



Touchscreen apps for child creativity: An evaluation of creativity apps designed for young children

Sophie A. Booton, Pinar Kolançali^{*}, Victoria A. Murphy

Department of Education, University of Oxford, UK

ARTICLE INFO

Keywords:

Elementary education
Touchscreen apps
Creativity
M-learning
21st century abilities

ABSTRACT

In the last decade, touchscreen apps have become a staple for educating and entertaining children. The number of apps that claim to support child creativity has been on the rise. Pedagogical or developmental qualities of these apps in supporting child creativity and learning, however, have not been extensively studied from a research angle. This study takes a systematic approach to reviewing a sample of commercially available apps to determine whether existing apps are well designed to support children's creativity. A sample of 152 Android apps targeting 4- to 12-year-olds which claimed to involve creativity was obtained from 3 sources. The quality of apps was rated using criteria based on research evidence about factors that support creativity (e.g., experimentation, modelling). The apps sampled tended to be for younger children, to be open-ended, for solo use and require in-app interaction, and involved mainly visual arts, and personal and social-emotional activities. Quality ratings for creativity were overall low, particularly in the areas of supports for convergent thinking and modelling creativity. Key app store data (review score, installs, payment and expert approval) were not a consistent predictor of creative app quality. Apps for older children also tended to score more highly. Thus, current apps for children's creativity are low in quality and app store data provides limited indication of their quality. Implications for parents, educators and app designers are considered.

1. Introduction

Creativity is a critical skill for children to acquire (Ananiadou & Claro, 2009). Several environmental factors that can impact creativity have been identified in experimental research (e.g., strategies for divergent thinking, and modelling of diverse examples, (Hoicka, Powell, Knight, & Norwood, 2018; Sun, Wang, & Wegerif, 2019). Using technological devices for creative activities such as drawing or playing with digital games supporting creative thinking have been shown to boost children's creativity (Bitu, Gali-non-Mélénec, & Molina, 2022; Xiong, Liu, & Huang, 2022). Meanwhile, children are increasingly using apps to learn informally (Marsh et al., 2015; Ofcom, 2021), and research evidence has shown that some app features play an important role in supporting child creativity and play (Marsh et al., 2018). Although many apps now claim to boost children's creativity, the characteristics of these apps have not been studied systematically. Thus, little is known about the potential of these new technologies to enable and support children's creativity, especially in an era where contemporary skills in creativity (e.g., coding) are ultimately technology-bound. For these reasons, understanding the qualities of these apps is, therefore, timely and necessary. To achieve this goal, the present study takes a systematic approach to reviewing a sample of commercially available apps to determine whether existing apps are well designed to

^{*} Corresponding author. 15 Norham Gardens, Oxford, OX2 6PY, UK.
E-mail address: pinar.kolançali@education.ox.ac.uk (P. Kolançali).

support children's creativity.

1.1. Mobile apps in Children's learning

Mobile touchscreen apps comprise a significant proportion of children's digital environment (Common Sense Media, 2017; Kabali et al., 2015; Ofcom, 2021). Apps for children make up around half of the top-selling apps on the market (Shuler, 2009), game and education categories being in the top three most downloaded categories both in the Apple and Google App stores (Statista, 2022a; 2022c). There is a growing role for mobile applications in children's learning, due to their ability to enable learning and play in informal contexts (Erstad, 2012); ease of use for young children (Marsh et al., 2015; Vavav, Cramariuc, & Schipor, 2015); and affordances such as multimedia, sensitivity to the physical context (Pachler, Bachmair, & Cook, 2010), speech recognition (Baker, 2017), and many more.

To assess the quality of apps available on the app market, researchers and policy-makers have created criteria to appraise the educational quality of apps for children by drawing on intervention research, education theory, and research with older technologies and touchscreen apps (Callaghan & Reich, 2018; Department for Education, 2019; Hillman & Marshall, 2009; Hirsh-Pasek et al., 2015; Kolak, Norgate, Monaghan, & Taylor, 2021; Papadakis, Kalogiannakis, & Zaranis, 2017). Apps on the market labelled as "educational" tend to be higher in quality than pure entertainment apps (Taylor, Kolak, Norgate, & Monaghan, 2022), and researchers have identified some commercially available apps as higher quality (Department for Education, 2019; Hirsh-Pasek et al., 2015; Meyer et al., 2021). However, the overall educational quality of these apps is nonetheless relatively low (Callaghan & Reich, 2018; Kolak et al., 2021; Meyer et al., 2021; Papadakis, Kalogiannakis, & Zaranis, 2018). Many apps on the market also contain a range of activities with diverse learning goals and domains (such as letter knowledge, memory, counting, and drawing), and such diversity of aims may have an adverse impact on app quality; indeed, award-winning apps are more likely to target just one or two specific domains (Vaala & Levine, 2015).

Given the variable quality of available apps, it is important to consider how parents and practitioners can make informed choices about which apps children use. To find apps, parents search using keywords, look within known categories in app stores (e.g., educational games), and check recommendations on parent-oriented websites (Marsh et al., 2015; Vaala & Levine, 2015). Critical for parents and educators' selection of apps is the information that is available in the app store, such as the average user review score, cost, and keywords. It appears that the information available in app titles and descriptions can be lacking in such key details as the educational domain of the app (e.g., mathematics) (Dubé, Kacmaz, Wen, Alam, & Xu, 2020) or target age (Papadakis et al., 2018; Vaala & Levine, 2015). However, where this information is available, it could possibly provide cues as to the quality of the app. Thus, it is important to examine whether this information is in fact predictive of the educational quality of apps.

Research has explored the relationship between certain app characteristics and app store information and the educational quality. One feature that parents might use to inform their judgement of app quality is cost. Some data suggested that apps that are paid for have slightly higher educational quality ratings than apps that are free to download (Meyer et al., 2021), although this relationship may be due to distraction from advertising being a key aspect of quality rating, as adverts tend to be present in free apps. Indeed, a study using different quality criteria did not replicate this finding (Kolak et al., 2021). Another factor is customer reviews. Two studies found that the average user rating in the app store had no relationship to the educational features or a quality score for educational apps (Dubé et al., 2020; Papadakis et al., 2018). Expert reviews can also be informative, with one study suggesting that the highest and lowest expert rated apps from two independent app rating websites differed in quality (Taylor, Kolak, Bent, & Monaghan, 2022). Another study also found no relationship between the number of app downloads and the quality of educational apps for pre-schoolers (Papadakis et al., 2018). The source of apps may also be significant, as one study found that literacy apps listed on an expert rating website (rather than a list on the app store) were more likely to contain stories or narratives than just closed-ended games (Vaala & Levine, 2015). Finally, one study examined the impact of learning domain (comparing mathematics and literacy apps) on the presence of app features for learning, finding no differences in modelling, feedback, reward systems, or touch interactions, although math apps were more likely to have adaptive levels of challenge (Callaghan & Reich, 2018).

Thus, a handful of studies have looked at how some app characteristics and app store data predict educational app quality. The results have been mixed and few studies compared the same specific variables or controlled for multiple variables to examine the predictors with the greatest impact. Many other factors could also be considered, including the target age, locus of app use (inside or outside the app), degree of open-endedness of activities (i.e., whether activities have correct answers or not), the recency with which an app has been updated, the presence or absence of an "expert approval" badge within the app store, the age of the app, its download size, and the relevance of the app description to the learning goal.

1.2. Mobile apps and creativity

As part of supporting learning more broadly, mobile applications have the potential to develop creativity specifically. Creativity can most simply be defined as the ability to produce ideas or products that are both novel to the creator and appropriate to the task (Plucker, Beghetto, & Dow, 2004). Creativity requires obtaining both domain-specific (Baer, 2015) and domain-general skills (Hong & Milgram, 2010) and involves both divergent thinking – the capacity to generate multiple and varied problems, ideas, or solutions – and convergent thinking – the ability to evaluate, select, and problem-solve to obtain the best solution (Guilford, 1967). The material environment has a fundamental influence on the expression of creativity. Affordances of the social and material environment provide resources, inspiration, modelling, and constraints for creative ideation and expression which encourage or discourage certain kinds of creation (Glăveanu, 2013; Kaufman & Glăveanu, 2019). Thus, the environment in which a child operates can be low or high in

potential for supporting creativity, and this is thought to be critical to their creative outcomes.

Indeed, research has suggested a wide variety of factors that may be present in the environment and affect children's expression of creativity. These studies (summarised in Table 1) reflect a range of methodologies and are conducted in primarily non-digital media contexts. They suggest that environments that support creativity should develop relevant (1) domain-specific competences. For example, to build creativity in poetry, support for poetry-related skills and knowledge is needed. They should also allow for and support (2) divergent thinking, by providing activities that can be completed in different ways, encouraging the generation of multiple different ideas, and providing strategies to do so, and (3) convergent thinking, by providing creative problem-solving activities and supports and strategies to work towards effective solutions. They should enable (4) experimentation through encouraging risk-taking and providing varied materials to use and provide (5) inspiration for creation or problem-solving through props to manipulate, pre-made materials to explore, and randomised prompts or materials. Environments should also provide many, good quality and diverse examples as (6) modelling of creative products and/or processes which encourage different responses to the task rather than imitation. Creative environments will also include opportunities and supports for (7) meta-cognition, including reflection on the creative process and evaluation of creative products, and build creative (8) self-efficacy through encouraging a growth mindset towards creativity and creating a self-affirming environment. Thus, several studies have considered how different factors in the physical environments of the laboratory or classroom might impact creativity, although few have investigated how these factors might play out in the context of digital environments.

