

Understanding how educational maths apps can enhance learning: A content analysis and qualitative comparative analysis

Laura A. Outhwaite¹  | Erin Early^{1,2} | Christothea Herodotou³ |
Jo Van Herwegen^{1,4} 

¹Centre for Education Policy and Equalising Opportunities, IOE, UCL's Faculty of Education and Society, London, UK

²School of Social Sciences, Education and Social Work, Queen's University Belfast, Belfast, UK

³Institute of Educational Technology, Open University, Milton Keynes, UK

⁴Department of Psychology and Human Development, IOE, UCL's Faculty of Education and Society, London, UK

Correspondence

Laura A. Outhwaite, Centre for Education Policy and Equalising Opportunities, IOE, UCL's Faculty of Education and Society, London, UK.

Email: l.outhwaite@ucl.ac.uk

Funding information

Nuffield Foundation, Grant/Award Number: FR-000000370

Abstract

Educational applications (apps) are ubiquitous within children's learning environments and emerging evidence has demonstrated their efficacy. However, it remains unclear what the active ingredients (ie, mechanisms), or combination of ingredients, of successful maths apps are. The current study developed a new, open-access, three-step framework for assessing the educational value of maths apps, comprised of type of app, mathematical content and app design features. When applied to a selection of available maths apps previously evaluated with children in the first 3 years of school (the final sample included 23 apps), results showed that practice-based apps were the most common app type tested ($n=15$). Basic number skills, such as number representation and relationships, were the most common area of mathematics targeted by apps ($n=21$). A follow-up qualitative comparative analysis showed observed learning outcomes with maths apps were enhanced when apps combined the following: a scaffolded and personalised learning journey (programmatically levelled) and explanations of why answers were right or wrong (explanatory feedback), as well as praise, such as 'Great job!' (motivational feedback). This novel evidence stresses the significance of feedback and levelling design features that teaching practitioners and

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *British Journal of Educational Technology* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

other stakeholders should consider when deciding which apps to use with young children. Directions for future research are discussed.

KEYWORDS

early years, iPad, mathematics

Practitioner notes

What is already known about this topic

- Educational apps have been shown to support maths attainment in the first 3 years of school.
- Several existing frameworks have attempted to assess the educational value of some of these maths apps.
- Emerging experimental evidence also demonstrates the benefits of specific app design features, including feedback and levelling.

What this paper adds

- Practice-based maths apps are the most common type of app previously evaluated with young children.
- These evaluated maths apps have mostly focused on basic number skills.
- The combination of explanatory and motivational feedback, with programmatic levelling (either dynamic or static), was a necessary condition for enhancing learning outcomes with maths apps.

Implications for practice and policy

- The inclusion of feedback and levelling in maths apps should be considered by app developers when designing apps, and by educational practitioners and parents when deciding which apps to use with their children.
- Further consideration is also needed for the development of educational apps that include a broad range of maths skills.

INTRODUCTION

Educational applications (apps) are interactive software designed to support learning and are primarily used on a hand-held touch screen tablet or smartphone device. These technologies are ubiquitous within children's school and home learning environments. In the United Kingdom, over 94% of children own or have access to touchscreen tablet devices. In other countries, such as South Africa, children are also more likely to have access to a tablet device, compared to a laptop or television (Marsh et al., 2020). Since the Covid-19 pandemic, the use of educational apps has increased with the aim to support learning and raise educational attainment (Department for Education, 2021a; Ofcom, 2020). However, with approximately 100 new apps for young children released every year since 2016 (Kanders et al., 2022), determining whether or what kind of apps provide a high-quality learning experience remains a significant challenge.

A recent systematic review identified 77 educational maths apps, which have been previously evaluated across 50 studies with children in the first 3 years of compulsory school

(Outhwaite et al., 2023). Overall, the studies predominately reported greater learning outcomes for children using the evaluated maths apps, compared to a range of control conditions. While this systematic review addressed if educational maths apps can be an effective learning tool, it remains unclear what the active ingredients (ie, mechanisms), or combination of ingredients, of successful maths apps are. To examine how different maths apps work, the underpinning pedagogy and app design features need to be examined (Griffith et al., 2020) and linked to the observed learning outcomes (Outhwaite et al., 2019).

The need to develop a new content analysis framework

To date, only one systematic review on educational apps (maths and literacy) has attempted to consider the potential relationships between the app design features and observed learning outcomes. Kim et al. (2021) scored 36 identified apps in their systematic review using five questions based on Hirsh-Pasek et al.'s (2015) theoretical framework. This framework proposed that high-quality educational apps should include active, engaged, meaningful and socially interactive learning with a specific learning goal. Questions included, 'do activities promote meaningful learning?', which were then rated on a 3-point Likert scale, from 'low (app contains many features that distract from learning)' to 'high (app promotes conceptual mastery that is consistently connected to a broader learning context)' (Kim et al., 2021, p. 4, [Supporting Information](#)). Moderation analyses showed no relationship between the quality of app score and the learning gains in the reported meta-analysis.

However, the questions used to assess the apps did not include specific app design features. For example, feedback, levelling, social interaction, task instruction, meaningful learning and solving problems are all prevalent across existing frameworks on the educational values of apps (Callaghan & Reich, 2018; Herodotou, 2021; Kolak et al., 2020; Meyer et al., 2021; Papadakis et al., 2017) (see [Table S1](#)). Of these specific app design features, feedback and levelling have been shown to be particularly important components of app-based instruction (Callaghan & Reich, 2021; Vanbecelaere et al., 2021).

Feedback in educational apps

Feedback within educational maths apps can be defined as immediate responses from the app, based on the actions, input and performance of the user (Tucker, 2015). Feedback can be understood according to its explanatory and motivational components. Explanatory feedback within the apps provides the user with an explanation of why their answer is correct/incorrect. Motivational feedback provides general feedback to the user such as 'You did it!' or 'Great job!' but is not directly associated with the answers or performance of the user (Herodotou, 2021).

Research shows children made more deliberate decisions while using educational apps with explanatory feedback, compared to a no feedback condition, where more trial-and-error responses were observed (Blair, 2013). Further experimental evidence shows preschool-aged children who received explanatory verbal feedback, made significantly fewer errors during a novel practice-based mathematics sorting game compared to children who received motivational non-verbal feedback (eg, cheering sound effects). Motivational verbal feedback (eg, 'Great job!') did not increase performance accuracy compared to other forms of feedback. However, the effect of explanatory, verbal feedback was no longer observed after increased engagement with the maths app (Callaghan & Reich, 2021). This is consistent with other studies demonstrating that once children can complete a learning task, the requirement for detailed feedback is reduced (Bartoschek et al., 2013).

