stacy et al. (2017) | USA, School (Practitioner) | TD (Not reported) | *IXL* | Qualitative | No control | Researcher developed | 200 | Mixed | NR | .75 | NA |
| Stubbé et al. (2016) | Sudan, School (Practitioner) | TD (703, 7-9 years) | *E-Learning Sudan* | QED | BAU | Researcher developed | 5,400 | Positive towards app | NR | .68 | 23.3 |
| Swicegood et al. (2015) | USA, School (Practitioner) | TD (40, 6-7 years) | *Addimal Adventure; Splash Math 2nd Grade* | Mixed methods | No control | Researcher developed | 4,800 | Mixed | NR | .77 | 0 |
| Tucker et al. (2016) | USA, School (Researcher) | TD (33, 7-8 years) | 6 apps | Qualitative | No control | NA - engagement only | 30-40 | Positive towards app | NA | .70 | 0 |
| Vanbecelaere et al. (2020) | Belgium, School (Practitioner) | TD (222, 6-7 years) | *Number Sense Game* | RCT | BAU | Researcher developed | 300 | Mixed | 1.58 | .71 | 33.9 |
| Watts et al. (2016) | USA, School (Researcher) | see Litster et al. (2019) | 11 apps | Mixed methods | No control | Researcher developed | see Litster et al. (2019) | Positive towards app | NA | .72 | NA |

| Study | Country, Setting (Implementer) | Final Sample (*n*, age) | App(s) | Method | Controls | Primary Maths Outcome Measure | Time on Task (minutes) | Overall results | Primary Effect Size* | Quality Score | Attrition Rate % |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Wu (2020) | USA, School (Researcher) | TD/LA (56, 4-5 years) | *MathAntics* | RCT | BAU | Researcher developed | 180 | Positive towards app | 1.89 | .68 | 11.1 |
| Zander et al. (2016) | Germany, School (NR) | TD (37, 8-11 years) | *Rotate It!* | QED | Active – non-digital | Researcher developed | NR | Mixed | NR | .64 | 27.5 |
| Zhang et al. (2020) | China, School (Practitioner) | TD (65, 8-9 years) | *Motion Math and Slice Fractions* | QED | BAU | Researcher developed | 120 | No difference | 2.31 | .64 | 14.5 |

**Within-subject effect sizes (Cohen's d) on primary mathematical outcome. Hedge's g correction applied for study samples equal to or less than 50 (reported in parenthesis).*

*BAU= business-as-usual; IPE = implementation process evaluation; LA = low achievers; NA = not applicable; NR= not reported; QED = quasi-experimental design; RCT = randomised control trial; SEND = special educational needs and disabilities; TD = typically developing.*

**Quality of Included Studies (RQ1)**

As shown in Table 1, across the 50 included studies, the quality ratings, relative to the study design, ranged from .45 to .93, with an average of .71. Twenty-seven of the included studies were classified as high quality, based on Kmet et al.'s (2004) criteria. Attrition rates were also relatively low across the included studies (see Table 1). Only seven studies reported attrition rates greater than 30% (Berkowitz et al., 2015; Judd & Klingberg, 2021; Lee & Choi, 2020; Parks & Tortorelli, 2020; Pitchford et al., 2019; Schaeffer et al., 2018; Vanbecelaere et al., 2020). Forty studies with a quantitative component were rated as using appropriate and justified analytic methods.

However, the quality of reporting standards across the 50 studies was relatively poor. As shown in Table 1, 15 studies did not sufficiently report data to enable calculations of within-subject effect sizes (Cohen's *d*) on children's learning gains in response to the maths app interventions. Similarly, intervention usage was not reliably or consistently reported across studies.

**Overview of Methods Used (RQ2)**

Experimental methods were the most common designs (20 RCTs and 13 QEDs) with mathematics as the primary outcome measure. These studies used a range of control groups, including business-as-usual (n = 12), active controls (n = 13), and multiple control groups (n = 8). None of the included studies included post-test only designs. Experimental designs are considered the gold standard form of evidence for understanding the efficacy of an intervention (Green et al., 2019). Implementation process evaluations (n = 4) are also valuable for providing insights into how the maths app interventions are most effectively implemented (Connolly et al., 2018) and feasibility studies (n = 7) are vital for establishing the viability of an intervention prior to scaling and investing in experimental studies (Green et al., 2019).

***Business-As-Usual Control Groups***

Of the 12 experimental studies with a business-as-usual control group, eight reported positive results[1] in favour of the maths app interventions compared to standard mathematical practice (Kosko & Ferdig, 2016; Nunes et al., 2019; Outhwaite et al., 2017 [Study 4 only]; Pitchford et al., 2019; Schacter & Jo, 2016; Spencer, 2013; Stubbé et al., 2016; Wu, 2020). Standard mathematical practice in these studies incorporated mathematical instruction typical for the classroom context and included, whole class teacher-led instruction, small group or one-to-one activities, play-based learning, and traditional, physical manipulatives. However, the specific mathematical activities completed by the business-as-usual control group were not explicitly differentiated.

[1] The description of overall results was based on the original studies reported effect sizes and *p* values.

The remaining four studies comparing the maths app intervention to standard practice reported mixed findings. Specifically, the training effects of the maths apps were limited to the targeted skills, such as number line estimation, magnitude comparison, and spatial skills, and did not transfer to broader mathematical competence (Cornu et al., 2019; Lee & Choi, 2020; Zhang et al., 2020) or maths anxiety (Vanbecelaere et al., 2020).

***Active Control Groups***

Thirteen studies used a range of active control groups in an experimental design (RCTs and QEDs). Five studies compared the maths app interventions to non-digital maths interventions (Grimes et al., 2020; Mattoon, 2015; Miller, 2018; Schacter & Jo, 2017; Zander et al., 2016). Importantly, unlike in the business-as-usual control groups, the different mathematical activities, and thus the potential mechanisms for learning, were differentiated. There were mixed outcomes across these five studies.

When comparing the maths app intervention to one-to-one and small peer group instruction, results showed immediate, near-transfer benefits for mathematical performance (Grimes et al., 2020; Schacter & Jo, 2017). However, the maths app intervention effects did not transfer to mathematical language skills (Grimes et al., 2020). The benefits of app-based maths instruction were also seen when children used a paper-based version of the task, and then the app-based version. Training effects were not observed when the app was used first, followed by the paper-based task (Zander et al., 2016).

Furthermore, learning gains with the maths app interventions were not statistically significant from gains made with traditional manipulatives and play-based learning (Mattoon, 2015; Miller, 2018). However, in both studies multiple apps were evaluated and the results per app were not disaggregated. As such, it is not clear which apps may or may not have supported learning. In addition, the frequency and duration of use for each app were not reported, thus it is also unclear which apps children engaged with and for long, which may have also impacted the observed results.

Eight studies compared the maths app intervention to other apps including maths (Hung et al., 2015; Judd & Klingberg, 2021; Schacter et al., 2016), reading/literacy (Berkowitz et al., 2015; Schaeffer et al., 2018), science (Parks & Tortorelli, 2020), and no educational or mathematical content (i.e., placebo) (Griffith et al., 2019; Hieftje et al., 2017). As the touch-screen tablet mode of delivery is consistent across the intervention and control groups, it enables the quality of the mathematical content to be assessed, rather than just the use of technology (i.e., a novelty effect).

Of these eight studies, six reported positive results in favour of the maths app interventions compared to the active controls (Berkowitz et al., 2015; Griffith et al., 2019; Hung et al., 2015; Judd & Klingberg, 2021; Schacter et al., 2016; Schaeffer et al., 2018). One study reported mixed results (Hieftje et al., 2017). Although the maths app intervention was shown to support children's numeration skills targeted by the app, the training effect did not transfer to other areas of mathematical development, including measurement, numerical operations, and problem solving (Hieftje et al., 2017).