There is growing evidence that digital applications offer a potential environment for supporting children's creativity. Constructionism has long suggested that creating and working with artefacts – digital or physical – can support self-directed learning (Papert, 1996), as long as this interaction is active, in that “the child programs the computer” rather than “the computer ... being used to program the child” (Papert, 1980). Indeed, children use a range of applications in their daily lives which have the potential to support play and creativity (Marsh et al., 2015) across varied learning domains, such as visual arts, story-telling, and coding. The majority (65%) of parents of 3 to 5-year-olds also cite encouraging play and creativity as a reason for downloading apps (Marsh et al., 2015). Several qualitative studies suggest the potential affordances of specific teacher or experimenter-chosen digital tools to support creativity. For example, one case study showed that a 3-year-old demonstrated more risk-taking in terms of choice of subject with a computer-based digital drawing app than when drawing on paper (Kucirkova & Sakr, 2015). A study of a class of 4 to 5-year-olds identified that children interpret pre-made images (e.g., stickers) and the spatio-temporal properties of digital drawing in varied ways to construct pictures and narratives (Sakr, Connelly, & Wild, 2016). Furthermore, naturalistic observational case studies have suggested that preschool children can use many types of creative thinking (exploration, involvement, enjoyment, and persistence) when playing with apps at home (Marsh et al., 2018), although this depended on the type of app: for example, more types of play were observed with drawing, building, and role-playing apps than with running and puzzle video games. Such studies have raised many intriguing hypotheses about the potential link between design features of digital technologies and children's creative behaviour.

Other studies have also shown how particular aspects of creativity apps may contribute to child creativity. One study, for example, demonstrated that developmentally appropriate creativity apps were more likely to increase children's engagement with the activities, although it may not lead to variations in children's creative production (Piotrowski & Meester, 2018). In a similar vein, using developmentally appropriate features in creative design activities such as drawing (e.g., features that do not amplify cognitive load on the user) has been found to improve children's originality in their creative production (Bitu et al., 2022). Another study has shown that a researcher-developed app game designed to support creative thinking strategies (e.g., divergent thinking) for preschool-aged children may contribute to the creative thinking skills of children, especially those who are in the older age group (Xiong et al., 2022).

Emerging evidence from research on apps and creativity suggests that we have limited knowledge on the aspects of creativity apps that may be used to cultivate child creativity. There has not to date been any quantitative assessment of the potential of existing apps on the market to support creativity. Indeed, no criteria have been devised to assess the quality of creativity apps specifically, as opposed to learning generally (Piotrowski & Meester, 2018) or for mathematics or literacy (Callaghan & Reich, 2018; Department for Education, 2019; Highfield & Goodwin, 2013). By mapping the literature about factors affecting creativity onto the features of mobile applications, we can assess the potential for supporting development in this area. This provides a first step to determining the quality of the existing app market and the predictive features of this quality.

Table 1
References for the literature on environmental features supporting creativity.

Feature	References
1 Domain-specific skills	Amabile & Pratt, 2016; Baer, 1996, 2015; Winner et al., 2020
2 Divergent thinking	Alfonso-Benlliure & Santos, 2016; Glăveanu, 2013; Khatena, 1971; Nusbaum & Silvia, 2011; Puckett Cliatt et al., 1980; Segundo Marcos et al., 2020; Sowden, Clements, Redlich, & Lewis, 2015; Ward & Kolomyts, 2019; Weatherford, Esparza, Tedder, & Smith, 2020
3 Convergent thinking	(Fessakis, Gouli, & Mavroudi, 2013, 2014; Knoblich, Ohlsson, Haider, & Rhenius, 1999; Medeiros, Partlow, & Mumford, 2014; Nielson & Powless, 2007)
4 Experimentation	(Guegan, Brechet, & Nelson, 2020; Richardson & Mishra, 2018; Robson, 2014; Webb & Rule, 2013; Winner et al., 2020)
5 Inspiration	(Alfonso-Benlliure & Santos, 2016; Baird et al., 2012; Dorin, 2013; Doron, 2016, 2017; Duarte-García & Sigal-Sefchovich, 2019; Dunn & Sweeney, 2018; Hoffmann & Russ, 2016; Kirsh, 2014; Leong et al., 2008; Sun et al., 2019)
6 Modelling	(Ali, Park, & Breazeal, 2021; Ho, Lin, Chen, & Lee, 2017; Hoicka et al., 2018; Ward & Kolomyts, 2019)
7 Meta-cognition	(Robson, 2014; van de Kamp, Admiraal, van Drie, & Rijlaarsdam, 2015; Winner et al., 2020)
8 Self-efficacy	(Karwowski et al., 2019; Warren, Mason-Apps, Hoskins, Azmi, & Boyce, 2018; Wen, Butler, & Koutstaal, 2013)

1.3. The present research

The primary objective of this study is to obtain a picture of the creativity app market for primary school children; the quality of these apps; and what factors predict quality. Thus, the research questions (RQs) were as follows.

Question 1. Are creativity apps more likely to contain certain app characteristics (e.g., in-app interactions, solo play, arts domain) than other app characteristics (e.g., out of app interactions, social play, sciences domain)?

Question 2. Are creativity apps higher quality in some areas than others?

Question 3. Does app store data (e.g., number of installs, customer review, cost, update recency) and app characteristics of creativity apps predict app quality?

To address these questions, a methodical review of the creativity app market for primary school aged children was undertaken, with app store data recorded and apps categorised for their app characteristics and rated for quality. Three search methods were applied to find applications, and these were screened for eligibility to obtain a sample of 152 creativity apps targeted at children. App characteristics were categorised, and apps were assessed for quality using pre-determined, evidence-based criteria.

2. Method

2.1. Design

A correlational design was used. The dependent variables were app quality ratings for creativity. The independent variables included data from the app store (number of installs, customer review score, expert approval, cost, number of reviews, age of the app, update recency, size, keyword tokens, keyword types); and app characteristics (target age, domain, subdomain, out-of-app engagement, social interaction, open-endedness, total number of app activities), of which 15 key predictors were pre-registered to be included in regression models. The full methods and analyses were pre-registered on the Open Science Framework (https://osf.io/uft7d/?view_

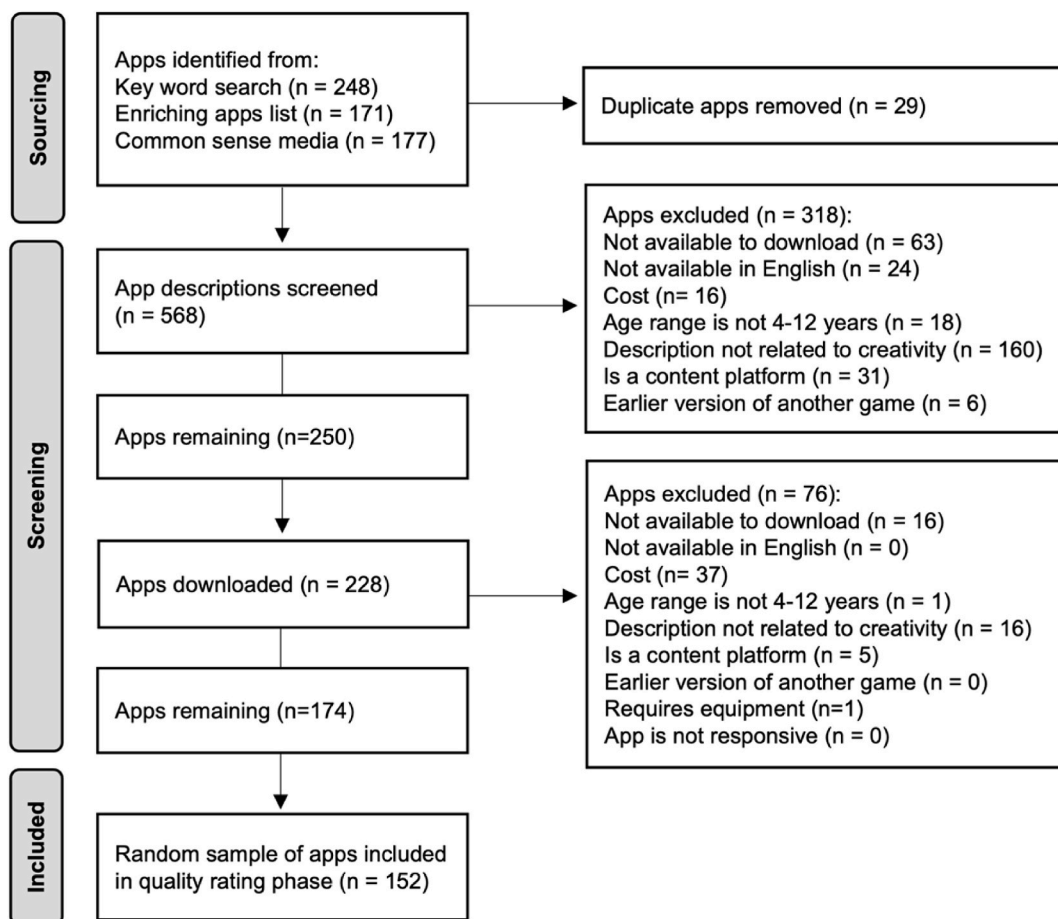


Fig. 1. Flow diagram of sampling creativity apps for children (Page et al., 2021).

only=1cab8f467b0b44c78a86233cf63197d4) and any exploratory deviations from this are noted.

A review of ethical considerations was carried out in accordance with the authors' university ethics procedures and the guidance provided by the Association of Internet Researchers (franzke, shakti, & Anja, 2020). No personal data from human participants was collected for this study, and only publicly available data was utilized. The research team followed existing good practices in digital research, such as only collecting the data necessary for analysis, and storing the data securely (Eynon, Fry, & Schroeder, 2017).