Levelling in educational apps

Levelling can be defined as tailoring learning content so that it accounts for and builds on children's prior knowledge and progression (Hsin & Wu, 2011; Magliaro et al., 2005). This can provide meaningful and accessible challenges, just beyond the child's current ability level (Inal & Cagiltay, 2007; Masterman & Rogers, 2002), in the zone of proximal development (Vygotsky, 1978).

Levelling can be implemented in educational maths apps in three ways: (1) participatory free form, (2) programmatic static and (3) programmatic dynamic (Kucirkova, 2018; Vandewaetere & Clarebout, 2014). Participatory free form levelling refers to apps that provide a suggested but not enforced sequence of learning content (Kucirkova, 2018). Evidence suggests young children can benefit from this form of levelling (Schenke et al., 2020), as they have greater agency and can be guided by their own interests. However, there is the risk that children may choose content that is too easy and thus hinder learning efficiency.

In contrast, programmatic levelling places children on a scaffolded and personalised learning journey. In programmatic static levelling, the learning content is tailored to a child based on an initial attainment assessment or is preselected by an adult. In programmatic dynamic levelling, the presented learning content is adapted in response to a child's performance while using the app (Vandewaetere & Clarebout, 2014). Experimental evidence shows app-based learning tasks were completed quicker and with increased accuracy when learning tasks gradually increased in difficulty (ie, programmatic levelling), compared to a non-strategic sequence of learning activities (ie, participatory free form levelling) (Callaghan & Reich, 2021; Hooshyar et al., 2018). Further research has found no overall group differences between dynamic and static forms of programmatic levelling in app-based instruction, but some individual differences were observed. Lower ability children appeared to benefit most from the predetermined and structured trajectory in the programmatic static levelling condition, whereas higher ability children thrived in the programmatic dynamic levelling condition. This is most likely due to the opportunity for higher ability children to skip learning content that had already been mastered, and thus, the app was able to provide more challenging content (Vanbecelaere et al., 2021).

Existing frameworks for the educational value of apps

Many existing frameworks assessing the educational value of apps have been developed. For example, Herodotou (2021) took a bottom-up approach and identified, through a literature review, app features that may facilitate or hinder learning across different subject areas including main figure (eg, a character or figure included in the app that communicates objectives, introduces concepts and models responses to activities), feedback, instructions, highlighting information, constraints (eg, restricting certain interactive features or limiting options), linking multiple representations, experimentation, as well as other features such as progression, sounds and language (Falloon, 2013; Moyer-Packenham et al., 2016). The proposed framework by Herodotou (2021) was then expanded through the observational study of 17 children using one practice-based maths app (*Moose Math*), which emphasised the importance of feedback, experimentation and learning theories in app design.

In contrast, other frameworks have been developed based on developmental and learning science theory (eg, Hirsh-Pasek et al., 2015) and thus adopted a top-down approach to coding (Callaghan & Reich, 2018; Kolak et al., 2020; Meyer et al., 2021; Papadakis et al., 2017). For example, Meyer et al. (2021) designed a detailed coding scheme and scoring system based on Hirsh-Pasek et al.'s (2015) four-pillar theoretical framework, which includes an

emphasis on off-screen social interaction. However, many of the items included in these frameworks are biased towards certain types of apps (see Kay & Kwak, 2018). For example, Meyer et al. (2021) list off-screen social interaction as a key component for successful outcomes. But this is a predominate feature of parent-based apps (eg, *Bedtime Math*) and is not aligned to the design principles of other types of apps, such as practice-based and game-based apps, which are typically designed to be individually used by the child. As such, when using this scoring scheme, parent-based apps will inherently gain a higher score compared to other types of apps. This is problematic, as no intervention studies to date have directly compared the learning outcomes of these different types of apps with these features (see Outhwaite et al., 2023). Thus, it is currently unknown whether this distinction between face-to-face interaction and in-app character engagement is appropriate within the context of app-based instruction.

A similar issue is also present in Kolak et al.'s (2020) framework, with the inclusion of the storyline items, which are a key component of game-based apps and are less relevant to other types of apps. To develop unbiased conclusions about the potential mechanisms underpinning learning outcomes with app-based maths instruction, it is important to develop framework items that are applicable to the different types of maths apps.

Overall, within the previous research examining the educational value of apps (Callaghan & Reich, 2018; Herodotou, 2021; Kolak et al., 2020; Meyer et al., 2021; Papadakis et al., 2017), it is assumed that all app design features, such as feedback and levelling, are equally important and combine in equivalent ways. However, some app design features may be necessary, sufficient or inconsequential. As such, the associated scoring systems within many of these frameworks are not suitable for making meaningful connections to intervention study outcomes. Therefore, a new framework for evaluating the educational value of maths apps is needed that affords more nuanced approaches and analyses to understand how specific maths app features interact (or not) with each other and support learning.

Current study

To address these limitations, the current study used a new three-step framework that extends the breadth and depth of previous research (ie, Callaghan & Reich, 2018; Herodotou, 2021; Kolak et al., 2020; Meyer et al., 2021; Papadakis et al., 2017) to examine: (1) the type of app, (2) mathematical content and (3) app design features of educational maths apps that have been evaluated in previous research. Specifically, the current study builds on the systematic review reported by Outhwaite et al. (2023) on educational maths apps with young children aged 4–7 years in the first 3 years of compulsory school. The current study addresses two research questions (RQs). First, this study asked, what type of apps, mathematical content and app design features have been included in maths apps previously evaluated? (RQ1). Using content analysis methods, the current study combined top-down (deductive) and bottom-up (inductive) approaches to answer RQ1. This enabled the current study to build on existing frameworks and ensure other app features, not currently included in existing frameworks, were also captured.

Second, this study asked, which app design features, or combination of design features, underpin children's learning outcomes with maths apps? (RQ2). To address the limitations of previous frameworks, the current study used a qualitative comparative analysis (QCA). QCA is an evidence synthesis approach, which is designed to identify configurations of factors (eg, app design features) that may be associated with a given outcome (eg, enhanced learning outcomes) (Thomas et al., 2014). It is a useful tool for synthesising components of complex interventions, where there are a limited number of evaluation studies identified through a systematic review, combined with a relatively large number of possible factors

that may influence the observed outcomes. Unlike previous moderation analysis (eg, Kim et al., 2021), QCA does not rely on a linear additive model. Instead, QCA seeks to examine how multiple possible factors can contribute to the observed outcomes, and so identifies 'causal recipes, not net effects' (Thomas et al., 2014, p. 4).