In contrast, Parks and Tortorelli (2020) found no difference in mathematical learning gains between the intervention (nine maths apps) and active control group (five other educational apps covering literacy, science, and maths). However, in this study the effects of each app were not disaggregated and both groups received some form of app-based mathematics instruction.

***Multiple Control Groups***

Eight experimental studies (RCTs and QEDs) included more than one control group. Two studies included a business-as-usual control group, as well as an active control group, consisting of other apps with reading (Hassler Hallstedt et al., 2018) or no educational content (Pitchford, 2015). This design enabled the effects of the maths app intervention, as a form of effective instruction, to be disentangled from maturation (business-as-usual) and the technology-based delivery (active controls) within the context of the same study. Significant, near-transfer learning gains in response to the maths app interventions were observed in both studies. While Pitchford (2015) found these benefits transferred to mathematical conceptual knowledge, not targeted by the intervention, Hassler Hallstedt et al. (2018) found no significant improvements in far-transfer outcomes between the different groups.

Ramani et al. (2019) also compared the maths app intervention to two active control groups, consisting of an app-based working memory and reading game. As domain-specific mathematical skills and domain-general working memory skills are closely intertwined (Allen et al., 2019), this design allowed closer examination of near and far-transfer effects of app-based instruction. Results showed children using the maths apps improved in numerical knowledge (near-transfer) compared to both active controls, as well as working memory (far-transfer), compared to the active controls using the reading game.

Five experimental studies used multiple control groups to compare different forms of implementation of the maths app intervention to understand how learning gains could be maximised. This included variations in time spent learning maths (Cary et al., 2020; Outhwaite et al., 2018), the language of instruction (Outhwaite et al., 2020), integration with other mathematical instructional materials (Pires et al., 2019), and additional links with parents at home (Schenke et al., 2020).

Results showed children's learning gains with maths apps were optimised when children used the apps for a longer, rather than shorter, duration (Cary et al., 2020). However, when focusing on achieving a balance between using the maths apps and other classroom activities, other research showed an extra 30 minutes per day learning mathematics, in the form of the app-based instruction had no additional benefit for children's basic and higher-order maths skills, compared to when the maths app intervention was integrated into the school day and thus did not take away from time spent with other subjects (Outhwaite et al., 2018).

Children also benefitted more from the maths app intervention, when it was implemented in their first language (L1), compared to their second language (L2; Outhwaite et al., 2020), as well as when it was used in combination with physical play manipulatives, rather than virtual play manipulatives (Pires et al., 2019). However, there was no added benefit of implementing the maths app intervention with children alongside a parent-focused companion app, compared to the maths app only (Schenke et al., 2020).

***Implementation Process Evaluations***

Implementation process evaluations (IPE) are typically used in combination with experimental methods (RCTs and QEDs) to further examine how a particular intervention works and under what circumstances (Humphrey et al., 2016; Pawson & Tilley, 1997). Within the 50 studies identified in this review, four RCTs included IPEs, which ranged from descriptions of implementation (Miller, 2018) to highlighting important implications for understanding how maths apps can be most effectively implemented to maximise children's learning outcomes (Nunes et al., 2019; Outhwaite et al., 2019) and understanding null results (Parks & Tortorelli, 2020).

Outhwaite et al. (2019) conducted observations of the maths app intervention sessions and interviews with the participating teachers and found 41% of the variance in children's learning outcomes (reported in Outhwaite et al., 2018) were explained by the established routine within the participating school contexts. This included implementing the intervention at a consistent time each day, having a member of staff responsible for the intervention implementation, having well-organised equipment, and having dedicated space within the classroom and a seating plan for the maths app intervention. In the larger-scale efficacy school-level RCT of the same maths app intervention, Nunes et al. (2019) triangulated qualitative data from observations, interviews, and questionnaires with teaching assistants. Results showed maths learning outcomes were greater when the teaching assistants perceived their role as an educator or guide, compared to an observer during implementation.

Similarly, when explaining the null findings, Parks and Tortorelli (2020) suggested that factors, such as lack of integration into the classroom setting, combined with a relatively short time on task (on average, 13 minutes per week) may have limited the success of the maths app intervention. Teachers also reported challenges with the logistics of charging tablet devices, difficulty downloading apps, and the need for training on how to integrate and implement the maths apps.

***Feasibility Studies***

In the current review, there were seven feasibility studies with typically developing children with mixed results. These studies are valuable for establishing if the maths app interventions were viable (Kalmpourtzis, 2014; Stacy et al., 2017) or not (Swicegood, 2015), and provided the initial evidence base needed for further larger-scale experimental studies (Outhwaite et al., 2017).

These studies also provide useful insights into age-related differences in children's engagement behaviours with the maths apps. Moyer-Packenham et al. (2016) and Watts et al. (2016) found 5-8-year-old children's efficiency and accurate performance in different mathematical skills significantly improved over time, in response to using a selection of 11 maths apps. In contrast, Bullock et al. (2017) assessed the feasibility of three of these apps with 4-year-old children and found most children did not make any significant progress in seriation and counting skills.

**Overview of Populations Used (RQ2)**

Most studies focused on the use of educational maths apps with typically developing children of all attainment levels (n = 43). A small number of studies considered whether educational maths apps can support children underachieving in mathematics (n = 8) and children with special educational needs and disabilities (n = 3).

***Children Identified as Underachieving in Mathematics***

Eight studies included a focus on children identified as underachieving in mathematics (Cary et al., 2020; Hassler Hallstedt et al., 2018; Hieftje et al., 2017; Kromminga & Codding, 2020; Nunes et al., 2019; Outhwaite et al., 2017 [Study 3 and Study 4 only]; Outhwaite et al., 2018; Wu, 2020). The methods used to identify these children varied across studies, which has important implications for the internal validity of the findings.

Four studies included exploratory sub-sample analyses to examine which children benefited the most from the maths app intervention (Cary et al., 2020; Hieftje et al., 2017; Outhwaite et al., 2018; Wu, 2020). In these studies, children were identified as underachieving in mathematics based on a statistical cut-off point applied to their pre-test maths assessment score. For example, Cary et al. (2020) analysed a sub-sample of participants when evaluating *KinderTEK*. 5-6-year-old children performing below age-expected levels in mathematics (n = 38) made significantly stronger learning gains with the maths app intervention (Cohen's $d$ = 1.97), compared to the control group. In contrast, children considered at-risk of being below age-expected levels in mathematics (n = 50) showed similar levels of improvement across the two experimental conditions. Similarly, both Outhwaite et al., (2018) and Wu (2020) found children identified as underachieving in mathematics made significantly greater improvements with *onebillion Maths 3-5 and Maths 4-6* (Cohen's $d$ = 4.03), and *MathemAntics* (Cohen's $d$ = 2.11), respectively, compared to their higher-attaining peers. Hieftje et al. (2017) also found the same pattern of results when examining *Knowledge Battle*. Collectively, this evidence may suggest that the maths app interventions were most beneficial for children underachieving in mathematics. However, this method of identifying children as underachieving is not considered rigorous, due to the regression to the mean phenomenon and can threaten the internal validity of the findings (Barnett et al., 2004).

With a more rigorous approach, four studies used different measures to identify children as underachieving in mathematics, relative to their peers and evaluate the primary outcome of the maths app intervention. In the current review, two studies used teacher reports (Outhwaite et al., 2017 [Study 3 and Study 4 only]; Nunes et al., 2019). One study used a screening assessment tool (Hassler Hallstedt et al., 2018) but this tool also included items used in the primary outcome measure that may have also been impacted by the regression to the mean phenomenon. One study used a combination of teacher reports and an independent screening assessment tool (Kromminga & Codding, 2020). Together, these four studies found children using the maths app intervention made significant gains in maths skills compared to the control conditions. However, in Kromminga and Codding's (2020) study, children underachieving in mathematics also benefited from the paper-based version of the app intervention.