2.2. App sampling strategy

A summary of the stages of the app review process is shown in Fig. 1. The full list of apps reviewed and their ratings can be found at https://osf.io/uf7d/?view_only=1cab8f467b0b44c78a86233cf63197d4.

2.2.1. Sample size

A priori power calculation (using G Power) was conducted (for linear multiple regression: fixed model, single regression coefficient) and showed that for a two-tailed regression with 15 predictors and an alpha level of 0.05, to obtain 80% power to detect a single regression coefficient with an effect size equivalent to partial $R^2 = 0.05$ (small to medium), a sample of 152 apps was needed. Thus, the sample size for the app review was set at 152 apps.

2.2.2. Search strategy

Android apps were sourced from the Android operating system (Google Play) instead of the competing iOS. Google Play has been identified as the largest app store (Statista, 2022b) and Android has been shown to be the most accessible mobile operating system reaching 75.9% of the mobile device users worldwide (International Data Corporation, 2022). Three approaches to searching for apps were used. Firstly, all results were extracted from Google Play's 'Enriching Apps' lists in the 'Kids category', because this list contains popular educational apps across many domains. Secondly, all results were extracted by searching Google Play for the key terms 'children kids creat*', to target apps labelled as creative. Thirdly, all results were extracted from Common Sense Media's (commonsensemedia.org) app reviews, which were filtered for Android, Creativity, and any ages from 4 to 12 years, to identify apps tagged by independent reviewers as related to creativity. All app titles were transferred to a spreadsheet and duplicates were identified, their source noted, and removed.

2.2.3. Exclusion criteria

App descriptions were then screened, and some excluded based on the following criteria.

- 1 Is not available to download in the UK on Android from Google Play.
 - 2 Is not available in English.
 - 3 Main app (including all key activities) is not available for free or one-off payment of £5 or less (e.g., requires a subscription or larger in app purchase).
 - 4 Target age range is not 4–12 years.
 - 5 App content is not related to creativity: the app description of core activities does not include use of any of the following keywords to describe app activities: creat*/original*/innovat*/divergent/idea*/experiment */invent*/imagin*/inquir*/design*/mak*/build*/craft*/problem-solv*
 - 6 Is a platform (e.g., video, music, ebook or gaming platform) for accessing highly diverse contents from different creators.
 - 7 Is an earlier version of another game included in the review.
 - 8 Requires special equipment other than a tablet device for most app features to be used (e.g., associated toys or books, music recording equipment).
- Once downloaded, apps were checked for eligibility again and could also be excluded if:
- 9 Is not responsive or crashes consistently once downloaded.

The number of apps excluded based on each criterion is shown in Fig. 1. To ensure that the exclusion criteria were applied consistently, a second rater categorised 20% of the apps ($n = 114$). Inter-rater agreement was high (Cohen's Kappa = .858).

2.3. Measures

2.3.1. App store data

Data for apps were extracted from the Google Play webpage of each application. These consisted of the average customer review score (ranging from 1.0 to 5.0); number of installs (in one of 19 categories, from 1+ to 1 billion +); the cost (free or paid); expert or teacher approval¹ (approved/not); the number of user reviews; the age of the app (days between release date and download date); the update recency (days between last update and download date); and size (file size in MB); keyword tokens (total number of creativity related keywords in app description, including repeats); and keyword types (number of different creativity related keywords in app

¹ During data collection, Google Play changed the labelling of their approval rating from "Expert approved" to "Teacher approved". The apps to which the label applied did not change.

description). If an app was free to download but required an in-app payment of £5 or less to access more than 50% of the key activities, this was purchased, and the app cost was coded as “paid”.

2.3.2. App characteristics

Five app characteristics were coded. These are explained in Table 2. To assess the reliability of the coding of app characteristics, a second rater categorised 20% of the screened apps ($n = 31$). Agreement was good for domain ($Kappa = .785$), and social interaction ($ICC = 0.800$). Whilst percentage agreement was relatively high for all other variables at this stage (63–94%), Cohen's Kappa and intraclass correlation values were low, primarily due to some categories having very low frequency. According to the pre-registration, discrepancies were discussed and resolved, and a second round of 20% of apps were coded. Inter-rater agreement at this point was good to excellent for all variables (out-of-app engagement: $ICC = 1.00$; 100% agreement; open-endedness: $ICC = 0.777$; 87% agreement; Age: $Kappa = .663$; 87% agreement). Remaining discrepancies were resolved through discussion.

2.3.3. Quality

2.3.3.1. Creative quality. As no previous research has assessed the quality of apps for supporting creativity, criteria were created for this study. Based on literature on factors supporting children's and adult's creativity (see Table 1), 8 factors were selected and criteria devised to evaluate the presence of app features that could enable these 8 factors. These were: domain-specific skills, divergent thinking, convergent thinking, experimentation, inspiration, modelling, meta-cognition, and creative self-beliefs. The criteria were tested out and refined over several rounds of coding with $N = 37$ apps. The full rubric is shown in Appendix A. To assess inter-rater reliability of the creative quality coding, a second rater scored 20% of the apps ($n = 31$). Inter-rater reliability ranged from good to excellent ($ICC = 0.738$ – 0.940), except for inspiration ($ICC = 0.464$). Discrepancies were discussed and resolved, and a second round of 20% of apps were coded for this variable. Inter-rater agreement at this point was good ($ICC = 0.950$). Remaining discrepancies were resolved through discussion.

To determine the underlying structure of creative app quality ratings, exploratory factor analysis was conducted, specifically principal component analysis (PCA) with varimax rotation, with the number of factors determined by parallel test and scree plot, and belonging to sub-scales determined by factor loadings >0.4 . The PCA identified 2 factors which explained in total 61% of the variance in the criteria. The first (which we will call divergent creativity) loaded on divergent thinking, experimentation, inspiration, meta-cognition, and creative self-efficacy. The second (which we will call convergent creativity) loaded on convergent thinking, domain-specific skills, and modelling. Reliabilities for the 2 sub-scales according to Cronbach's alpha were good ($\alpha = 0.838, 0.604$). Averages of the raw scores were used to calculate each subscale score.

2.3.3.2. General quality. A set of criteria based on previous research (Hirsh-Pasek et al., 2015; Meyer et al., 2021) were devised to assess features of apps known to support learning generally, though not specifically creativity. More details about these criteria, score calculation, and results relating to these are included in supplementary materials. Three variables were used as control variables in our regression analyses here: general: meaning (comprising active learning, meaningful learning, and social supports), general: usability (comprising engagement and design suitability) and general: structure (comprising scaffolding and feedback).

2.4. App rating procedure

Both raters used a Samsung Galaxy Tab A tablet 8.0" (2019) to access the apps. Screened apps were downloaded to the raters' devices. One rater pre-selected which parts of apps would be coded using the following criteria. If apps specified (either in the Google Play app description or within the app) which activities were designed to support creativity, only these activities were included. If not, all activities could be included. If apps contained more than 5 activities, or more than 3 examples or levels (e.g., coloring templates), 5 activities and 3 examples per activity were sampled to assess using random number generation.

When reviewing apps, raters were required to close all other apps, to have internet access and sound turned on, and to allow access to the camera or gallery if prompted by the app. Raters accessed the menus, parent area, settings and the pre-selected activities and examples. Raters produced varied responses (correct, incorrect, delayed, “rule-breaking” e.g., drawing outside the lines), tested out all available functions (e.g., hints, erase, change tool etc.), progressed to the end of activities to view feedback and rewards, and repeated activities to assess adaptation between sessions. For apps with multiple activities which differed, raters provided an averaged rating across activities.

2.5. Data preparation

Data were first checked for outliers and values 3 SD above or below the mean for all variables were removed. Missing data (due to outliers or information absent from the app store) was limited ($n \leq 4$ cases for all variables) and was imputed via multiple imputation with chained equations. Variables with high skewness were transformed using a process of increasing strength of transformations (square root, cube root, logarithmic, reciprocal), with the transform that produces the lowest level of skewness selected (Osborne, 2010).

Table 2
Summary of app characteristics.

Design feature	Description	Type	Values/labels
Target age	Target age range of the app based on design and content	Categorical (mutually exclusive)	4-7 (younger) 8-12 (older)
Domain	Domains covered by app activities reviewed (one per activity in the app)	Binary categorical (x7) (non-mutually exclusive)	Visual arts Performing arts Language Science, Technology, Engineering & Math (STEM) Humanities Personal, social and emotional Cognitive-motor skills
Subdomain	Subdomains covered by app activities	Categorical (non-mutually exclusive)	34 categories e.g., drawing, poetry, architecture, biology
Out of app engagement	To what extent activities are within the app or outside the app	Ordinal scale	Mostly in app (−1) Mixed (0) Mostly out of app (1)
Social interaction	To what extent the app contains content which can be used alone vs. promotes social interaction	Ordinal scale	Mostly solo (−1) Mixed (0) Mostly social (1)
Open-endedness	To what extent the app contains tasks which are open-ended or closed-ended.	Ordinal scale	Mostly closed (−1) Mixed (0) Mostly open (1)

IRR = Inter-Rater Reliability.

3. Results

3.1. RQ1: are creativity apps more likely to contain certain characteristics than others?

The descriptive statistics for the app characteristics across the 152 included apps are shown in Table 3. To examine whether some app characteristics were statistically more frequent than others, comparisons were made between categories. For target age, a chi-squared goodness-of-fit test confirmed that significantly more apps had content primarily suitable to younger children (those aged 4–7) compared to older children (those aged 8–12) ($\chi^2(1) = 53.29$, $p < .001$).