METHODS

The current study builds on the systematic review reported in Outhwaite et al. (2023). A systematic search of academic and grey literature identified 50 studies that have evaluated 77 educational maths apps with 23,981 children in the first 3 years of compulsory school, across 18 countries. A narrative synthesis of this evidence highlighted that most studies evaluated the selected maths app(s) with typically developing children ($n=43$) within a randomised control trial or quasi-experimental design ($n=33$). Most studies found some positive benefits on mathematical learning outcomes ($n=46$). The current study expanded on this evidence by examining the underpinning pedagogy of the identified apps (RQ1) and assessing how app design features may combine to support learning outcomes (RQ2).

The current study protocol was preregistered on the Open Science Framework (osf.io/pzkmh) and ethical approval was granted by the IOE ethics committee (REC 1376).

Inclusion criteria

For the purposes of this study, an educational app was defined as interactive software primarily used on a hand-held touch screen tablet or smartphone device. The current study focused on children in the first 3 years of compulsory school.

To be eligible for inclusion in the content analysis, the maths apps identified in the systematic review Outhwaite et al. (2023) had to meet the following criteria:

- An app needed to be the individual focus of an intervention study that had mathematical attainment as the primary outcome measure, which was measured before (pre-test) and after (post-test) the intervention period. If multiple apps were included in one study, the results needed to be reported separately for each app, or app-specific learning data needed to be provided by the lead author of the study, following communication from the research team.
- Additionally, apps needed to be commercially available and accessible for download from the Apple and/or Google Play store(s) in the United Kingdom. If the app was not commercially available (eg, it has been developed by a research team for the purposes of the study), the lead author was contacted to request access. Access to the apps was required for later coding (see below).

App screening

As shown in Figure S1, of the 77 apps identified in the systematic review (Outhwaite et al. 2023), 23 apps met the eligibility criteria (references included in Supporting Information). Thirty-three apps identified in the systematic review could not be linked to learning outcomes for the content analysis. This was the result of the apps being included in interventions with multiple apps and the respective effect of each not being disaggregated (eg, Parks & Tortorelli, 2020). The remaining 21 apps were excluded due to: the app was unavailable

and no response was received from the lead author/app developer when emailed to request access (including two follow-up emails) ($n=7$), the app was no longer available ($n=6$), the app was unavailable on the Apple or Google Play stores in the United Kingdom ($n=6$) and the app was not available in English ($n=2$).

App coding

Each of the included apps ($n=23$) was then coded following a three-step process: (1) the type of app, (2) the mathematical content covered by the app and (3) the design features. This three-step coding process was designed to extend the breadth and depth of previous research (ie, Callaghan & Reich, 2018; Herodotou, 2021; Kolak et al., 2020; Meyer et al., 2021; Papadakis et al., 2017) and the details are outlined below.

Type of app

First, the 23 apps were classified by type. This was based on Kay and Kwak (2018) taxonomy of different types of educational apps. While these categories were exclusive (ie, there was no overlap across the different categories), they were not exhaustive (ie, the categories did not cover all types of educational apps currently available); an important feature of qualitative coding (Braun & Clarke, 2006). As such, the taxonomy was expanded to include parent-based apps (see Table 1).

Mathematical content

Second, the mathematical content within the 23 included apps was catalogued, based on four areas of mathematical development relevant to the first 3 years of school: (1) number representation and relationships (11 description points including transcoding, number bonds and number line estimation), (2) counting (7 description points including one-to-one correspondence, cardinality and skip counting), (3) arithmetic (10 description points including addition, subtraction and arithmetic symbols and language) and (4) shape, patterns and measurement (10 description points including working with patterns, shape recognition and sequence of events) (see Table S2).

These four initial categories and their description points were developed based on current theories of mathematical development, which highlight the importance of a range of specific maths skills (Butterworth, 2005; Clements & Sarama, 2009; Gilmore et al., 2018). They were also aligned to best practice guidelines for effective early mathematics teaching (Clark et al., 2020), as well as content covered in mathematical curriculums in the first 3 years of school in England (Department for Education, 2021b, 2013) and the USA (Common Core State Standards Initiative, n.d.).

Included apps were dichotomously coded (not present [0] or present [1]) for the four outlined areas of mathematical development. Apps had to meet at least one of the description points, for the area of mathematical development to be coded as present. To capture more detail about how well the different areas of mathematical development were covered within the included apps, each of the description points was also dichotomously coded (not present [0] or present [1]). There was also the opportunity to include other mathematical skills, not currently listed, to ensure exhaustive coding.

TABLE 1 Summary of the different types of apps with descriptions and content analysis results.

Type of app	Description	Number of apps	Name of included apps
Practice-based	Designed to support the acquisition of learning content, such as mathematical facts and concepts, through targeted practice. Primarily designed to be used by children individually and is self-paced. Child is the user of the learning content	15	<i>Fingu; Friends of Ten; Intro to Math; IXL; KinderTEK; Know Number Lite; MathemAntics; Maths 3–5; Maths 4–6; Math Shelf; Montessori Numbers for Kids; Native Numbers; Pink Tower; Splash Math 2nd Grade; Teaching Number Lines</i>
Game-based	Same as practice-based apps, with the addition that learning content is embedded within broader immersive player narrative	4	<i>Addimal Adventure; Slice Fractions; Vektor; Zorbit's Math Adventure</i>
Constructive	Designed to encourage the exploration and active manipulation of mathematical ideas and concepts. Child is the user of the learning content	2	<i>Montessori Bead Skip Counting; 100s Board</i>
Productive	Designed to support children to produce their own content, for example, to present their own ideas on a particular maths topic; they are creators of their own learning content	1	<i>Quizlet Plus</i>
Parent-based	App content is primarily designed for parents/caregivers and encourages offline interactions and learning opportunities with children	1	<i>Bedtime Math</i>

Design features

Third, the presence of five app design features in the 23 maths apps were dichotomously coded (not present [0] or present [1]): (1) feedback, (2) levelling, (3) social interaction, (4) task instruction and (5) meaningful learning and solving problems. A description of each design feature (Table S1) and how it was coded (Table S4) is included in the Supporting Information. This list of five app design features was collated based on their predominant presence in existing frameworks evaluating the educational value of apps (Callaghan & Reich, 2018; Herodotou, 2021; Kolak et al., 2020; Meyer et al., 2021; Papadakis et al., 2017; see Table S1). Additional design features were also identified through reviewer engagement with the app and by re-visiting the associated papers and applying the principles of an intervention component analysis (Sutcliffe et al., 2015) (see Table S4).