### *Special Educational Needs and Disabilities*

Three studies assessed the feasibility of maths app interventions with children with special educational needs and disabilities (SEND), including Down Syndrome and attention deficit hyperactivity disorder (ADHD), as well as vision loss, and emotional, behavioural, communication and learning difficulties (Ahmad et al., 2014; Pecora, 2015; Pitchford et al., 2018).

Within the included studies, most children with SEND showed improvements in mathematical skills in response to app-based instruction. However, the average progress rate was twice as slow for children with SEND relative to their mainstream peers. Moreover, the extent of children's additional needs significantly predicted their progress rate; children with hearing and/or language difficulties made slower progress compared to other SEND profiles (Pitchford et al., 2018). Further observations also indicated children with SEND sometimes faced challenges engaging with the maths apps, such as becoming frequently distracted, lack of interest, unfocused, and interrupted by schedule changes (Ahmad et al., 2014; Pecora, 2015).

## Overview of Outcomes Used (RQ2)

### *Mathematical Outcomes*

Included studies mostly focused on mathematical learning outcomes (n = 46), which were primarily assessed with researcher developed outcome measures (n = 32). Only 14 studies used a standardised assessment of mathematical attainment as the primary outcome measure (See Table 1).

Twenty-four studies sufficiently reported data to calculate within-subject effect sizes (Cohen's *d*) on children's learning gains in response to the maths app interventions and afford comparisons on the magnitude of effect sizes across studies. Effect sizes ranged from .05 to 3.34, with 11 studies reporting Cohen's *d* effect sizes equal to or greater than one (Grimes et al., 2020; Hassler Hallstedt et al., 2018; Kosko & Ferdig, 2016; Mattoon et al., 2015; Outhwaite et al., 2017; Outhwaite et al., 2020; Pitchford, 2015; Schacter & Jo, 2016; Vanbecelaere et al., 2020; Wu, 2020; Zhang et al., 2020).

Although most of these 11 studies used a RCT design (n = 6; Grimes et al., 2020; Hassler Hallstedt et al., 2018; Kosko & Ferdig, 2016; Pitchford, 2015; Vanbecelaere et al., 2020; Wu, 2020), studies with small sample sizes, such as those with fewer than 250 students, are more likely to produce inflated effect sizes than larger sample sizes (Cheung & Slavin, 2013). Of these 11 studies, nine had final sample sizes less than 250 children (Grimes et al., 2020; Kosko & Ferdig, 2016; Mattoon et al., 2015; Outhwaite et al., 2017; Outhwaite et al., 2020; Schacter & Jo, 2016; Vanbecelaere et al., 2020; Wu, 2020; Zhang et al., 2020). Hedge's *g* corrections were applied for studies with sample sizes equal to or less than 50 (Lin, 2018; see Table 1), and these effect sizes remained above one (Grimes et al., 2020; Kosko & Ferdig, 2016; Mattoon et al., 2015; Outhwaite et al., 2017).

Effect sizes can also be impacted by the outcome measure used. Of these 11 studies, seven used researcher developed measures as the primary outcome measure (Kosko & Ferdig, 2016; Outhwaite et al., 2017; Pitchford, 2015; Schacter & Jo, 2016; Vanbecelaere et al., 2020; Wu, 2020; Zhang et al., 2020), which are more likely to have larger effect sizes, compared to standardised assessment tools (Cheung & Slavin, 2013).

Only three studies used a RCT design with a large sample size (greater than 250 children) and a standardised measure of mathematical attainment (Berkowitz et al., 2015; Hassler Hallstedt et al., 2018; Outhwaite et al., 2018). Hassler Hallstedt et al. (2018) evaluated the *Chasing Planets* app for 19-20 weeks with 281 low-achieving students and reported an effect size of 1.19. However, there are important caveats to the internal validity of these findings due the crossover of items between the outcome measure and the measure used to identify children as underachieving in mathematics. As such, the large effect size may be inflated.

The two other studies with rigorous experimental designs reported effect sizes less than one. Berkowitz et al. (2015) evaluated the *Bedtime Math* app for 22 weeks with 278 children and reported an effect size of .82. Outhwaite et al. (2018) evaluated *onebillion Maths 3-5 and Maths 4-6* for 12 weeks with 389 children and reported an effect size of .78.

### *Long-Term Mathematical Outcomes*

Most included studies focused on mathematical outcomes measured immediately after the intervention period (n = 42). In the current review, five studies included a delayed post-test assessment of children's mathematical attainment with mixed results. Three studies showed the mathematical gains on the primary outcome measure were sustained at follow-up (Hassler Hallstedt et al., 2018; Outhwaite et al., 2017; Schaeffer et al., 2018). The timing of delayed post-tests in these studies ranged from five months (Outhwaite et al., 2017) to two years (Schaeffer et al., 2018). Two studies found the effects of the maths app intervention faded after one to (Ramani et al., 2019) two months (Vanbecelaere et al., 2020).

### *Time on Task*

Thirty-nine studies reported approximate time on task for the maths app interventions (see Table 1). For studies that reported time on task as a range (e.g., 600-1,350 minutes in Parks & Tortorelli, 2020), the intervention usage was estimated based on the mean average of the two reported values. On average, children used the maths apps for 797 minutes (13 hours), which ranged from 24 minutes

(Ginsburg et al., 2019) to 5,400 minutes (90 hours; Stubbé et al., 2016). Within these 39 studies, 23 studies provided sufficient data to calculate within-group effect sizes. A small, positive, but not statistically significant relationship was observed between intervention usage and within-subject effect sizes on mathematical primary outcomes ($r = .30$, $p = .160$).

***Use of In-App Data***

Four studies used log data automatically collected by the maths apps. For example, Hasanah et al. (2017) examined children's in-app behaviour, including the average number of steps made by children and the number of errors made. Pitchford et al. (2018) collected in-app data on progress through the app content and examined how this related to learning gains. Two studies used in-app data collection to measure children's mathematical skills (Broda et al., 2019; Judd & Klingberg, 2021). This innovative use of data collection demonstrates proof-of-concept for conducting intervention studies remotely and for improving the quality of collecting data, particularly relating to usage and children's engagements with the apps.

***Non-Attainment Outcomes***

Alongside mathematical learning outcomes, 13 studies also included a range of non-attainment outcome measures, including enjoyment, preference modality, intrinsic motivation, and flow experience. Five studies reported that children mostly enjoyed using the evaluated maths app interventions (Berggren & Hedler, 2014; Ginsburg et al., 2019; Hieftje et al., 2017; Nunes et al., 2019; Pitchford et al., 2018). Furthermore, Hieftje et al. (2017) observed a positive relationship between enjoyment and children's learning gains in mathematics.

While Griffith et al. (2019) also found children enjoyed using the selection of seven maths apps, enjoyment ratings were not significantly different from that of other apps with no educational content. No significant differences were also observed in children's preferences between app-based learning and non-digital based maths interventions (Kromminga & Codding, 2020; Pecora, 2015; Swicegood, 2015).

Four studies examined the impact of educational maths apps on children's intrinsic motivation and flow experience (i.e., sustained attention) with mixed results. Hung et al. (2015) found greater flow experience, but not intrinsic motivation, with the maths app in the treatment group, compared to controls. The remaining three studies found no group differences in intrinsic motivation or flow experience (Grimes et al., 2020; Spencer, 2013; Zander et al., 2016).