For domains of learning, a Cochran's Q test confirmed that some domains were more common in the apps than others ($\chi^2(6) = 162.28$, $p < .001$). As shown in Table 3, visual arts were by far the most common, followed by personal and social education, and cognitive skills, language and STEM, with performing arts and humanities being the least common. The subdomains represented by apps are shown in Fig. 2. The subdomains of activities which were most common across apps were personal, drawing, perceptual, and crafts.

For each of the three app characteristics with ordinal data (engagement, interaction, activity), a chi-squared goodness-of-fit test was first performed with 3 groups. Significantly more apps involved primarily or only in-app engagement rather than mixed or out of app ($\chi^2(2) = 231.09$, $p < .001$). Significantly more apps were intended primarily or only for solo use rather than mixed or primarily social use (0.7%) ($\chi^2(2) = 269.24$, $p < .001$). Significantly more apps involved primarily open-ended activities rather than mixed or

Table 3
Descriptive statistics for app characteristics across the 152 included apps.

Target age	Older		Younger			
	N	%	N	%		
	131	79.6	21	20.4		
Domain	Yes		No			
	N	%	N	%		
Visual arts	76	50.0	76	50.0		
Performing arts	6	3.9	146	96.1		
Language	14	9.2	138	90.8		
STEM	15	9.9	137	90.1		
Humanities	3	2.0	149	98.0		
PSE	44	28.9	108	71.1		
Cognitive skills	30	19.7	122	80.3		
Design	Low		Mixed		High	
	N	%	N	%	N	%
Out of app engagement	139	91.4	8	5.3	5	5.3
Social interaction	146	96.1	5	3.3	1	0.7
Open-endedness	42	27.6	28	18.4	82	53.9

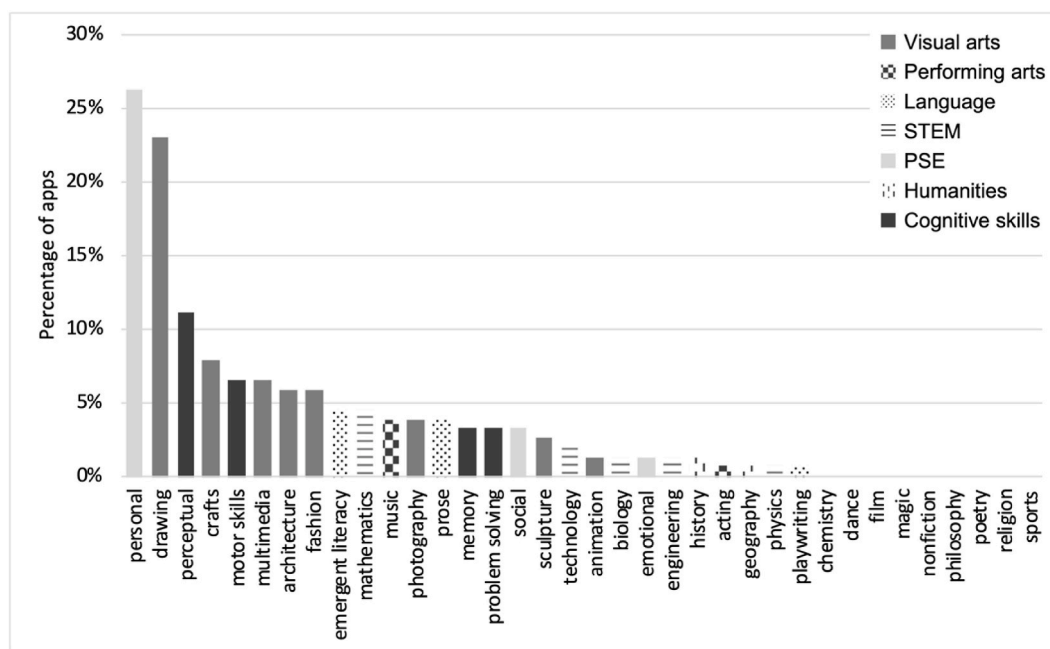


Fig. 2. Percentage of apps containing at least one activity for each subdomain coded.

primarily closed-ended activities ($\chi^2(2) = 31.00, p < .001$).

3.2. RQ2: are creativity apps higher quality in some areas than others?

Descriptive statistics for the app quality ratings, including subscales and each criterion item are shown in Table 4. Creative quality was low, with average scores for both divergent ($t(151) = -8.03, p < .001$) and convergent creativity ($t(151) = -15.21, p < .001$) significantly below the midpoint of the scale of 1 according to one-sample t-tests. To assess whether apps scored higher in creative quality on some items than others, a one-way repeated measures ANOVA with 8 levels was conducted, which confirmed a significant effect of criterion ($F(7,151) = 49.792, p < .001$). Planned comparisons with t-tests demonstrated that: experimentation > domain-specific skills > divergent thinking = metacognition = inspiration = creative self-efficacy > convergent thinking = modelling (see Appendix B for full test results).

3.3. RQ3: do app store data and app characteristics of creativity apps predict app quality?

Descriptive statistics for app store data are shown in Table 5. Zero-order correlations between variables included in regression models are shown in Table 6. Correlations for app store variables included only for descriptive purposes and four additional exploratory variables are shown in Appendix D (the total number of app activities, and the search source(s) of the app). It is notable that for both divergent and convergent creative quality, there were significant positive correlations with number of relevant keywords, number of reviews, and being listed on common sense media, and significant negative correlations with being listed in the keyword search, total number of app activities, and the cognitive skills domain. Other particularly strong correlations were positive associations between divergent creativity and open-endedness, and convergent creativity and target age.

To assess which factors contribute most to creative app quality, two hierarchical multiple linear regressions were conducted predicting convergent and divergent creativity scores. App data variables were entered in the first step, followed by all app characteristics, followed by general learning quality scores.²

The results of the regression models for creativity are shown in Table 7. For divergent creativity, model 1, which contained app store data only, explained 2.1% of the variance, which was not significant ($F(4,147) = 0.81, p = .524$), and none of the individual coefficients were significant. In the 2nd model, app characteristics of the apps were added: this explained a further 56.9% of the variance, and the change in R squared was significant ($F(8,139) = 24.18, p < .001$). In this model, there was a significant effect of age, and open-endedness, indicating that apps for older children and with more open-ended activities tended to score more highly than

² The only deviations to these models from the pre-registration were 1) to remove the 'social interaction' characteristic, due to its overlap with the general learning criterion of social supports, and 2) to remove the humanities and performing arts domains, due to their very low frequencies ($n = 3$ and $n = 6$ respectively).

Table 4

Descriptive statistics for creative app quality ratings.

	Min.	Max.	Mean	Std. Dev.
Creativity: convergent	0.00	2.00	0.48	0.425
Creativity: divergent	0.00	1.60	0.68	0.491
Domain-specific skills	0	2	0.87	0.523
Divergent thinking	0	2	0.61	0.515
Convergent thinking	0	2	0.30	0.598
Experimentation	0	2	1.15	0.867
Inspiration	0	2	0.55	0.550
Modelling	0	2	0.26	0.581
Metacognition	0	2	0.59	0.635
Creative self-efficacy	0	2	0.51	0.515

Table 5

Descriptive statistics for app store data.

	Min	Max	Mean	SD
Review score	3.0	4.9	4.108	0.312
Paid	0	1	0.45	0.499
Number of installs	6	17	12.02	2.078
Update recency	0	293	60.2743	70.54
Keyword tokens	1	15	4.08	3.418
Expert approval	0	1	0.47	0.501
Number of reviews	0	1,237,413	35142.59	121,907
Age of app	5	137	58.07	30.94
Download size	2.60	463.0	116.9673	85.13
Keyword types	1	6	2.01	1.136
Number of app activities	1	60	3.7961	7.369
Source: enriching apps list	0	1	0.24	0.431
Source: keyword search	0	1	0.40	0.492
Source: common sense media	0	1	0.38	0.487

those for younger children or with more closed-ended activities in terms of divergent creativity. In the 3rd and final model, convergent creativity quality and the three general quality sub-scores were included as predictors and this added a further 8.1% of the variance, which was again significant ($F(4,135) = 8.29, p < .001$). After controlling for these predictors, the effect of activity remained, but the effect of age was no longer significant. Furthermore, the effects of some domains and out of app engagement became significant and negative: specifically, there was a negative effect for the domains of language and STEM. This suggests that after controlling for other aspects of app quality, these domains and out of app engagement are associated with lower divergent creative quality scores. Finally, general meaning score emerged as a predictor of divergent creativity. The full model predicted 64.4% of the variance in divergent creative quality.

For convergent creativity, model 1, containing app store data only, explained 5.0% of the variance, which was not significant ($F(4,147) = 1.93, p = .109$), and only the coefficient for review score was significant, having a small positive effect. In the 2nd model, app characteristics of the apps were added: this explained a further 28.2% of the variance, and the change in R squared was significant ($F(8,139) = 7.32, p < .001$). In this model, the effect of review score was no longer significant. There was however a significant positive effect of target age, and negative effect of the cognitive skills domain. In the 3rd and final model, divergent creativity quality and the three general quality sub-scores were included as predictors and this added a further 28.9% of the variance, which was again significant ($F(4,135) = 25.77, p < .001$). After controlling for these predictors, the effect of target age and cognitive skills remained, but the negative effect of open-endedness became significant, such that more open-ended activities was associated with lower convergent creativity score. General meaning and general structure scores were also positive predictors of convergent creativity quality. The full model predicted 57.6% of the variance in convergent creative quality.