Coding procedure

Included apps were coded using a data collection form developed for this study (see [Table S4](#)). To understand the app design features, reviewers (first and second authors) played with each app for 20–30 minutes and engaged with a minimum of 10 different activities from across the app. During play, reviewers gave both correct and incorrect answers to understand how each maths app responded to user behaviour. The first reviewer (second author) completed the coding procedure for all the included maths apps. A second reviewer (first author) repeated the coding procedure for 20% (randomly selected) of the apps. There was excellent agreement between the reviewers ($\kappa=0.85$). Any disagreements were resolved through discussion.

Effect sizes on learning outcomes

To capture children's learning outcomes with the evaluated maths apps that met the inclusion criteria, the following data were extracted from the relevant studies:

- Group mean and standard deviation for pre-test and post-test mathematical attainment scores for the intervention group(s).
- Final sample size of the intervention group(s).
- Intervention intensity: number of weeks the intervention was implemented for and the number and length of sessions per week.
- Whether the mathematical assessment tool used as the outcome measure was standardised or researcher developed.

The extracted data on pre-test and post-test mathematical attainment scores were used to calculate within-subject (ie, pre-test to post-test) effect sizes for the progress made in maths attainment over the duration of the intervention period for the intervention group only. Maths attainment was measured using a range of standardised and research developed assessment tools (see [Table S5](#)). The final sample size was used to calculate the confidence interval for the observed within-subject effect size. Hedges' *g* corrections were applied for final sample sizes equal to or <50 (Lin, 2018) (see [Table 2](#)).

In line with the preregistered protocol, a highly effective intervention in the QCA was defined as within-subject effect size >1. In cases where maths apps were evaluated in multiple studies (eg, *onebillion Maths 3–5 and Maths 4–6*; *Math Shelf*), the most robust study (eg, a randomised control trial with the largest sample size) that had sufficiently reported data to calculate the within-subject effect size was used as an indication of children's learning outcomes (eg, Outhwaite et al., 2018; Schacter & Jo, 2017).¹

Qualitative comparative analysis

To address RQ2, a QCA was conducted to understand which of the specific app design features, or combination of features, were necessary or sufficient² for enhancing children's learning outcomes with educational maths apps. To be included in the QCA analysis, apps needed to be the individual focus of an intervention study that had mathematical attainment as the primary outcome measure and included, sufficiently reported outcome measures before (pre-test) and after (post-test) the intervention period. If the study included multiple maths apps and the reported results were not disaggregated for each app, the identified apps were excluded. This was because there was no guarantee that these apps, with

TABLE 2 Summary of the included apps in the qualitative comparative analysis (QCA) with potential conditions and classifications of highly effective intervention sets.

Study	App	Potential conditions					Outcome (within-group effect size)			
		Feedback—explanatory	Feedback—motivational	Levelling-programmatic (dynamic or static)	Social interaction	Task instruction	Cohen's d	Hedge's g^a	Highly effective intervention Set ^b	Sample size (treatment group only)
Berkowitz et al. (2015)	<i>Bedtime Math</i>	0	0	1	1	0	0.82	n/a	0	Unknown
Cary et al. (2020)	<i>KinderTek</i>	1	1	0	1	1	0.88	n/a	0	58
Spencer (2013)	<i>Know Number Lite</i>	0	1	0	0	0	0.56	n/a	0	114
Schacter and Jo (2017)	<i>Math Shelf</i>	0	1	1	1	0	0.20	n/a	0	231
Wu (2020)	<i>MathemAntics</i>	1	1	1	1	1	1.89	1.84	1	28
Grimes et al. (2020)	<i>Native Numbers</i>	1	1	1	0	1	1.10	1.06	1	24
Outhwaite et al. (2018)	<i>Onebillion (Maths 3–5 and Maths 4–6)</i>	1	1	0	1	1	0.78	n/a	0	126
Kosko and Ferdig (2016)	<i>Zorbit</i>	1	1	1	1	1	1.45	1.41	1	27

^aHedge's g correction applied to effect sizes where study sample sizes were <50.

^bEffect sizes >1.00 coded as 1. Effect sizes <1.00 coded as 0.

combined results, would have the same app design features and thus, the identification of active ingredient/s would be unclear.

Eight apps were suitable for this analysis (see [Figure S1](#)). The associated included studies focused on typically developing children aged between 4 and 7 years old with an experimental design (randomised control trial or quasi-experimental design) (Berkowitz et al., 2015; Cary et al., 2020; Grimes et al., 2020; Kosko & Ferdig, 2016; Outhwaite et al., 2018; Schacter & Jo, 2017; Spencer, 2013; Wu, 2020). Two studies reported subgroup analyses on children identified as underachieving in mathematics (Cary et al., 2020; Wu, 2020). However, for the purposes of this analysis, the extracted within-group effect sizes focused on the whole sample. Details of the eight studies, including intervention intensity and outcome measures, are included in [Table S5](#).

RESULTS

To address RQ1, an overview is presented for the content analysis, which describes the types of apps, mathematical content and app design features, that were included in a sample of 23, previously evaluated, maths apps. To address RQ2, a QCA is presented, which examines the app design features, or combination of design features that were associated with enhanced learning outcomes with the eight eligible maths apps.

Content analysis: Type of app (RQ1)

As outlined in [Table 1](#), practice-based apps were the most popular type of app that has been evaluated ($n=15$). Productive ($n=1$) and parent-based apps ($n=1$) were the least common.

Content analysis: Mathematical content (RQ1)

Number representation and relationships were the most common area of maths development covered within the included apps ($n=21$). There were 18 apps that targeted counting skills, 12 apps targeted arithmetic and 13 apps targeted shape, patterns and measurement. [Table S2](#) summarises the specific mathematical skills included within these overall areas of mathematical development.

Four apps also included additional mathematical areas that were not categorised in the initial coding (*IXL*; *Slice Fractions*; *Splash Maths 2nd Grade*; *Vektor*). *IXL* included the exploration of probability, data and graphs. *Slice Fractions* included the addition and subtraction of fractions. *Splash Maths 2nd Grade* explored how to read data from line, picture and bar graphs. *Vektor* included visuo-spatial working memory tasks and shape rotation tasks.