**External Validity of the Current Evidence (RQ3)**

The external validity of the current evidence base was mixed. Although included studies were conducted in 18 countries, the majority were conducted in the USA (n = 26). Only eight studies examined the same maths app intervention in three different countries, including England (Nunes et al., 2019; Outhwaite et al., 2017; Outhwaite et al., 2018; Outhwaite et al., 2019), Brazil (Outhwaite et al., 2020), and Malawi (Pitchford, 2015; Pitchford et al., 2018; Pitchford et al., 2019).).

However, in most included studies the maths app intervention was implemented in the classroom learning environment (n = 46) by teaching practitioners (n = 26), rather than by researchers (n = 20) (see Table 1). But only four studies were implemented in the home learning environment by parents (Berkowitz et al., 2015; Griffith et al., 2019; Kosko & Ferdig, 2016; Schaeffer et al., 2018).

**Where are the Current Evidence Gaps? (RQ4)**

Across the 50 studies, eight key findings and associated gaps were identified in the evidence base. Table 2 illustrates the directness of the included studies aligned with the eight themes identified through the thematic narrative synthesis. These gaps should be addressed in future research to advance understandings of app-based mathematics instruction and ensure optimal learning outcomes for all children.

***Table 2 Directness of 50 included studies to themes identified in the gaps in the current evidence.***

| Study | Near- and far-transfer benefits | Children underachieving in mathematics | Special educational needs and disabilities | Role of age and language | Time on task and immediate and sustained learning gains | Innovative methods for data collection | Cross-cultural comparisons | Home learning environment |
|---|---|---|---|---|---|---|---|---|
| Ahmad et al. (2014) | | | ✓ | | | | | |
| Berggren & Hedler (2014) | | | | | | | | |
| Berkowitz et al. (2015) | | | | | | | | ✓ |
| Broda et al. (2019) | | | | | | ✓ | | |
| Bullock et al. (2017) | | | | ✓ | | | | |
| Cary et al. (2020) | | ✓ | | | | | | |
| Cornu et al. (2019) | ✓ | | | | | | | |
| Ginsburg (2019) | | | | | | | | |
| Griffith et al. (2019) | | | | | | | | ✓ |
| Grimes et al. (2020) | ✓ | | | | | | | |
| Hassler Hallstedt et al. (2018) | ✓ | ✓ | | | ✓ | | | |
| Hasanah et al. (2017) | | | | | | ✓ | | |
| Hieftje et al. (2017) | ✓ | ✓ | | | | | | |
| Hung et al. (2015) | | | | | | | | |
| Judd & Klingberg (2021) | | | | | | ✓ | | |
| Kalmpourtzis (2014) | | | | | | | | |
| Kosko & Ferdig (2016) | | | | | | | | ✓ |

| Study | Near- and far-transfer benefits | Children underachieving in mathematics | Special educational needs and disabilities | Role of age and language | Time on task and immediate and sustained learning gains | Innovative methods for data collection | Cross-cultural comparisons | Home learning environment |
|---|---|---|---|---|---|---|---|---|
| Kromminga & Codding (2020) | | ✓ | | | | | | |
| Lee & Choi (2020) | ✓ | | | | | | | |
| Litster et al. (2019) | | | | | | | | |
| Mattoon et al. (2015) | | | | | | | | |
| Miller (2018) | | | | | | | | |
| Moyer-Packenham et al. (2016) | | | | | | | | |
| Nunes et al. (2019) | | ✓ | | | | | ✓ | |
| Outhwaite et al. (2017) | | ✓ | | | ✓ | | ✓ | |
| Outhwaite et al. (2018) | ✓ | ✓ | | | | | ✓ | |
| Outhwaite et al. (2019) | | | | | | | ✓ | |
| Outhwaite et al. (2020) | | | | ✓ | | | ✓ | |
| Parks & Tortorelli (2020) | | | | | | | | |
| Pecora (2015) | | | ✓ | | | | | |
| Pires et al. (2019) | | | | | | | | |
| Pitchford (2015) | ✓ | | | | | | ✓ | |
| Pitchford et al. (2018) | | | ✓ | ✓ | | ✓ | ✓ | |
| Pitchford et al. (2019) | | | | | | | ✓ | |
| Ramani et al. (2019) | ✓ | | | | ✓ | | | |
| Schacter & Jo (2016) | | | | | | | | |
| Schacter & Jo (2017) | ✓ | | | | | | | |
| Schacter et al. (2016) | | | | | | | | |
| Schaffer et al. (2018) | | | | | ✓ | | | ✓ |
| Schneke et al. (2020) | | | | | | | | |
| Spencer (2013) | | | | | | | | |
| Stacy et al. (2017) | | | | | | | | |
| Stubbé et al. (2016) | | | | | | | | |
| Swicegood (2015) | | | | | | | | |
| Tucker et al. (2016) | | | | | | | | |
| Vanbecelaere et al. (2020) | ✓ | | | | ✓ | | | |
| Watts et al. (2016) | | | | | | | | |
| Wu (2020) | | ✓ | | | | | | |
| Zander et al. (2016) | | | | | | | | |
| Zhang et al. (2020) | ✓ | | | | | | | |

***Near- and Far-Transfer Benefits***

Nine experimental studies (RCTs or QEDs) explicitly examined differences between learning gains in mathematical skills targeted by the maths app intervention (near-transfer), and other relevant mathematical and cognitive outcomes, not included in the intervention (far-transfer) with mixed results. Three studies found the benefits of the maths app interventions transferred to other skills (far-transfer), including higher-order and conceptual mathematical skills, such as mathematical reasoning and problem solving (Outhwaite et al., 2018; Pitchford, 2015), as well as working memory (Ramani et al., 2019). Pitchford and Outhwaite (2019) also reported far-transfer benefits to children's attention skills, independent of mathematics, in a secondary analysis of data reported in Pitchford's (2015) pupil-level RCT in Malawi. This study was not included in the 50 studies, as it did not have mathematics as a primary outcome measure, and so fell beyond the scope of the eligibility criteria.

In contrast, six studies found the benefits of the maths app interventions did not transfer to broader mathematical skills, including more complex arithmetic, problem-solving, and measurement skills (Cornu et al., 2019; Hassler Hallstedt et al., 2018; Hieftje et al., 2017; Lee & Choi, 2020; Vanbecelaere et al., 2020). Training effects also did not transfer to children's mathematical language skills (Grimes et al., 2020) or maths anxiety (Vanbecelaere et al., 2020).

As these studies represent the best experimental evidence currently available, the mixed findings may indicate that variations in the breadth of the mathematical content and quality of the design features included in the evaluated apps may, in part, explain the variations in the observed near- and far-transfer training effects. However, no firm conclusions can be drawn based on the evidence to date. Further research is needed to systematically evaluate these underlying pedagogical features within the maths apps to identify potential mechanisms that underpin children's learning in this technology-based context.

Furthermore, it is vital that future studies include an appropriately powered sample size within a randomised experimental design to address the possibility of inflated effect sizes and provide rigorous evidence to inform education policy and practice (Cheung & Slavin, 2013). Future studies should also use standardised near- and far-transfer outcome measures that are sensitive to capture change over time but are also not biased towards the treatment group.

***Children Identified as Underachieving in Mathematics***

While there is emerging evidence that educational maths apps can provide supplementary learning opportunities for children identified as underachieving in mathematics, the quality and rigour of this evidence was varied.

Five studies used the same assessment tool to identify children underachieving in mathematics, as well as measure the outcomes of the intervention (Cary et al., 2020; Hassler Hallstedt et al., 2018; Hieftje et al., 2017; Outhwaite et al., 2018; Wu, 2020). This poses a threat to the internal validity of these findings, due to the possibility of regression to the mean (Barnett et al., 2004). This is because extreme values, such as a low mathematics score at pre-test may be due to measurement error and are more likely to be followed by a less extreme score closer to the participant's true mean when assessed again at post-test (Barnett et al., 2004; Stigler, 1997). This makes it very difficult to disentangle a regression to the mean effect from a genuine intervention effect (Yudkin & Stratton, 1996).