4. Discussion

This methodical review of the creativity app market for primary school aged children aimed to examine the app characteristics of existing creativity apps for children, their quality, and the relation between their quality and app store data and app characteristics. A large sample of 152 apps targeted at children whose descriptions contained creativity-related keywords were reviewed using evidence-based criteria to address three research questions.

4.1. RQ1: are creativity apps more likely to contain certain characteristics than others?

In our sample of creativity apps for children, certain app characteristics were indeed more common than others. With regards to target age, more apps were primarily suitable for younger (4–7 years) than older (8–12 years) children in our age range. This is

Table 6

Correlations between variables included in the regression models, including creative and general app quality, app store data, and app characteristics,. *p < .05 1 = Creativity: divergent.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2 Creativity: convergent	.327*	–														
3 General: meaning	.480*	.675*	–													
4 General: usability	.154	.086	.219*	–												
5 General: structure	–.227*	.322*	.086	.152	–											
6 Review score	.118	.168*	.082	.059	.137	–										
7 Number of installs	–.023	.074	–.076	–.313*	.069	–.052	–									
8 Paid	–.056	–.114	–.035	.449*	.117	.003	–.343*	–								
9 Expert approval	.038	–.118	.113	.410*	–.008	–.030	–.234*	.259*	–							
10 Target age	.150	.433*	.338*	–.027	.234*	.089	.083	–.094	–.317*	–						
11 Domain: Visual arts	.271*	.028	.044	–.212*	–.294*	.030	.160*	–.344*	–.079	.049	–					
12 Domain: Language	–.154	.063	.206*	–.033	–.042	–.216*	–.211*	–.104	–.120	.008	–.182*	–				
13 Domain: STEM	–.180*	.233*	.175*	.131	.437*	.144	.019	.013	.040	.216*	–.243*	–.105	–			
14 Domain: PSE	–.004	–.181*	–.203*	.050	–.036	–.080	.044	.213*	.121	–.323*	–.493*	–.153	–.163*	–		
15 Domain: Cognitive	–.432*	–.297*	–.331*	–.063	.230*	–.059	.092	.119	–.073	–.005	–.099	.014	.002	–.171*	–	
16 Out of app engage.	–.008	.224*	.335*	–.081	–.180*	.078	–.094	–.034	–.049	.330*	.096	.184*	–.042	–.183*	–.142	–
17 Open-endedness	.731*	.130	.394*	.260*	–.423*	.047	–.152	.017	.169*	.035	.228*	–.071	–.203*	.024	–.572*	.060

Table 7
Results of hierarchical multiple regression models predicting creative quality.

		Creativity divergent			Creativity convergent		
		Beta	t	p	Beta	t	p
1	(Constant)		0.057	.955		−0.434	.665
	Review score	0.119	1.450	.149	0.168	2.079	.039
	Number of installs	−0.032	−0.365	.715	0.036	0.411	.682
	Paid	−0.081	−0.919	.359	−0.081	−0.923	.357
2	Expert approval	0.055	0.647	.518	−0.084	−0.991	.324
	(Constant)		−0.806	.421		−1.086	.279
	Review score	0.079	1.374	.172	0.119	1.629	.106
	Number of installs	0.023	0.366	.715	0.122	1.500	.136
	Paid	−0.017	−0.262	.794	0.012	0.146	.884
	Expert approval	−0.016	−0.264	.792	0.018	0.229	.819
	Target age	0.175	2.624	.010	0.347	4.063	<.001
	Domain: Visual arts	0.103	1.265	.208	−0.044	−0.422	.673
	Domain: Language	−0.050	−0.755	.451	0.099	1.167	.245
	Domain: STEM	−0.070	−1.063	.290	0.119	1.409	.161
	Domain: PSE	0.052	0.647	.519	−0.105	−1.029	.305
	Domain: Cognitive skills	−0.042	−0.584	.560	−0.329	−3.545	<.001
	Out of app engagement	−0.111	−1.825	.070	0.039	0.504	.615
	Open-endedness	0.668	9.177	<.001	−0.019	−0.204	.839
3	(Constant)		−0.715	.476		−0.915	.362
	Review score	0.052	0.983	.327	0.053	0.927	.355
	Number of installs	−0.02	−0.334	.739	0.060	0.962	.338
	Paid	−0.018	−0.288	.774	−0.039	−0.580	.563
	Expert approval	−0.065	−1.063	.290	−0.119	−1.822	.071
	Target age	0.071	1.087	.279	0.143	2.087	.039
	Domain: Visual arts	0.110	1.47	.144	−0.025	−0.304	.762
	Domain: Language	−0.136	−2.165	.032	−0.055	−0.803	.423
	Domain: STEM	−0.134	−2.141	.034	−0.021	−0.307	.760
	Domain: PSE	0.088	1.203	.231	−0.068	−0.854	.395
	Domain: Cognitive skills	0.030	0.438	.662	−0.261	−3.637	<.001
	Out of app engagement	−0.176	−2.911	.004	−0.004	−0.058	.954
	Open-endedness	0.562	6.965	<.001	−0.252	−2.594	.011
	General: meaning	0.298	3.533	<.001	0.588	7.361	<.001
	General: usability	−0.017	−0.255	.799	0.043	0.616	.539
	General: structure	−0.021	−0.288	.774	0.193	2.847	.005
	Creativity: divergent				0.136	1.48	.141
	Creativity: convergent	0.114	1.415	.159			

supported by previous research, which suggested that there is a disproportionate amount of app content for preschool children (Sari, Takacs, & Bus, 2019; Vaala & Levine, 2015). This finding suggests that there is a scarcity of apps targeting creativity skills that are developmentally more suitable for older children (e.g., coding). Although evidence from a small sample of studies shows that digital games that target such skills can also have a positive impact on the cognitive development of younger children (Blumberg et al., 2019, p. pp1), developmental maturity has been found to play an important role in children's learning of these skills (Bati, 2022). However, the attempts to produce digital content or improve the quality of existing apps that aim to promote these skills for older children remain insufficient (Blumberg et al., 2019, p. pp1). This study, therefore, contributes to the literature by highlighting the need in the creativity app market for apps that are optimised for the developmental needs of children in both early and middle childhood, especially given that the developmental optimisation of these apps plays a key role in child engagement (Bitu et al., 2022; Piotrowski & Meester, 2018).

In terms of domains of learning, most creativity apps contained at least one activity within the visual arts. This could be due to the link between art and creativity in both lay perception and education (Department for Education, 2021; NASUWT, 2009) and the range of activities that the visual arts cover (drawing, crafts, multimedia, fashion design, etc.). Apps with activities relating to personal and social education were the second most frequent: the majority of these were related to personal development, including cooking and role play. Perhaps surprisingly, cognitive skills were the third most common domain covered, with perceptual and motor skills being frequently represented. Anecdotally, these apps seemed to be those that contained a range of different activities and claimed to support “problem-solving” or did not make it clear which activities were linked to creativity in their descriptions. Other domains (language, STEM, performing arts, and humanities) were less common in this sample of apps. Literature has shown that activities in digital games that support learning in domains related to STEM fields (e.g., programming and coding) can support child creativity as well as cognitive development (for a review of studies see, Bati, 2022; Papadakis, 2022). Apps that support learning in these areas can be used as accessible resources for cultivating foundational skills for children from backgrounds disproportionately underrepresented in STEM fields (e.g., Bower et al., 2022). The findings from this study, however, show that the scope of creativity apps remains narrow and that most apps are produced to support artistic skills.

For app engagement, the vast majority involved primarily in-app engagement rather than mixed or out of app engagement. This is

perhaps not surprising, given that on-screen interactions allow more data to be collected about users, more evidence for engagement, and more time for advertising. The small number of apps with off-screen engagement involved either craft instructions, writing prompts, or taking photos or videos to edit. It has been suggested that learning benefits can arise from apps linking to real-life activities and environments (Hirsh-Pasek et al., 2015), including activating prior knowledge, more active participation in learning, and greater transfer of knowledge.

In terms of social interaction, many more apps were intended primarily or only for solo use rather than mixed or primarily social use. The lack of opportunities for social interaction has been noted by other authors (Kolák et al., 2021; Meyer et al., 2021; Vaala & Levine, 2015), and this study confirms that this is also an issue with creativity apps. This design feature may reflect conceptions that parents want to use screen time as “babysitting” (Beyens & Eggermont, 2014), but in fact, parents report often co-using devices with their children (Marsh et al., 2015; Ofcom, 2021; Pearce et al., 2022). The small number of apps that did allow for social interaction were either games to be played with another player outside of the app or included an online multiplayer mode. The research evidence on the effects of individual or interactive play on children is divided. Although some studies show that app features supporting interactivity between children and adults may contribute to child learning (e.g., Sheehan, Pila, Lauricella, & Wartella, 2019), other studies have also found that children who played more independently demonstrated better learning outcomes after playing with digital creativity games (e.g., Lennon, Pila, Flynn, & Wartella, 2022). Hence, this finding can only suggest that most creativity apps are designed to promote individual play without denoting further implications on the benefits of such game characteristics for children.

Finally, just over half of apps involved primarily or only open-ended activities, which was more than mixed or primarily closed-ended activities. This is counter to previous research with general educational apps for children that suggested a predominance of closed-ended “drill” or quiz activities (Highfield & Goodwin, 2013; Vaala & Levine, 2015), suggesting that apps for creativity are designed to be more open-ended. It should be noted, however, that over a quarter of apps are still primarily or only closed-ended. Although open-ended games have been shown to promote curiosity and support conceptual learning (Kang & Liu, 2022; Moore, 2017), the amount of guidance provided in these games has been shown to play a significant role in the experiences of children (Flynn, Richert, & Wartella, 2019), especially for those who are identified as academically at-risk (Kang & Liu, 2022). In this regard, the findings from the current study suggest that the popularity of open-ended activities in creativity apps can offer a medium for cultivating curiosity and conceptual learning in children; however, the benefits of these apps may vary depending on the children’s characteristics.