Content analysis: App design features (RQ1)

[Table S3](#) summarises the app design features included within the sample of maths apps. Most maths apps included explanatory (ie, explaining why an answer is right or wrong) and motivational feedback (ie, 'Great job!') ($n=12$).

The levelling of maths content was most commonly provided in a suggested, but not enforced sequence of activities (ie, participatory free form levelling) ($n=12$). In contrast, four apps provided learning content that was tailored to the child based on an initial assessment

or learning content that was preselected by an adult (ie, programmatic static levelling). Five apps offered learning content that was adaptive to the individual child based on their performance while using the app (ie, programmatic dynamic levelling).

Ten apps included an in-app character who provided task instructions and, in some cases, modelled the learning activity. Only two apps included support for adult–child interactions.

Evidence of task instructions was mixed. In seven apps, instructions could not be repeated, and no instructions were available in five apps. However, six apps provided task instructions, which could be repeated by the child, as often as required.

Most apps provided the opportunity to practice basic maths skills in isolation ($n=17$), but very few offered practice in multiple basic maths skills in relation to each other ($n=3$). No apps embedded the maths skills practice within a real-life context or were applied to solve novel problems.

Three apps also included additional design features that were not categorised in the initial coding (*Maths 3–5*; *Maths 4–6*; *Pink Tower*). *Maths 3–5* and *Maths 4–6* included end-of-topic quizzes, which were designed to assess and monitor children's progress through the app. *Pink Tower* provided children with multiple opportunities to respond but constrained incorrect moves to facilitate accuracy.

Qualitative comparative analysis (RQ2)

Preliminary analyses

To ensure the relative heterogeneity of the included studies, preliminary analyses showed no significant association was observed between intervention intensity (ie, time spent on the app) and the observed within-subject effect sizes, $r=-0.12$, $p=0.822$ (see [Table S5](#)). An independent samples *t*-test also showed no significant differences in observed within-subject effect sizes based on if the outcome measure was standardised or researcher developed, $t(6)=0.307$, $p=0.769$ (see [Table S5](#)).

Qualitative comparative analysis selected conditions

The QCA aimed to identify configurations of app design features that may be associated with enhanced learning outcomes. Due to the small number of maths app interventions available for this analysis ($n=8$), the number of included app design features within the QCA needed to be restricted to three (see [Table 2](#)). To explain differences between learning outcomes reported in the associated studies, the QCA also needed to include app design features that were considered the most salient (Thomas et al., 2014). As such, feedback (explanatory and motivational) and levelling (programmatic and participatory) were the chosen features to be included as conditions in the QCA because of the observed variation across the sample of apps (see [Table 2](#)). These features have also been the focus of previous experimental research (eg, Callaghan & Reich, 2021; Vanbecelaere et al., 2021), which the current study can build upon.

Specifically, Model 1 focused on the potential combinations of explanatory feedback, motivational feedback and programmatic levelling. Next, Model 2 further examined the potential combinations of the different types of programmatic levelling (ie, dynamic and static) as separate conditions and collapsed the different forms of feedback features into one condition (ie, a fuzzy set—see below).

Model 1

Model 1 included the following three conditions: explanatory feedback, motivational feedback and programmatic levelling (either dynamic or static), all entered as dichotomous variables (yes = 1, no = 0). As shown in [Table 3](#), results indicated the combination of explanatory and motivational feedback present within the apps, together with programmatic levelling (static or dynamic, rather than participatory levelling) was a necessary condition for highly effective maths apps. Although motivational feedback on its own was not associated with highly effective maths apps, importantly, it was not a hindrance on children's learning outcomes. It is also important to highlight that the differential effects of explanatory and motivational feedback could not be disentangled in this analysis. This is because within this sample, there were no maths apps that included explanatory feedback only.

Model 2

Model 1 focused on programmatic levelling, which can be dynamic or static. To further explore any differential effects between these different types of levelling, the three conditions included in Model 2 were programmatic dynamic levelling, programmatic static levelling (both entered as dichotomous variables) and feedback entered as a fuzzy set. Within a QCA approach, a fuzzy set can include multiple categories to allow for partial membership in a specified set, whereby a score of 1 represents full membership in the set, 0 indicates full non-membership in the set and a score of 0.5 signifies neither fully in, nor fully out of the set (Thomas et al., 2014). In the context of model 2, explanatory and motivational feedback was given a stronger weighting (score of 1), compared to motivational feedback only (score of 0.5) and no feedback (score of 0; see [Table 4](#)).

Consistent with Model 1, results showed the combination of explanatory and motivational feedback combined with programmatic dynamic levelling was a necessary condition for highly effective maths apps. The same pattern of results was also observed for programmatic static levelling. This suggests that there were no differential effects between the different types of programmatic levelling for enhancing children's learning outcomes within app-based mathematics instruction.

DISCUSSION

The current study uses a new three-step framework for analysing the educational value of maths apps for children in the first 3 years of compulsory school. Specifically, the content analysis framework examined the type of app, mathematical content and app design features within 23 educational maths apps. These maths apps have been evaluated in previous research in terms of their impact on young children's learning outcomes (Outhwaite et al., 2023). The reported QCA extended on these descriptive results to examine which specific app design features, or combination of features, were associated with enhanced learning gains. To our knowledge, this is the first study to examine how educational maths apps are designed to support learning with a comprehensive (top-down and bottom-up) approach that was inclusive of different types of apps. Moreover, the innovative use of QCA provides actionable recommendations for app designers, educational practitioners and other stakeholders designing, using and evaluating maths apps with young children.

TABLE 3 Summary of Model 1 in the qualitative comparative analysis (QCA).

	Feedback-explanatory	Feedback-motivational	Levelling-programmatic (dynamic or static)	Number of apps	Membership in 'highly effective intervention' set	Raw consistency
1	1	1	1	3 (<i>MathemAntics; Native Numbers; Zorbit</i>)	1	1.00
1	1	0	0	2 (<i>KinderTek; onebillion</i>)	0	n/a
0	1	1	1	1 (<i>Math Shelf</i>)	0	n/a
0	1	0	0	1 (<i>Know Number Lite</i>)	0	n/a
0	0	1	1	1 (<i>Bedtime Math</i>)	0	n/a

TABLE 4 Summary of Model 2 in the qualitative comparative analysis (QCA).