In contrast, only three studies in the current review used an external method for identifying children as underachieving in mathematics, which was completely independent from the outcome variable (Kromminga & Codding, 2020; Outhwaite et al., 2017 [Study 3 and Study 4 only]; Nunes et al., 2019). This helps to minimise the regression to the mean problem (Yudkin & Stratton, 1996), and thus greater confidence can be placed in these three studies, compared to the five using other identification methods.

However, within these three studies, only one used a RCT design with a business-as-usual control group and a large sample size (Nunes et al., 2019). As such, more large-scale rigorous intervention studies are needed to evaluate the use of educational maths apps with young children identified as underachieving in mathematics. The use of an assessment measure for identifying these children, independent of the outcome measure, should be considered best practice when conducting future research.

***Special Educational Needs and Disabilities***

Research with children with SEND was restricted to feasibility level evidence and no studies worked with children with dyscalculia. However, it is important to acknowledge that the inclusion criteria for the current review was restricted to the first three years of formal education. As many special educational needs are not formally diagnosed until children are older and the SEND code of practice ranges from 0-25 (DfE, 2020), there may be more developed evidence that fell beyond the scope of this review.

Nevertheless, further research is needed to examine the use of educational maths apps with children with specific mathematical difficulties, such as dyscalculia, Down Syndrome, and Williams Syndrome. For example, Sella et al. (2021) showed children with Down syndrome made significant improvements in specific numerical skills and mental calculation, in response to using the computer-based game, The Number Race, for 10 weeks, compared to a reading active control group. Although this study fell beyond the scope of the current review, it provides a useful demonstration of how interventions initially evaluated with typically developing children (Sella et al., 2016), can be translated for different population groups. As such, similar research specific to educational maths apps on touch-screen tablets is needed to identify how such interventions can be successfully implemented and adapted to meet the needs of different groups of children and ensure equitable access to effective and evidence-based mathematical instruction.

***Role of Age and Language***

Consistent with previous systematic reviews (Griffith et al., 2020; Herodotou, 2018), the current synthesis also suggests that younger children may face more barriers when accessing educational maths apps (e.g., Bullock et al., 2017). However, this chronological approach, does not necessarily capture how the suitability of educational maths apps may differ based on children's individual abilities. Two studies showed children's language skills influenced their progress through the maths app intervention, in that children with better proficiencies in the language of instruction made more progress through the maths apps, compared to those with weaker language skills (Pitchford et al., 2018; Outhwaite et al., 2020). Further studies conducted with the same maths app intervention in South Africa and England with children with English as an additional language found most enjoyed accessing the maths app and benefited from accessing the content in their first language (L1), with minimal impact on teacher time and resources (Pitchford et al., 2021). This study fell beyond the scope of the current review and was not included in the final sample of 50 studies. Overall, this evidence collectively suggests that as young children are still developing their language and vocabulary skills, it may be important to consider their strengths and weaknesses in these areas, when deciding if and which apps to use, rather than just their age.

Furthermore, different types of apps may be better suited to support children's developing language skills in the context of app-based maths instruction. For example, apps that are designed with a strong focus on social interaction (Hirsh-Pasek et al., 2015), may provide additional linguistic scaffolds that are not included in other types of apps that place an emphasis on individual use by the child. As such, future research should further examine the role of language and how this may interact with age, when children use different types and features of maths apps.

***Time on Task and Immediate and Sustained Learning Gains***

Although a small, positive, but not statistically significant relationship was observed between intervention usage (i.e., time on task) and learning outcomes ($r$ = .30), this result should be treated tentatively. This is because usage was not reliably and consistently reported across studies, and the quality of reporting standards across the 50 studies was relatively poor, which limited the calculations of within-subject effect sizes on learning outcomes.

In studies where usage was sometimes adequately reported, it gave insights into explaining the observed long-term mathematical outcomes, in response to the maths app interventions. Three studies reported learning gains that were maintained when assessed at a 5-month- to- 2-year- follow-up (Hassler Hallstedt et al., 2018; Outhwaite et al., 2017; Schaeffer et al., 2018). Importantly, in these studies, the maths app intervention was implemented over a sustained period, with time on task ranging from 900 to 1,698 minutes. In contrast, the two studies that found the effects of the maths app intervention faded at one to two months later, were implemented for a shorter duration of 100-300 minutes (Ramani et al., 2019; Vanbecelaere et al., 2020). Overall, this highlights the need for more reliable and consistent reporting of maths app usage, which can be supported by innovative methods for data collection.

### *Innovative Methods for Data Collection*

Previous systematic reviews on educational apps identified the lack of studies utilising in-app data on learner analytics (Herodotou, 2018). The current review identified four studies that used log data to understand children's in-app behaviour and examine learning outcomes (Broda et al., 2019; Hasanah et al., 2017; Judd & Klingberg, 2021; Pitchford et al., 2018). These studies demonstrate proof-of-concept for innovative methods for data collection with educational maths apps. Future studies should build on this evidence base to make effective use of the in-app data automatically generated by the maths apps, particularly for rigorously examining the relationship between usage and mathematical outcomes. This data could also be utilised to conduct high-quality intervention studies remotely, for example during periods of home learning and for working with hard-to-reach populations, such as those with SEND.

### *Cross-Cultural Comparisons*

Although 18 different countries were represented across the 50 included studies in the current review, only one maths app intervention was evaluated in different high- (England), middle- (Brazil), and low- (Malawi) income country contexts (Nunes et al., 2019; Outhwaite et al., 2017; Outhwaite et al., 2018; Outhwaite et al., 2019; Outhwaite et al., 2020; Pitchford, 2015; Pitchford et al., 2018; Pitchford et al., 2019). Across these eight studies, results were positive in favour of the maths app intervention, compared to a range of control groups. As such, it demonstrates that this maths app intervention can be effectively used to support children's mathematical development, particularly for girls in countries where gender differences in standard practice may typically hinder their learning progress (Pitchford et al., 2019). This has important implications for addressing global educational challenges, including issues faced because of schools around the world being closed for significant periods of time for most children during the Covid-19 pandemic. Data from the United Nations (2021) suggests that the impact of the pandemic has wiped out 20 years of educational gains, thus limiting progress towards the education Sustainable Development Goals. To continue supporting children's learning and development, particularly in response to Covid-19, there has been a greater emphasis on the implementation of children's education by parents at home, for which technology can play an important role. However, it is essential these interventions are effective, evidence based, and distributed in an equitable way. In response, further rigorous research is needed to evaluate educational maths apps around the world, particularly in the home learning environment.

### *Home Learning Environment*

The home learning environment plays a vital role in child development (Toth et al., 2020), yet parents typically engage in maths activities at home once a week, compared to every day for reading

(Organisation for Economic Co-operation and Development [OECD], 2020). Parents also often report their own maths anxieties, which can impact their children's mathematical outcomes (Maloney et al., 2015).

In the current review, only four studies were conducted at home with parents (Berkowitz et al., 2015; Griffith et al., 2019; Kosko & Ferdig, 2016; Schaeffer et al., 2018). One additional study involved parents in the maths app intervention implemented in the classroom (Schenke et al., 2020). As such, more research is needed to understand how maths app interventions can work in the home to empower parents and support children's learning. This could be achieved through apps that encourage off-screen play and maths talk (Berkowitz et al., 2015; Schaeffer et al., 2018). No apps in the current review included a separate parent area that communicated children's progress. This may support parental involvement in children's mathematical learning, but further research is needed.