4.2. RQ2: are creativity apps higher quality in some areas than others?

Creative quality was overall low according to our criteria. This extends previous research suggesting that educational apps tend to be low-quality for supporting learning in literacy or generally (Callaghan & Reich, 2018; Kolák et al., 2021; Meyer et al., 2021) by showing that current creativity apps are not strongly evidence-based either. Lines of evidence have been consistent across studies investigating the quality of apps, and the lack of change in game production has been linked to the discrepancy between the expectations of researchers and game developers (Colliver, Hatzigianni, & Davies, 2020). Researchers, policymakers, and organisations are guided by pedagogical frameworks in their assessments of apps. This agenda also resonates with the expectations of parents when choosing apps for their children (Broekman, Piotrowski, Beentjes, & Valkenburg, 2016). Game developers, however, are often guided by their motivation to increase their revenues, even though this means the game production may not respond to the needs or expectations of the users (Colliver et al., 2020). This study contributes to the literature by showing that the quality of creativity apps does not differ from other educational apps produced for children.

The data also confirmed that creativity apps for children performed better on some quality features than on others. Supports for experimentation and domain-specific skills were most common: thus, existing creativity apps tend to contain some opportunities for children to develop relevant domain-specific skills and knowledge (such as motor skills or colour names for drawing or realistic scenes for role playing) and some variety in materials or ways to use them. This is important, as previous research demonstrates that the ability to be creative in a domain is linked to skills in the specific domain (Baer, 2015) and that children need opportunities to experiment in order to be creative (Robson, 2014). It should be noted that scores for these two variables were still around the midpoint of the scale, and the criteria were generous, indicating basic provision for these criteria. Supports for convergent thinking and modelling were particularly rare, meaning that very few apps allowed or scaffolded children to solve creative problems that had a correct solution or provided examples of creative products or processes. This is problematic, as convergent thinking is an important aspect of creativity to nurture (Cropley, 2006), and the provision of many and diverse examples has been suggested to boost children’s creativity (Hoicka et al., 2018).

4.3. RQ3: do app store data and app characteristics of creativity apps predict app quality?

The results suggest that some app store data and app characteristics can explain some variance in creative app quality, although a significant amount of variance remains to be accounted for. Some predictors were similar across convergent and divergent quality scores. The key app store data (review score, installs, payment, and expert approval) explained a negligible amount of the variance in both of these scores. Some previous studies also found no evidence for an effect of review scores, installs, or cost on the quality of educational apps (Dubé et al., 2020; Kolák et al., 2021; Papadakis et al., 2018). This study adds to this evidence by showing that even with a large sample of apps, these factors are not strong predictors of creative quality specifically, especially when these factors are considered alongside each other.

However, other app store data did relate to both convergent and divergent creative quality scores. Apps that contained more

creativity-related keywords in their app descriptions and were listed by common sense media scored more higher, whereas those sourced from a keyword search scored lower. This finding is consistent with other studies (Taylor, Kolak, Bent, & Monaghan, 2022; Vaala & Levine, 2015) and suggests that independent, expert reviews are a positive indicator of creative app quality, whereas user or expert reviews in the app store itself may in fact be misleading. Relevant keywords in app descriptions, however, provide some indication of creative quality.

Other factors that correlated with both convergent and divergent creative quality scores included the total number of app activities, cognitive skills domain, target age, and general meaning score. Apps with more different activities scored lower, perhaps due to the fact that only some activities were intended to support creativity and/or the difficulty in maintaining quality when creating a large amount of app content. Those in the cognitive skills domain also scored lower, likely because these activities tended not to involve creative skills despite having creativity related keywords (e.g., problem-solving) in their app descriptions. Furthermore, target age and general meaning scores emerged as significant positive predictors in regression analyses after controlling for other factors, suggesting that apps for older children and those with more meaningful content scored higher for both aspects of creativity.

However, some factors were related to divergent and convergent thinking differentially. First, open-endedness emerged as a positive predictor of divergent creativity and a negative predictor of convergent thinking. Apps higher in convergent creative quality were also more likely to be higher in general learning structure (better scaffolding and feedback). The two facets of creativity also showed different associations with domains of learning, with divergent creativity scores lower for STEM and language activities. Overall, these findings may be explained by the fact that divergent thinking tasks require a diversity of responses, whereas convergent thinking tasks require a correct answer. This may make it easier to create open-ended divergent thinking activities and closed-ended convergent thinking activities with clear scaffolding and feedback, especially in the STEM and language domains. However, this suggests that attention should be given to how to provide useful scaffolding and feedback in divergent creativity activities and successfully integrate divergent thinking into STEM and language activities as both directions in creativity have been shown to support distinct aspects of learning and respond differently to the diverse needs of children (Flynn et al., 2019; Kang & Liu, 2022).

4.4. Implications, limitations and future research

The results presented here have significant implications for parents and practitioners when selecting creative apps for children. They imply that caregivers should source creative apps from independent media review sites rather than from app stores themselves and look for relevant keywords in the app description. Having said that, they should also adopt a skeptical approach to app store information, avoiding reliance on expert approval, number of installs, review score, and claims in app descriptions, and instead checking app content to seek apps with fewer activities that allow children to do open-ended creation.

The findings also have implications for app developers in terms of designing higher quality apps and marketing them accurately. Firstly, it is recommended that app developers carefully consider the accuracy of app descriptions, using only relevant keywords, and if there are multiple activities within an app, being specific about which activities are related to which area of development. Secondly, it is suggested that apps focus attention on a smaller number of activities and areas of learning to increase quality. Thirdly, there is a gap in the market for certain kinds of creativity apps, particularly those for older children, in domains other than visual arts and socio-emotional and involving real-world interactions. Finally, there is a critical need for more evidence-based design of creativity apps for children, particularly in terms of modelling and convergent thinking, but there is significant room for improvement in all aspects of design.

One of the limitations of this study is that the quality criteria are only as accurate as the current literature on supports for children's creativity. Because there have been relatively few experimental studies with children, especially in a digital learning context, the criteria chosen may not fully reflect the features needed to enable children's creativity. Further, some of the criteria that might be important (such as intrinsic motivation) but are subjective could not be reliably rated. Whilst a much larger sample of apps was included here than in previous app content analysis studies, the sample is still necessarily limited, and this prevents us from including all potentially relevant variables in regression models. Furthermore, while this study is one of the very few quantitative studies of creativity apps for children, it is correlational and, so a causal relation between predictors and app quality cannot be inferred. Lastly, this study only included apps that were available in English. The implications from this study, therefore, should be drawn with caution. Although less than 1% of the apps from the initial sample were excluded due to being only available in a language other than English, suggesting that most apps were produced in English, this study still needs to be replicated for app markets that target different language groups to draw cross-culturally valid implications. Future research needs to both expand on the evidence base for designing creativity apps for children and test the impact of these apps on children's creativity in controlled experiments. More experimental research is required that manipulates specific features of the learning environment, particularly in an app context, to examine how this affects children's expression of creativity. Furthermore, longer-term interventions with evidence-based apps to see whether they affect the development of children's creativity over time would also allow the causal impacts of apps to be assessed. Lastly, further research needs to be conducted to determine if the issues with quality in creativity apps also apply to apps produced in other languages. It is possible that the mass-production of apps in English may be leading to a lower proportion of high-quality apps in the app market. In markets with smaller scales of production, this may not be the case. With more evidence on these issues and knowledge exchange between researchers and game developers, research can contribute to the development of widely accessible apps with high quality features for children across the world.

4.5. Conclusion

In conclusion, children's apps with a creativity goal are more commonly targeted at younger children and tend to cover visual arts activities, be for solo use, be open-ended, and have primarily in app interactions. These apps are overall low quality, but particularly so in regards to supporting convergent thinking and modelling creativity. Most information from the app store, including review score, installs, payment and expert approval, does not provide a consistent indication of app quality, although depending on where the app was sourced, the number of app activities and relevant keywords were correlated with quality. Apps for older children are higher quality, but the effect of learning domain and open-endedness of activities differs between supports for convergent and divergent creativity. More work is therefore needed first to investigate the quality of apps available in other languages and second to improve the quality of creativity apps for children and the relevance of the information in the app store for parents.

Credit author statement

Sophie A. Booton: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Funding acquisition, Project administration. Pinar Kolancali Methodology, Formal analysis, Investigation, Data curation, Writing – review & editing. Victoria A. Murphy: Supervision, Funding acquisition, Writing – review & editing.

Data availability

Provided via OSF.

Acknowledgements

This work was funded by Ferrero International, creators of Appplaydu. The authors would also like to thank Birsu Kandemirci, Elena Hoicka for their feedback on the methodology, Daniela Singh for her help with developing the criteria, and members of the LiFT team for their support with preparation and writing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compedu.2023.104811>.

Appendix B. Scoring rubrics for creative quality rating with added examples of apps included in the final app sample

Domain-specific skills.