Levelling—dynamic	programmatically	Levelling—static	programmatically	Feedback ^a	Number of apps	Membership in 'highly effective intervention' set	Raw consistency
1	0	0	1	1	2 (Native Numbers; Zorbit)	1	1.00
0	1	1	1	1	1 (MathemAntics)	1	1.00
0	0	0	1	1	2 (KinderTek; onebillion)	0	n/a
1	0	0	0.5	0.5	1 (Math Shelf)	0	n/a
0	0	0	0.5	0.5	1 (Know Number Lite)	0	n/a
0	1	1	0	0	1 (Bedtime Math)	0	n/a

^a0 = no feedback; 0.5 = motivational feedback only; 1 = explanatory and motivational feedback.

Type of apps

The current content analysis identified that to date, practice-based maths apps are the most common type of app to be evaluated ($n=15$) (see [Table 1](#)). This is consistent with broader reviews, which show a greater emphasis on the evaluation of mobile learning where the child is the user of the learning content, rather than producers, collaborators and creators of knowledge (Crompton et al., 2017). Productive apps (eg, *Quizlet Plus*), which align with these characteristics, were identified in the current study as one of the least common types of app to have been evaluated in the context of early maths development.

Similarly, only one parent-based maths app (eg, *Bedtime Math*) has been evaluated. Unlike the other types of apps, parent-based apps have the opportunity to facilitate the principles of guided play. These types of apps provide caregivers with ideas for how to engage in playful learning opportunities in maths with their children. Guided play, which is initiated by adults, and led by children, has been shown to have a greater effect on children's early maths skills, compared to direct instruction and free play (Skene et al., 2022). The principles of direct instruction, such as deliberately sequenced small units of information taught explicitly with repeated practice, are more characteristic of practice-based apps, such as *Maths 3–5 and Maths 4–6* (Outhwaite et al., 2018). In the current study, the reported effect sizes were descriptively larger for the parent-based app *Bedtime Math* (evaluated in a randomised control trial with >250 participants, Cohen's $d=0.82$), compared to the practice-based apps *Maths 3–5 and Maths 4–6* (evaluated in a randomised control trial with >250 participants, Cohen's $d=0.78$).

Mathematical content

The included apps primarily targeted basic skills in understanding number representations and relationships ($n=21$) with number representation with Arabic digits, verbal and/or written number word recognition as the most common ($n=19$) (see [Table S2](#)). This is also consistent with previous research showing a greater emphasis on basic counting skills in the classroom (von Spreckelsen et al., 2019) and a relatively reduced focus on spatial skills in formal classroom activities (Gilligan-Lee et al., 2022). However, mathematical development is complex and consists of a broad range of skills that children need to master (Butterworth, 2005; Clements & Sarama, 2009; Gilmore et al., 2018). As such, future app design for young children should consider a broad range of mathematical skills that is not limited to basic number skills (see [Table S2](#)), and evaluation studies should examine the role that maths apps can play in the development of these skills.

Feedback and levelling

Explanatory (with motivational) feedback ($n=12$) was the most common form of feedback within the included maths apps (see [Table S3](#)). Similarly, participatory free form levelling ($n=12$) was the most common form of levelling. In contrast, previous research has shown these app design features are rare within commercially available educational apps that have not been empirically evaluated but were rated the most popular on the app stores (Callaghan & Reich, 2018). This may, in part, reflect other evidence demonstrating that when teaching practitioners are choosing which educational apps to use with their children, they often place less value on feedback and learning theory, compared to scaffolding features and content aligned with taught curriculums (Montazami et al., 2022).

Results from the QCA demonstrated that variations in how feedback and levelling were implemented within the app design were associated with differences in the within-subject effect sizes of children's progress following the maths app intervention. Specifically, the QCA suggests that the combination of explanatory and motivational feedback, together with programmatic levelling (either dynamic or static), was a necessary condition within the app design for enhancing children's learning outcomes with app-based maths instruction. However, it is important to highlight that within the current sample, there were no educational maths apps that had explanatory feedback and programmatic levelling, and not motivational feedback. This means that the relative contributions of each type of feedback cannot be fully disentangled and should be the focus of future research using experimental methods. Nevertheless, individually, these app design features were sufficient for enhanced learning outcomes. Unlike previous reviews (eg, Kim et al., 2021), the current results support the role of well-designed educational apps for children's outcomes. In particular, these results are consistent with other experimental research that showed children made significantly fewer errors and completed app-based maths learning tasks quicker and with increased accuracy when explanatory feedback and programmatic levelling were present (Callaghan & Reich, 2021).

Limitations

Although the eight studies included in the QCA can be considered to have good levels of rigour using an experimental design (randomised control trial or quasi-experimental design) with a relatively heterogenous sample of typically developing children, some caution should be taken with these results due to the possibility of inflated effect sizes. Although Hedge's *g* corrections were applied as appropriate to sample sizes <50 (Lin, 2018), only two studies had overall final sample sizes over 250 children (Berkowitz et al., 2015; Outhwaite et al., 2018). Previous reviews have identified this benchmark as best practice for avoiding inflated effect sizes (Cheung & Slavin, 2013). In this QCA, both studies were not classified as a highly effective intervention set (within-subject effect size >1), relative to the other included studies.

Likewise, of the three studies that were identified in the highly effective intervention set (Grimes et al., 2020; Kosko & Ferdig, 2016; Wu, 2020), two used a researcher developed assessment of mathematical attainment as the primary outcome measure. Within this QCA, these issues cannot be statistically controlled for. Although the preliminary independent samples *t*-test showed no significant differences between effect sizes reported from studies using researcher developed or standardised assessment measures, it is still a potential caveat to consider when interpreting the study results.

Furthermore, the population characteristics within the included studies were predominantly typically developing children in English-speaking countries, United States and England (see Table S5). Despite the advantages of a relatively heterogenous sample for synthesising across studies, it may limit the generalisability of the current findings to other population characteristics, such as children with special educational needs and disabilities, including mathematical learning difficulties, as well as children from low socio-economic status and ethnic minority backgrounds, and those with English not as their first language. Future research will benefit from working with children from diverse backgrounds to ensure innovations with educational apps address, rather than exacerbate, inequalities in learning. Overall, the QCA was limited by the small number of apps that were eligible for inclusion ($n=8$). The remaining 15 apps could not be included because the associated studies did not provide sufficient information to calculate the necessary effect sizes on maths attainment outcomes. This also hindered the ability to include other types of apps (eg, productive apps)

in the QCA. Future research can address these limitations through improving the reporting standards of evaluation studies to support the synthesis of evidence, which is vital for advancing the field and informing evidence-based decisions.