## Conclusion

The current systematic review was inclusive of quantitative, qualitative, and mixed methods studies to capture via a narrative synthesis the emerging evidence on whether educational maths apps can support learning, under what circumstances apps work, and for who apps benefit the most. Most studies were implemented in the classroom by teaching practitioners with typically developing children, highlighting the usability and external validity of the evaluated maths apps for this group of children. The best available experimental evidence (RCTs and QEDs) mostly demonstrated positive results towards the maths app intervention for typically developing children, compared to a range of control groups, including standard mathematical practice (business-as-usual), and other educational apps (active control).

However, only two of these studies met the highest standards for rigorous methods, including, a RCT design with a large sample size (greater than 250 children), a standardised measure of mathematical attainment, and sufficiently reported data to calculate within-subject effect sizes (Berkowitz et al., 2015; Outhwaite et al., 2018). As such, although the current evidence base demonstrates promising and externally valid findings, more high-quality rigorous research is needed.

To further enhance the evidence base in this field, eight directions for future research were identified. 1) There is a need to further examine how maths apps can support skills targeted by the app (i.e., near-transfer), as well as broader mathematical skills and related non-attainment outcomes (i.e., far-transfer). 2) Future studies need to examine how maths apps can support children underachieving in mathematics, and these children should be reliably identified in ways that do not threaten the internal validity of the findings. 3) Additional studies are also needed that focus on children with SEND. 4) Within-child factors, such as the child's age and language skills, should also be considered to understand variability in learning with different types and features of maths apps. 5) Future studies should incorporate reliable and consistent reporting of children's maths app usage and examine how this may underpin immediate and sustained learning gains. 6) Innovative methods of data collection can be used to support these efforts by capturing in-app data on learner analytics and working with hard-to-reach populations. 7) Further cross-cultural studies are also needed to examine the use educational maths apps in different educational, cultural, and economic contexts. 8) Finally, additional studies are required to evaluate the use of maths apps in the home learning environment and the role parents play in supporting children's learning. Overall, the current evidence combined with future studies incorporating these eight directions, will help support evidence-based decisions and ensure that maths apps can add value to young children's learning.


## Acknowledgements

The project has been funded by the Nuffield Foundation (grant number FR-000000370), but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org. We would like to thank the members of our advisory board, Prof Alison Clark-Wilson, Prof Natalia Kucirkova, Dr Alison O'Mara-Eves, Prof Victoria Simms, Ms Caroline Wootton, and Ms Lucy Williams.


## Conflict of Interest

The authors have no relevant financial or non-financial interests or completing interests to disclose.

## CRediT Author Statement

**Laura A. Outhwaite:** conceptualisation, methodology, formal analysis, writing- original draft, writing- review and editing, project administration, and funding acquisition. **Erin Early**: methodology, formal analysis, writing- original draft, and writing- review and editing. **Christothea Herodotou:** conceptualisation, methodology, writing- review and editing, and funding acquisition. **Jo Van Herwegen:** conceptualisation, methodology, writing- review and editing, project administration, and funding acquisition.

## References

***50 studies included in the narrative synthesis**


*Ahmad, W. F. W., Muddin, H. N. B. I., & Shafie, A. (2014). Number skills mobile application for down syndrome children. In *2014 International Conference on Computer and Information Sciences (ICCOINS)* (p. 1-6).

Aubrey, C., Godfrey, R., & Dahl, S. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal*, *18*(1), 27-46.

Barnett, A. G., Van Der Pols, J. C., & Dobson, A. J. (2004). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215-220.

*Berggren, J., & Hedler, C. (2014). CamQuest: Design and evaluation of a tablet application for educational use in preschools. In *Proceedings of the 2014 conference on Interaction design and children* (p. 185-188).

*Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, S. C., & Beilock, S. L. (2015). Math at home adds up to achievement in school. *Science*, *350*(6257), 196-198.

Blumberg, M., & Pringle, C. D. (1982). The missing opportunity in organizational research: Some implications for a theory of work performance. *Academy of management Review*, *7*(4), 560-569.

*Broda, M., Tucker, S., Ekholm, E., Johnson, T. N., & Liang, Q. (2019). Small fingers, big data: Preschoolers' subitizing speed and accuracy during interactions with multitouch technology. *The Journal of Educational Research*, *112*(2), 211-222.

*Bullock, E. P., Shumway, J. F., Watts, C. M., & Moyer-Packenham, P. S. (2017). Affordance access matters: Preschool children's learning progressions while interacting with touch-screen mathematics apps. *Technology, Knowledge and Learning*, *22*(3), 485-511.

*Cary, M. S., Kennedy, P. C., Shanley, L., & Clarke, B. (2020). Learning Gains From the KinderTEK® iPad Math Program: Does Timing of a Preventative Intervention Matter?. *Journal of Special Education Technology*, *36*(4), 321-335.

Chaudron, S. (2015). *Young Children (0-8) and Digital Technology a Qualitative Exploratory Study across Seven Countries.* Available from: http://www.lse.ac.uk/media@lse/research/ToddlersAndTablets/RelevantPublications/Young-Children-(0-8)-and-Digital-Technology.pdf.

Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88-113.

*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.*Knowledge and Learning*, *22*(3), 485-511.

Davis-Kean, P. E., Domina, T., Kuhfeld, M., Ellis, A., & Gershoff, E. T. (2021). It matters how you start: Early numeracy mastery predicts high school math course-taking and college attendance. *Infant and Child Development*, e2281.

Department for Education. (2017). *Bold Beginnings.* Available from: https://www.gov.uk/government/publications/reception-curriculum-in-good-and-outstanding-primary-schools-bold-beginnings.

Department for Education. (2019). *Realising the potential of technology in education.* Available from: https://www.gov.uk/government/publications/realising-the-potential-of-technology-in-education

Department for Education. (2020). *SEND code of practice: 0 to 25 years.* Available from: https://www.gov.uk/government/publications/send-code-of-practice-0-to-25.

*Ginsburg, H. P., Wu, R., & Diamond, J. S. (2019). MathemAntics: a model for computer-based mathematics education for young children/MathemAntics: un modelo de enseñanza de matemáticas asistida por ordenador para niños. *Infancia y Aprendizaje*, *42*(2), 247-302.

Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., ... & Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, *3*(1), 2-29.

Griffith, S. F., Hagan, M. B., Heymann, P., Heflin, B. H., & Bagner, D. M. (2020). Apps as learning tools: a systematic review. *Pediatrics*, *145*(1).

*Griffith, S. F., Hanson, K. G., Rolon-Arroyo, B., & Arnold, D. H. (2019). Promoting early achievement in low-income preschoolers in the United States with educational apps. *Journal of Children and Media*, *13*(3), 328-344.

*Grimes, K. R., Park, S., McClelland, A., Park, J., Lee, Y. R., Nozari, M., ... & Zaparolli, B. (2020). Effectiveness of a Numeracy Intelligent Tutoring System in Kindergarten: A Conceptual Replication. *Open Science Framework Pre-print.* Available from: https://osf.io/n2x7b/download.

*Hasanah, N., Hayashi, Y., & Hirashima, T. (2017). An analysis of learner outputs in problem posing as sentence-integration in arithmetic word problems. *Research and practice in technology enhanced learning*, *12*(1), 1-16.

*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.

Herodotou, C. (2018). Young children and tablets: A systematic review of effects on learning and development. *Journal of Computer Assisted Learning*, *34*(1), 1-9.

*Hieftje, K., Pendergrass, T., Kyriakides, T. C., Gilliam, W., & Fiellin, L. (2017). An evaluation of an educational video game on mathematics achievement in first grade students. *Technologies*, *5*(2), 30.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting Education in "Educational" Apps: Lessons from the Science of Learning. *Psychological Science, 16*(1), 3-34.