Criteria		
0- Unsatisfactory App provides no, very basic or mostly inaccurate content for developing domain-specific skills or knowledge relevant to creativity in that domain, such that the child is unlikely to learn much from it. Examples from app sample In Funny Food DRESS UP games for toddlers and kids!, users add body parts, clothes and accessories to food characters which provides no opportunity to learn any relevant skills or knowledge. In Shapes and Colours – Kids games for toddlers, users dragged objects to match colours or sizes but this does not allow them to learn anything about colour which could be applied to creativity.	1- Limited App provides mostly accurate content for developing domain-specific skills or knowledge relevant to creativity in that domain but this is limited (e.g., low quality, limited in scope, implicit, unrealistic, and/or not <u>transferable</u>) In Toca Kitchen Sushi, children use realistic ingredients and tools to make sushi, but there is no explicit teaching and the skill would not be easy to transfer to making sushi in real life. In Bogga Vacation - App For Kids, content supports young children's understanding of going on holiday (e.g., packing, going to the airport), but the content is limited in scope as the same events are replayed every time.	2- Good App provides good, explicit, and accurate content for developing domain-specific skills or knowledge relevant to creativity in that domain. In MarcoPolo Ocean, children are explicitly taught facts about sea creatures. In Grasshopper: Learn to Code, explicit instructions and feedback are given to support learning to code.

Divergent thinking.

Criteria		
0- Unsatisfactory App does not require divergent thinking (I.e., only simple choices, no requirement for multiple ideas/solutions/problems to find).	1- Limited App allows for divergent thinking (I.e., allowing for multiple ideas/solutions/problems to find) but does not explicitly support or encourage this.	2- Good App explicitly prompts for divergent thinking (I.e., encourages multiple ideas/solutions/problems to find to the same activity) and/or supports divergent thinking through effective divergent thinking strategies (e.g., embedding strategies in game

(continued on next page)

(continued)

Criteria		
design, prompting to use strategies, tools e.g., mindmaps)		
Examples from app sample		
In Sago Mini Space, children move a character to activate animations but there is no requirement for them to generate ideas.	In Hovercraft - Build Fly Retry, users can build hovercrafts using individual cubes in whatever way they wish.	In Rory's Story Cubes, children are explicitly instructed to interpret the cubes in different ways.
In OK Play: Create your story!, children make simple choices between options and templates to generate a story but they are not able to provide their own ideas.	In Sago Mini Fun Fair, children can explore a fun fair scene by moving characters and objects to role play or tell stories.	In Bad Piggies, children can earn badges by solving levels using different strategies and tools.

Convergent thinking.

Criteria		
0- Unsatisfactory	1- Limited	2- Good
App does not require or allow for convergent thinking (e.g., simple stimulus-response actions, habitual responding, simple visual matching, activities with no correct answers) or has no instructions.	App provides instructions/tutorials for convergent thinking tasks but no further supports for convergent thinking <u>or</u> further supports require an in-app purchase.	App provides one or more supports for convergent thinking which <u>does not require an in-app purchase</u> (e.g. hints, instructions to view a problem from different perspectives, scaffolded evaluation of ideas, useful tools for the problem at hand e.g., a dictionary, inventory, map)
Examples from app sample		
In Toca Life: Farm, children can play freely with the characters and props but there are no instructions to prompt convergent problem-solving.	In Toontastic 3D, video instructions are given for each part of a story plot but no further supports are provided.	In Terraria, targets are given towards building a society and free supports are given to help with this, such as an inventory and map.
In the Mekorama puzzle level builder, children can create their own levels, but there are no instructions and guidance to do so.	In My Town: Discovery, there are character needs which the child can provide for but no further supports.	In Moose Math by Duck Moose, children can tap a character to receive a maths hint adapted to their attempts to solve the problem so far.

Experimentation.

Criteria		
0- Unsatisfactory	1- Limited	2- Good
App contains no features that allow user experimentation or risk-taking (or on average 0 per activity/example).	App contains either: one feature per activity/example that allows user experimentation or risk-taking. Or Both erase/undo/replace AND restart ONLY (no varied modes/materials).	App contains two or more features per activity/example that allow user experimentation or risk-taking which <u>has to include at least one of</u> : varied materials (in terms of diversity 2+ dimensions – e.g. colour, size, style, material, modality etc. - and quantity (20+ total options); open-ended e.g., photo/audio upload counts as 20+) <u>and/or</u> varied modes of interacting with the materials to produce different products (e.g., making a story, or picture, or video; open-ended play)
Examples from app sample		
In Kids Kitchen, children tap and drag objects on screen to make dishes (e.g. a milkshake) but the actions and order are fixed and the same each time.	In Arduino Science Journal, children can edit the images and text of their science experiment record, start a new one, and edit old ones, but varied materials or ways of interacting are not provided.	In Stop Motion Studio, children can edit their own multimedia, including importing photos, videos, and audio and adding drawings to create varied products.
In The Room: Old Sins, the puzzles to be solved by finding and using objects have to be completed in a pre-determined way.	In Kids Music Instruments Sounds, the game can be replayed and different music can be made, but the materials (instruments and notes) available were too limited.	In Coloring Cars McQueen, children can select from many colours and pen sizes to colour their templates and can replay to colour in different ways.

Inspiration.

Criteria		
0- Unsatisfactory	1- Limited	2- Good
App includes no supports for inspiration (or on average 0 per activity/example).	App includes <u>one</u> support for inspiration for children's creation per activity/example (from: pre-made elements such as stickers, templates, phrases/words to choose from and/or props that can be moved, placed, rotated and/or sensory observation and/or randomised elements.)	App includes <u>two or more</u> supports for inspiration per activity/example.

(continued on next page)

(continued)

Criteria		
Examples from app sample		
In FlipaClip: Cartoon Animation, children can draw, upload and edit files to make an animation but no pre-made elements (e.g., backgrounds or props) are provided. In Toca Train, children ride a train around a scene and there are no supports for sensory observation or props to play with.	In Dollicon: Cute Doll Avatar Maker, a randomise button generates a random combination of look elements which can then be edited to inspire children's choices, but there are no other supports for inspiration. In Toca Life: Vacation, there are objects and characters which serve as pre-made elements to inspire the child's open-ended role play.	In Rory's Story Cubes, dice are rolled to randomise the images used to inspire a story, and can be moved and rotated to help with structuring the narrative. In BBC CBeebies Get Creative - Build, paint and play!, each game has a number of pre-made elements, such as stickers, stencils, or blocks and a random wand tool which adds a random element to the child's picture.

Modelling.

Criteria		
0- Unsatisfactory The app does not contain any examples or modelling of creative processes (steps leading to an outcome or workings-out) or products (final outcomes) (or on average 0 per activity).	1- Limited The app contains modelling of creative processes OR products which are free to access, with at least one example per activity on average. <u>Or</u> The app contains only one example each of modelling of creative processes AND product per activity.	2- Good The app contains modelling of creative processes AND products which are free to access, with two or more examples of each per activity on average.
Examples from app sample		
In Magic Slate - Colour & Draw free, there are no examples of drawings made or videos or instructions for drawing processes to see how effects could be made. In Pettson's Inventions there is no modelling of problem-solving strategies.	In Kiloblocks, several premade objects (e.g. trees, caravan) are present which provide a model for end products to create. In Stop Motion Studio, only one example of a premade animation for which the process can be seen is included.	In ScratchJr, there are several premade projects with editable code for children to remix. In Origami for kids: easy paper schemes there are numerous step-by-step instructions for completing origami models.

Meta-cognition.

Criteria		
0- Unsatisfactory App contains no features that enable reflection on creation (or on average 0 per activity/example).	1- Limited App contains one type of feature that enables reflection on creation per activity/example (from: record of full creative product can be saved AND ability to restart activity, and/or record of creative process, and/or reflection questions/prompts)	2- Good App contains two or more types of features that enable reflection on creation per activity/example.
Examples from app sample		
In Art of Glow, children draw shapes and animations but there is no photo button or gallery to save creations to. In Sago Mini Neighborhood Blocks, the game is saved but cannot be restarted to allow children to make changes.	In Avokiddo Emotions, children can take photos of their play with animals and props which are saved in the device gallery, but there are no further supports for meta-cognition. In Miga Town, the game is saved so the child can look back at it later.	In ChatterPix Kids by Duck Moose, creations can be viewed in the gallery and edited to review the process. In Hoopa City, children can save their creations, and have a book of formulas for elements they previously generated as a reflection prompt.

Self-Beliefs.

Criteria		
0- Unsatisfactory App does not contain any supports for child's creative self-beliefs (or on average 0 per activity/example).	1- Limited App contains one support for child's creative self-efficacy per activity/example (from: praise or indicators for effort rather than ability or performance, presence of <u>child</u> character(s), directly identifying the child as a creator, displaying/using the child's creation elsewhere in the app.)	2- Good App contains two or more supports for child's creative self-efficacy per activity/example.
Examples from app sample		
In Blackboard App - Simply Best, pictures can be saved to the device but are not displayed in a gallery in the game, and there is no praise or child characters to interact with. In Thinkrolls 2 - Logic Puzzles, performance but not effort is rewarded and there is no display of creations.	In Dinosaur Bus - Create a Car! Games for kids, children's decorated vehicle is used in a mini-game at the end. In Rainbow Ice Cream Sandwich Maker Ice Cream Shop, children are identified as chefs as praise at the end of activities (e.g., "Very creative work chef!").	In Toontastic 3D, children's cartoons are displayed in a gallery and child characters are included that can be animated.

Appendix C. T-test comparisons between creativity criterion scores

Pair	MD	SD	t	df	p
Experimentation – Domain-specific skills	0.283	0.909	3.84	151	<.001*
Domain-specific skills – Divergent thinking	0.257	0.636	4.98	151	<.001*
Divergent thinking - Metacognition	0.026	0.563	0.58	151	.565
Divergent thinking - Inspiration	0.066	0.497	1.63	151	.105
Divergent thinking – Creative self-efficacy	0.105	0.589	2.21	151	.029
Divergent thinking – Convergent thinking	0.309	0.730	5.22	151	<.001*
Metacognition - Convergent thinking	0.283	0.825	4.23	151	<.001*
Inspiration - Convergent thinking	0.243	0.797	3.76	151	<.001*
Creative self-efficacy - Convergent thinking	0.204	0.817	3.08	151	.002
Creative self-efficacy - Modelling	0.250	0.730	4.22	151	<.001*
Convergent thinking - Modelling	0.051	0.694	0.82	151	.414

* significant at Bonferroni corrected alpha level of $p \leq .001$.