Future research

Additionally, the current results can be used to inform the development of clear hypotheses in future research on how educational maths apps work to support learning of young children. In particular, future research should consider individual differences and the dynamics of feedback and levelling. For example, examining how these design features can be most optimally integrated for different groups of children, such as those with mathematical learning difficulties who may require more support than their typically developing peers (Vanbecelaere et al., 2021). In-app data could also be insightful for understanding the real-time dynamics of how children respond to the feedback and levelling features, and if this should be further adapted, as they progress and master the learning content (Bartoschek et al., 2013; Callaghan & Reich, 2021).

Implications and conclusion

This study provides a new, open-access framework for assessing the educational value of maths apps based on the type of app, the mathematical content and app design features included. Within the educational maths apps previously evaluated with young children, current findings demonstrate that there is an overall, relatively narrow focus on basic number skills. These skills are most frequently taught within the context of a practice-based app, where children are the users of the learning content. Further research is needed to evaluate the impact of different types of maths apps, particularly productive apps, where children are creators of their learning content and parent-based apps, which encourage off-screen playful interactions between children and their caregivers. The inclusion of more complex maths skills, which encompass the breadth and depth of holistic mathematical development are also needed.

The current study also found that the combination of explanatory and motivational feedback with programmatic levelling play an important role in enhancing children's learning outcomes with app-based maths instruction. This evidence suggests the feedback and levelling design features of specific apps should be considered by stakeholders when deciding which apps to use with young children. Nevertheless, further research is needed to examine how individual differences in response to these features may vary over time, and how these dynamics can be effectively implemented to ensure all children, regardless of their ability level, are able to access and learn from educational maths apps.

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.

FUNDING INFORMATION

The project has been funded by the Nuffield Foundation (grant number FR-000000370).

CONFLICT OF INTEREST STATEMENT

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

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Open Science Framework at www.osf.io/pzkmh.

ETHICS STATEMENT

Ethical approval for this study was granted by the IOE ethics committee (REC1376).

ORCID

Laura A. Outhwaite  <https://orcid.org/0000-0002-4454-7775>

Jo Van Herwegen  <https://orcid.org/0000-0001-5316-1818>

ENDNOTES

¹ Two maths apps were evaluated in multiple studies: (1) *onebillion Maths 3–5 and Maths 4–6* were evaluated in six other studies in the United Kingdom, Malawi and Brazil (Nunes et al., 2019; Outhwaite et al., 2017, 2018, 2020; Pitchford et al., 2018, 2019) and (2) *Math Shelf* was evaluated in two other studies in the United States (Schacter et al., 2016; Schacter & Jo, 2016). The effect sizes observed across these additional studies were largely consistent with those observed in the studies included in the current study (i.e. <1 and therefore not classified as a highly effective intervention set, see Table 2). The only exception was Outhwaite et al. (2017) which reported effect sizes greater than 1 across four quasi-experimental pilot studies with small sample sizes and researcher developed tools. As Outhwaite et al. (2017) were not considered the most robust of the seven studies evaluating the *onebillion* maths apps, it was not included in the current study.

² In QCA, necessary condition(s) are those that must be present all or most of the time for a defined outcome to occur. The absence of necessary condition(s) prevents the occurrence of the defined outcome. In contrast, sufficient condition(s) are those that when present lead to the defined outcome all or most of the time (Thomas et al., 2014).

REFERENCES

- Bartoschek, T., Schwering, A., Li, R., & Münzer, S. (2013). Ori-Gami—An app fostering spatial competency development and spatial learning of children. In *Proceedings of the 15th AGILE International Conference on Geographic Information Science*, Leuven, Belgium. Springer-Verlag.
- 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.
- Blair, K. P. (2013). Learning in critter corral: Evaluating three kinds of feedback in a preschool math app. In *Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 372–375).
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, 46(1), 3–18.
- Callaghan, M. N., & Reich, S. M. (2018). Are educational preschool apps designed to teach? An analysis of the app market. *Learning, Media and Technology*, 43(3), 280–293.
- Callaghan, M. N., & Reich, S. M. (2021). Mobile app features that scaffold pre-school learning: Verbal feedback and leveling designs. *British Journal of Educational Technology*, 52(2), 785–806.
- 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.
- 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.
- Clark, A., Henderson, P., & Gifford, S. (2020). *Improving Mathematics in the Early Years and Key Stage 1*. https://educationendowmentfoundation.org.uk/public/files/Publications/Maths/EEF_Maths_EY_KS1_Guidance_Report.pdf.
- Clements, D. H., & Sarama, J. (2009). *Learning and teaching Early math*. Routledge.