Hubber, P. J., Outhwaite, L. A., Chigeda, A., McGrath, S., Hodgen, J., & Pitchford, N. J. (2016). Should touch screen tablets be used to improve educational outcomes in primary school children in developing countries?. *Frontiers in Psychology*, *7*, 839.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). *Implementation and process evaluation (IPE) for interventions in educational settings: A synthesis of the literature.* Available from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf.

*Hung, C. Y., Sun, J. C. Y., & Yu, P. T. (2015). The benefits of a challenge: student motivation and flow experience in tablet-PC-game-based learning. *Interactive Learning Environments*, *23*(2), 172-190.

Jordan, N.C. & Levine, S.C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews*. *15(*1), 60-68.

*Judd, N., & Klingberg, T. (2021). Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nature Human Behaviour*, 1-7.

Kabali, H. K., Irigoyen, M. M., Nunez-Davis, R., Budacki, J. G., Mohanty, S. H., Leister, K. P., & Bonner, R. L. (2015). Exposure and use of mobile media devices by young children. *Pediatrics*, *136*(6), 1044-1050.

*Kalmpourtzis, G. (2014). Teaching of Spatial Thinking in Early Childhood Through Game-Based Learning: The use of the Ipad. In *InECGBL20 14–8th European Conference on Games Based Learning: ECGBL 2014* (p. 231).

Kmet, L. M., Cook, L. S., & Lee, R. C. (2004). *Standard quality assessment criteria for evaluating primary research papers from a variety of fields.* Available from: https://era.library.ualberta.ca/items/48b9b989-c221-4df6-9e35-af782082280e.

*Kosko, K., & Ferdig, R. (2016). Effects of a tablet-based mathematics application for pre-school children. *Journal of Computers in Mathematics and Science Teaching*, *35*(1), 61-79.

*Kromminga, K. R., & Codding, R. S. (2020). A comparison of 1: 1 flashcards and a tablet app on student mathematics proficiency. *Journal of Behavioral Education*, 1-26.

*Lee, H. K., & Choi, A. (2020). Enhancing early numeracy skills with a tablet-based math game intervention: a study in Tanzania. *Educational Technology Research and Development*, *68*(6), 3567-3585.

Lin, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PloS one*, *13*(9), e0204056.

*Litster, K., Moyer-Packenham, P. S., & Reeder, R. (2019). Base-10 blocks: A study of iPad virtual manipulative affordances across primary-grade levels. *Mathematics Education Research Journal*, *31*(3), 349-365.

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, *16*(8), 404-406.

Marsh, J., Murris, K., Ng'ambi, D., Parry, R., Scott, F., Thomsen, B. S., … & Woodgate, A. (2020). *Children, Technology and Play.* Billund: The LEGO Foundation.

*Mattoon, C., Bates, A., Shifflet, R., Latham, N., & Ennis, S. (2015). Examining Computational Skills in Prekindergarteners: The Effects of Traditional and Digital Manipulatives in a Prekindergarten Classroom. *Early Childhood Research & Practice*, *17*(1), n1.

*Miller, T. (2018). Developing numeracy skills using interactive technology in a play-based learning environment. *International Journal of STEM Education*, *5*(1), 1-11.

*Moyer-Packenham, P. S., Bullock, E. K., Shumway, J. F., Tucker, S. I., Watts, C. M., Westenskow, A., ... & Jordan, K. (2016). The role of affordances in children's learning performance and efficiency when using virtual manipulative mathematics touch-screen apps. *Mathematics Education Research Journal*, *28*(1), 79-105.

*Nunes, T., Malmberg, L. E., Evans, D., Sanders-Ellis, D., Baker, S., Barros, R., ... & Evangelou, M. (2019). *Onebillion: Evaluation Report.* Education Endowment Foundation (EEF): London, UK.

Organisation for Economic Co-operation and Development. (2020). *Early Learning and Child Well-being: A Study of Five-year-Olds in England, Estonia, and the United States.* Available from:

https://www.oecd.org/education/school/early-learning-and-child-well-being-3990407f-en.htm.

*Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J. (2018). Raising early achievement in math with interactive apps: A randomized control trial. *Journal of Educational Psychology*, *111*(2), 284.

*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., Gulliford, A., & Pitchford, N. J. (2019). A new methodological approach for evaluating the impact of educational intervention implementation on learning outcomes. *International Journal of Research & Method in Education*, *43*(3), 225-242.

*Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2020). Language counts when learning mathematics with interactive apps. *British Journal of Educational Technology*, *51*(6), 2326-2339.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, *372*(71), doi: 10.1136/bmj.n71.

*Parks, A.N. and Tortorelli, L. (2020). Impact of a district-wide one-to-one technology initiative on kindergartners' engagement and learning outcomes. *Journal of Research in Childhood Education,* doi: 10.1080/02568543.2020.1809578.

Pawson, R., & Tilley, N. (1997). Realistic evaluation. In S. Matthieson (Ed.), *Encyclopedia of Evaluation* (p. 359-367). London: Sage.

*Pecora, J. (2015). *What is the effect of specific technology apps on student achievement, focus, and attention in a Kindergarten Special Education Mathematics classroom? (Masters Dissertation).* New Jersey: Caldwell University.

*Pires, A. C., González Perilli, F., Bakała, E., Fleisher, B., Sansone, G., & Marichal, S. (2019). Building blocks of mathematical learning: Virtual and tangible manipulatives lead to different strategies in number composition. *Frontiers in Education*, *4*(81), doi:10.3389/feduc.2019.00081.

Pitchford, N. J. (2022). Customised E-Learning Platforms. In T. Madon, A. J. Gadgil, R. Anderson, L. Casaburi, K. Lee, & A. Rezaee (Eds.), *Introduction to Development Engineering.* London: Springer.

*Pitchford, N. J., Chigeda, A., & Hubber, P. J. (2019). Interactive apps prevent gender discrepancies in early-grade mathematics in a low-income country in sub-Sahara Africa. *Developmental Science*, *22*(5), e12864.

Pitchford, N. J., Gulliford, A., Outhwaite, L. A., Davitt, L. J., Katabua, E., & Essien, A. A. (2021). Using Interactive Apps to Support Learning of Elementary Maths in Multilingual Contexts: Implications for Practice and Policy Development in a Digital Age. In *Multilingual Education Yearbook 2021* (pp. 135-153). Springer, Cham.

*Pitchford, N.J. (2015). Development of early mathematical skills with a tablet intervention: a randomized control trial in Malawi. *Frontiers in Psychology, 6*(485), doi: 10.3389/fpsyg.2015.00485.

*Pitchford, N.J., Kamchedzera, E., Hubber, P.J., & Chigeda, A.L. (2018). Interactive apps promote learning of basic mathematics in children with Special Educational Needs and Disabilities. *Frontiers in Psychology, 9*(262), doi: 10.3389/fpsyg.2018.00262.

Pitchford, N. J., & Outhwaite, L. A. (2019). Secondary benefits to attentional processing through intervention with an interactive maths app. *Frontiers in Psychology*, *10*, 2633.

*Ramani, G. B., Daubert, E. N., Lin, G. C., Kamarsu, S., Wodzinski, A., & Jaeggi, S. M. (2020). Racing dragons and remembering aliens: Benefits of playing number and working memory games on kindergartners' numerical knowledge. *Developmental Science*, *23*(4), e12908.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943-973.

*Schacter, J., & Jo, B. (2016). Improving low-income preschoolers mathematics achievement with Math Shelf, a preschool tablet computer curriculum. *Computers in Human Behavior*, *55*, 223-229.

*Schacter, J., & Jo, B. (2017). Improving preschoolers' mathematics achievement with tablets: A randomized controlled trial. *Mathematics Education Research Journal*, *29*(3), 313-327.

*Schacter, J., Shih, J., Allen, C. M., DeVaul, L., Adkins, A. B., Ito, T., & Jo, B. (2016). Math shelf: A randomized trial of a prekindergarten tablet number sense curriculum. *Early Education and Development*, *27*(1), 74-88.

*Schaeffer, M. W., Rozek, C. S., Berkowitz, T., Levine, S. C., & Beilock, S. L. (2018). Disassociating the relation between parents' math anxiety and children's math achievement: Long-term effects of a math app intervention. *Journal of Experimental Psychology: General*, *147*(12), 1782-1790.

*Schenke, K., Redman, E. J., Chung, G. K., Chang, S. M., Feng, T., Parks, C. B., & Roberts, J. D. (2020). Does "Measure Up!" measure up? Evaluation of an iPad app to teach preschoolers measurement concepts. *Computers & Education*, *146*, 103749.

Sella, F., Onnivello, S., Lunardon, M., Lanfranchi, S., & Zorzi, M. (2021). Training basic numerical skills in children with Down syndrome using the computerized game "The Number Race". *Scientific Reports*, *11*(1), 1-14.

Sella, F., Tressoldi, P., Lucangeli, D., & Zorzi, M. (2016). Training numerical skills with the adaptive videogame "The Number Race": A randomized controlled trial on preschoolers. *Trends in Neuroscience and Education*, *5*(1), 20-29.

Simms, V., McKeaveney, C., Sloan, S. & Gilmore, C. (2019). *Interventions to Improve Mathematical Achievement in Primary School-Aged Children*. Available from: https://www.ulster.ac.uk/__data/assets/pdf_file/0003/417864/web-00553-05_UU_A4_Report_v7.pdf.

*Spencer, P. (2013). iPads: Improving numeracy learning in the early years. In *Proceedings of the 36th Annual Conference of the Mathematics Education Research Group of Australasia (MERGA)* (p.610-617).

*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), doi:10.3389/fpsyg.2017.00179.

Stigler, S. M. (1997). Regression towards the mean, historically considered. *Statistical Methods in Medical Research*, *6*(2), 103-114.

*Stubbé, H., Badri, A., Telford, R., van der Hulst, A., & van Joolingen, W. (2016). E-Learning Sudan, Formal Learning for Out-of-School Children. *Electronic Journal of e-Learning*, *14*(2), 136-149.

*Swicegood, G.P. (2015). *An investigation of the impact of iPad usage on elementary mathematical skills and attitudes (PhD Thesis).* Missoula: University of Montana.

Tamim, R. M., Borokhovski, E., Pickup, D., & Bernard, R. M. (2015). *Large-scale, government-supported educational tablet initiatives*. British Columbia: Commonwealth of Learning.

Toth, K., Sammons, P., Sylva, K., Melhuish, E., Siraj, I., & Taggart, B. (2020). Home learning environment across time: the role of early years HLE and background in predicting HLE at later ages. *School Effectiveness and School Improvement*, *31*(1), 7-30.

*Tucker, S. I., Moyer-Packenham, P. S., Westenskow, A., & Jordan, K. E. (2016). The complexity of the affordance–ability relationship when second-grade children interact with mathematics virtual manipulative apps. *Technology, Knowledge and Learning*, *21*(3), 341-360.

UNESCO. (2017). *More Than One-Half of Children and Adolescents Are Not Learning Worldwide.* Available from: http://uis.unesco.org/sites/default/files/documents/fs46-more-than-half-children-not-learning-en-2017.pdf

UNESCO. (2020). *The State of Broadband 2020: Tackling digital inequalities- A decade for action.* Available from: http://handle.itu.int/11.1002/pub/8165dc3c-en.

United Nations. (2021). *The Sustainable Development Goals Report.* Available from: https://unstats.un.org/sdgs/report/2021/The-Sustainable-Development-Goals-Report-2021.pdf

Van Yperen, N. W., Blaga, M., & Postmes, T. (2015). A meta-analysis of the impact of situationally induced achievement goals on task performance. *Human Performance*, *28*(2), 165-182.

*Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020). The effects of two digital educational games on cognitive and non-cognitive math and reading outcomes. *Computers & Education*, *143*, 103680.

Vega, V. & Robb., M. (2019). *The Common Sense Census: Inside the 21st Century Classroom*. San Francisco: Common Sense Media.

*Watts, C. M., Moyer-Packenham, P. S., Tucker, S. I., Bullock, E. P., Shumway, J. F., Westenskow, A., ... & Jordan, K. (2016). An examination of children's learning progression shifts while using touch screen virtual manipulative mathematics apps. *Computers in Human Behavior*, *64*, 814-828.

Watts, T. W., Clements, D. H., Sarama, J., Wolfe, C. B., Spitler, M. E., & Bailey, D. H. (2017). Does Early Mathematics Intervention Change the Processes Underlying Children's Learning?. *Journal of Research on Educational Effectiveness*, *10*(1), 96-115.

*Wu, R. (2020). The effect of the math application *MathemAntics* on preschoolers' math performance (*PhD Thesis).* New York: Columbia University.

Yudkin, P. L., & Stratton, I. M. (1996). How to deal with regression to the mean in intervention studies. *Lancet*, *347*, 241-243.

*Zander, S., Wetzel, S., & Bertel, S. (2016). Rotate it!–Effects of touch-based gestures on elementary school students' solving of mental rotation tasks. *Computers & Education*, *103*, 158-169.

*Zhang, L., Shang, J., Pelton, T., & Pelton, L. F. (2020). Supporting primary students' learning of fraction conceptual knowledge through digital games. *Journal of Computer Assisted Learning*, *36*(4), 540-548.

## Appendix

**Preliminary scoping searches, 18th January 2021**

**Search string 1** – ("early years" OR preschool* OR kindergarten OR "primary school" OR "elementary school" OR "young children") AND ("educational app" OR "interactive app" OR tablet OR "tablet technology" OR "iPad app" OR "math* app") AND (math* OR number* OR "number sense" OR arithmetic* OR measurement OR geomet* OR shape)

**Search string 2-** ("early years" OR preschool* OR **kindergart***  OR "primary school" OR "elementary school" OR "young children" OR "**young pupils**" OR "**young students**") AND ("educational app" OR "interactive app" OR tablet OR "tablet technology" OR "iPad app" OR "**android app**" OR "math* app") AND (math* OR number* OR "number sense" OR arithmetic* OR measurement OR geomet* OR shape)

**Search string 3-** ("early years" OR preschool* OR **kindergart***  OR "primary school" OR "elementary school" OR "young children" **OR "young pupils" OR "young students" OR child* OR pupils OR students**) AND ("educational app" OR "interactive app" OR tablet OR "tablet technology" OR "iPad app" OR "**android app**" OR "math* app") AND (math* OR number* OR "number sense" OR arithmetic* OR measurement OR geomet* OR shape)

***Table A3 Results of the preliminary scoping searches***

| Database | Platform | Search 1 | Search 2 | Search 3 |
|---|---|---|---|---|
| PsycINFO | Ovid | 82 | 43 | 25 |
| ERIC | ProQuest | 96 | 9 | 221 |
| Medline | Ovid | 33 | 21 | 29 |
| Scopus | Elsvier | 1,582 | 2,012 | 1,807 |
| Science Citation Index-Expanded | Web of Science (Core Collection) | 78 | 78 | 607 |
| Social Science Citation Index (SSCI) | Web of Science | 123 | 126 | 373 |
| Arts and Humanities Citation Index | Web of Science | 4 | 4 | 10 |
| Emerging Sources Citation Index | Web of Science | 38 | 39 | 188 |
| PubMed | | 348 | 348 | 1,401 |
| British Education Index (BEI) | EBSCO | 34 | 35 | 96 |
| Australian Education Index (AEI) | ProQuest | 9 | 9 | 31 |
| **Total** | | **2427** | **2724** | **4788** |

@cepeo_ucl

ucl.ac.uk/ioe/cepeo