- Common Core State Standards Initiative. (n.d.). *Common core state standards for mathematics*. http://www.corestandards.org/wp-content/uploads/Math_Standards1.pdf
- Crompton, H., Burke, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A systematic review. *Computers & Education*, *110*, 51–63.
- Department for Education. (2013). *Mathematics programmes of study: Key stages 1 and 2*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335158/PRIMARY_national_curriculum_-_Mathematics_220714.pdf
- Department for Education. (2021a). *Education technology (EdTech) survey 2020–2021*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1057817/Education_Technology_EdTech_Survey.pdf
- Department for Education. (2021b). *Statutory framework for the early years foundation stage*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974907/EYFS_framework_-_March_2021.pdf
- Falloon, G. (2013). Young students using iPads: App design and content influences on their learning pathways. *Computers & Education*, *68*, 505–521.
- Gilligan-Lee, K. A., Bradbury, A., Bradley, C., Farran, E. K., Van Herwegen, J., Wyse, D., & Outhwaite, L. A. (2022). Spatial thinking in practice: A snapshot of teacher's spatial activity use in the early years' classroom. *PsyArXiv*. <https://doi.org/10.31234/osf.io/zqc2x>
- Gilmore, C., Göbel, S. M., & Inglis, M. (2018). *An introduction to mathematical cognition*. Routledge.
- 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), e20191579.
- Grimes, K. R., Park, S., McClelland, A., Park, J., Lee, Y. R., Nozari, M., Umer, Z., & Zapparoli, B. (2020). Effectiveness of a numeracy intelligent tutoring system in kindergarten: A conceptual replication. *Open Science Framework Pre-Print*. <https://osf.io/n2x7b/download>
- Herodotou, C. (2021). MAD learn: An evidence-based affordance framework for assessing learning apps. In *Proceedings of the 7th International Conference of the Immersive Learning Research Network (iLRN)* (pp. 1–8). <https://doi.org/10.23919/iLRN52045.2021.9459325>
- 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.
- Hooshyar, D., Yousefi, M., & Lim, H. (2018). A procedural content generation-based framework for educational games: Toward a tailored data-driven game for developing early English reading skills. *Journal of Educational Computing Research*, *56*(2), 293–310.
- Hsin, C. T., & Wu, H. K. (2011). Using scaffolding strategies to promote young children's scientific understandings of floating and sinking. *Journal of Science Education and Technology*, *20*(5), 656–666.
- Inal, Y., & Cagiltay, K. (2007). Flow experiences of children in an interactive social game environment. *British Journal of Educational Technology*, *38*(3), 455–464.
- Kanders, K., Hickman, M., & Bazalgette, L. (2022). *Could toddler tech help to get more children school ready?* <https://www.nesta.org.uk/project/mapping-parenting-technology/could-toddler-tech-help-to-get-more-children-school-ready/>
- Kay, R., & Kwak, J. Y. (2018). Comparing types of mathematics apps used in primary school classrooms: An exploratory analysis. *Journal of Computers in Education*, *5*(3), 349–371.
- Kim, J., Gilbert, J., Yu, Q., & Gale, C. (2021). Measures matter: A meta-analysis of the effects of educational apps on preschool to grade 3 children's literacy and math skills. *AERA Open*, *7*, 23328584211004183.
- Kolak, J., Norgate, S. H., Monaghan, P., & Taylor, G. (2020). Developing evaluation tools for assessing the educational potential of apps for preschool children in the UK. *Journal of Children and Media*, *15*(3), 410–430.
- 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.
- Kucirkova, N. (2018). Personalised learning with digital technologies at home and school: Where is children's agency? In *Mobile technologies in children's language and literacy*. Emerald Publishing Limited.
- Lin, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PLoS ONE*, *13*(9), e0204056.
- Magliaro, S. G., Lockee, B. B., & Burton, J. K. (2005). Direct instruction revisited: A key model for instructional technology. *Educational Technology Research and Development*, *53*(4), 41–55.
- Marsh, J., Murriss, K., Ng'ambi, D., Parry, R., Scott, F., Thomsen, B. S., Bishop, J., Bannister, C., Dixon, K., Giorza, T., Peers, J., Titus, S., Da Silva, H., Doyle, G., Driscoll, A., Hall, L., Hetherington, A., Krönke, M., Margary, T., ... Woodgate, A. (2020). *Children, technology and play*. The LEGO Foundation.
- Masterman, E., & Rogers, Y. (2002). A framework for designing interactive multimedia to scaffold young children's understanding of historical chronology. *Instructional Science*, *30*(3), 221–241.
- Meyer, M., Zosh, J. M., McLaren, C., Robb, M., McCaffery, H., Golinkoff, R. M., Hirsh-Pasek, K., & Radesky, J. (2021). How educational are “educational” apps for young children? App store content analysis using the four pillars of learning framework. *Journal of Children and Media*, *15*, 1–23.

- Montazami, A., Pearson, H. A., Dubé, A. K., Kacmaz, G., Wen, R., & Alam, S. S. (2022). Why this app? How educators choose a good educational app. *Computers & Education*, *184*, 104513.
- Moyer-Packenham, P. S., Bullock, E. K., Shumway, J. F., Tucker, S. I., Watts, C. M., Westenskow, A., Anderson-Pence, K. L., Maahs-Fladung, C., Boyer-Thurgood, J., Gulkilik, H., & 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., Bryant, P., & Evangelou, M. (2019). *Onebillion: Evaluation report*. Education Endowment Foundation (EEF).
- Ofcom. (2020). *Children's media lives: Life in lockdown*.
- 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>
- 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), 284–298.
- 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.
- Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2017). Designing and creating an educational app rubric for preschool teachers. *Education and Information Technologies*, *22*(6), 3147–3165.
- Parks, A. N., & 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*, *35*, 602–615. <https://doi.org/10.1080/02568543.2020.1809578>
- 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-Saharan Africa. *Developmental Science*, *22*(5), e12864.
- 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). <https://doi.org/10.3389/fpsyg.2018.00262>
- 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.
- 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.
- Skene, K., O'Farrelly, C. M., Byrne, E. M., Kirby, N., Stevens, E. C., & Ramchandani, P. G. (2022). Can guidance during play enhance children's learning and development in educational contexts? A systematic review and meta-analysis. *Child Development*, *93*(4), 1162–1180.
- 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)* (pp. 610–617).
- Sutcliffe, K., Thomas, J., Stokes, G., Hinds, K., & Bangpan, M. (2015). Intervention component analysis (ICA): A pragmatic approach for identifying the critical features of complex interventions. *Systematic Reviews*, *4*(1), 140.
- Thomas, J., O'Mara-Eves, A., & Brunton, G. (2014). Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: A worked example. *Systematic Reviews*, *3*(1), 1–14.
- Tucker, S. I. (2015). *An exploratory study of attributes, affordances, abilities, and distance in children's use of mathematics virtual manipulative iPad apps* [Doctoral dissertation, Utah State University]. All Graduate Theses and Dissertations. <https://digitalcommons.usu.edu/etd/4592>
- Vanbecelaere, S., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2021). The effectiveness of an adaptive digital educational game for the training of early numerical abilities in terms of cognitive, noncognitive and efficiency outcomes. *British Journal of Educational Technology*, *52*(1), 112–124.
- Vandewaetere, M., & Clarebout, G. (2014). Advanced technologies for personalized learning, instruction, and performance. In *Handbook of research on educational communications and technology* (pp. 425–437). Springer.

- von Spreckelsen, M., Dove, E., Coolen, I., Mills, A., Dowker, A., Sylva, K., Ansari, D., Merkley, R., Murphy, V., & Scerif, G. (2019). Let's talk about maths: The role of observed "maths-talk" and maths provisions in preschoolers' numeracy. *Mind, Brain, and Education*, 13(4), 326–340.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wu, R. (2020). *The effect of the math application MathemAntics on preschoolers' math performance* [PhD Thesis]. Columbia University.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Outhwaite, L. A., Early, E., Herodotou, C., & Van Herwegen, J. (2023). Understanding how educational maths apps can enhance learning: A content analysis and qualitative comparative analysis. *British Journal of Educational Technology*, 00, 1–22. <https://doi.org/10.1111/bjet.13339>