Appendix D. Correlations between raw scores for creative quality criteria

	C1	C2	C3	C4	C5	C6	C7
C2 Divergent thinking	.275**						
C3 Convergent thinking	.367**	.128					
C4 Experimentation	.246**	.745**	.017				
C5 Inspiration	.208*	.588**	.031	.677**			
C6 Modelling	.294**	.227**	.313**	.265**	.189*		
C7 Metacognition	.156	.573**	.085	.729**	.479**	.402**	
C8 Creative self-efficacy	-.021	.356**	-.071	.410**	.269**	.155	.509**

* $p < .05$ ** $p < .001$ C1 = Domain-specific skill.

Appendix E. Correlations between additional app store data variables and creative app quality

	1	2	3	4	5	6	7	8	9	10
2 Creativity: convergent	.327*	–								
3 Age of app	.018	.220*	–							
4 Update recency	.025	.001	.233*	–						
5 Download size	.125	.052	-.167*	-.169*	–					
6 Keyword tokens	.332*	.164*	.070	.012	.178*	–				
7 Keyword types	.285*	.151	.168*	.076	.153	.815*	–			
8 No. App activities	-.511*	-.318*	-.048	-.198*	.003	-.245*	-.212*	–		
9 Source: list	-.018	-.066	-.282*	.061	-.025	-.128	-.013	-.041	–	
10 Source: keyword	-.315*	-.258*	-.048	-.211*	-.228*	-.182*	-.212*	.388*	-.433*	–
11 Source: CSM	.391*	.348*	.294*	.145	.282*	.314*	.353*	-.390*	-.351*	-.643*

* $p < .05$.

References

- franzke, aline, shakti, B., & Anja, Z. (2020). *Internet Research: Ethical Guidelines 3.0*. Charles and the Association of Internet Researchers. <https://aoir.org/reports/ethics3.pdf>.
- Alfonso-Benlliure, V., & Santos, M. R. (2016). Creativity development trajectories in Elementary Education: Differences in divergent and evaluative skills. *Thinking Skills and Creativity*, 19, 160–174. <https://doi.org/10.1016/j.tsc.2015.11.003>
- Ali, S., Park, H. W., & Breazeal, C. (2021). A social robot's influence on children's figural creativity during gameplay. *International Journal of Child-Computer Interaction*, 28, Article 100234. <https://doi.org/10.1016/j.ijcci.2020.100234>
- Amabile, T. M., & Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior*, 36, 157–183.
- Ananiadou, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries. In *OECD education working papers*.
- Baer, J. (1996). The effects of task-specific divergent-thinking training. *The Journal of Creative Behavior*, 30(3), 183–187. <https://doi.org/10.1002/j.2162-6057.1996.tb00767.x>.
- Baer, J. (2015). The importance of domain-specific expertise in creativity. *Roeper Review*, 37(3), 165–178. <https://doi.org/10.1080/02783193.2015.1047480>
- Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science*, 23(10), 1117–1122. <https://doi.org/10.1177/0956797612446024>
- Baker, E. B. A. (2017). Apps, iPads, and literacy: Examining the feasibility of speech recognition in a first-grade classroom. *Reading Research Quarterly*, 52(3), 291–310. <https://doi.org/10.1002/rrq.170>
- Bati, K. (2022). A systematic literature review regarding computational thinking and programming in early childhood education. *Education and Information Technologies*, 27(2), 2059–2082. <https://doi.org/10.1007/s10639-021-10700-2>
- Beyens, I., & Eggermont, S. (2014). Putting young children in front of the television: Antecedents and outcomes of parents' use of television as a babysitter. *Communication Quarterly*, 62(1), 57–74. <https://doi.org/10.1080/01463373.2013.860904>

- Bitu, F., Galinon-Mélénec, B., & Molina, M. (2022). Modifying sensory afferences on tablet changes originality in drawings. *Frontiers in Psychology*, 13. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.806093>.
- Blumberg, F. C., Deater-Deckard, K., Calvert, S. L., Flynn, R. M., Green, C. S., Arnold, D., et al. (2019). Digital games as a context for children's cognitive development: Research recommendations and policy considerations. *Social Policy Report*, 32(1), pp1–33. <https://doi.org/10.1002/sop2.3>
- Bower, C. A., Zimmermann, L., Verdine, B. N., Pritulsky, C., Golinkoff, R. M., & Hirsh-Pasek, K. (2022). Enhancing spatial skills of preschoolers from under-resourced backgrounds: A comparison of digital app vs. concrete materials. *Developmental Science*, 25(1), e13148–n. <https://doi.org/10.1111/desc.13148>. /a.
- Broekman, F. L., Piotrowski, J. T., Beentjes, H. W. J., & Valkenburg, P. M. (2016). A parental perspective on apps for young children. *Computers in Human Behavior*, 63, 142–151. <https://doi.org/10.1016/j.chb.2016.05.017>
- 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. <https://doi.org/10.1080/17439884.2018.1498355>
- Colliver, Y., Hatzigianni, M., & Davies, B. (2020). Why can't I find quality apps for my child? A model to understand all stakeholders' perspectives on quality learning through digital play. *Early Child Development and Care*, 190(16), 2612–2626. <https://doi.org/10.1080/03004430.2019.1596901>
- Common Sense Media. (2017). *The Common Sense census: Media use by kids age zero to eight*. https://doi.org/10.1163/9789004250901_013
- Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal*, 18(3), 391–404. https://doi.org/10.1207/s15326934crj1803_13
- Department for Education. (2019). *Educational Criteria for early years apps*.
- Department for Education. (2021). *Statutory framework for the early years foundation stage*.
- Dorin, A. (2013). Chance and complexity: Stochastic and generative processes in art and creativity. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/2466816.2466837>
- Doron, E. (2016). Short term intervention model for enhancing divergent thinking among school aged children. *Creativity Research Journal*, 28(3), 372–378. <https://doi.org/10.1080/10400419.2016.1195616>
- Doron, E. (2017). Fostering creativity in school aged children through perspective taking and visual media based short term intervention program. *Thinking Skills and Creativity*, 23, 150–160. <https://doi.org/10.1016/j.tsc.2016.12.003>
- Duarte-García, M. A., & Sigal-Seřchovich, J. R. (2019). Working with Electroacoustic Music in Rural Communities: The use of an interactive music system in the creative process in primary and secondary school education. *Organised Sound*, 24(3), 228–239. <https://doi.org/10.1017/S135577181900030X>
- Dubé, A. K., Kacmaz, G., Wen, R., Alam, S. S., & Xu, C. (2020). Identifying quality educational apps: Lessons from 'top' mathematics apps in the Apple App store. *Education and Information Technologies*, 25(6), 5389–5404. <https://doi.org/10.1007/s10639-020-10234-z>
- Dunn, J., & Sweeney, T. (2018). Writing and iPads in the early years: Perspectives from within the classroom. *British Journal of Educational Technology*, 49(5), 859–869. <https://doi.org/10.1111/bjet.12621>
- Erstad, O. (2012). The learning lives of digital youth-beyond the formal and informal. *Oxford Review of Education*, 38(1), 25–43. <https://doi.org/10.1080/03054985.2011.577940>
- Eynon, R., Fry, J., & Schroeder, R. (2017). The ethics of Internet research. In *Handbook of online research methods* (pp. 47–72). Sage.
- Fessakis, G., Gouli, E., & Mavroudi, E. (2013). Problem solving by 5–6 years old kindergarten children in a computer programming environment: A case study. *Computers in Education*, 63, 87–97. <https://doi.org/10.1016/j.compedu.2012.11.016>
- Fessakis, G., Lappas, D., & Mavroudi, E. (2014). Could computer games-based problem solving positively affect the development of creativity in young children? A mixed method case study. In L. K. Heider, & M. R. Jalongo (Eds.), *Young children and families in the information age: Applications of technology in early childhood* (Vol. 10, pp. 207–226). Springer. https://doi.org/10.1007/978-94-017-9184-7_12
- Flynn, R. M., Richert, R. A., & Wartella, E. (2019). Play in a digital world: How interactive digital games shape the lives of children. *American Journal of Play*, 12(1), 54–73.
- Glăveanu, V. P. (2013). Rewriting the language of creativity: The five A's framework. *Review of General Psychology*, 17(1), 69–81. <https://doi.org/10.1037/a0029528>
- Guegan, J., Brechet, C., & Nelson, J. (2020). Dreamlike and playful virtual environments to inspire children's divergent thinking. *Journal of Media Psychology*. <https://doi.org/10.1027/1864-1105/a000279>
- Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.
- Highfield, K., & Goodwin, K. (2013). Apps for mathematics learning: A review of 'educational' apps from the iTunes app store. *The 36th Annual Conference of Mathematics Education Research Group of Australasia*, 2009, 378–385.
- Hillman, M., & Marshall, J. (2009). Evaluation of digital media for emergent literacy. *Computers in the Schools*, 26(4), 256–270. <https://doi.org/10.1080/07380560903360186>
- 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 in the Public Interest*, 16(1), 3–34. <https://doi.org/10.1177/1529100615569721>
- Ho, T. K., Lin, H., Chen, C., & Lee, L. (2017). The development and evaluation of a tablet painting application for enhancing the artistic expression skills of students through reflection. *Computers in Education*, 115, 56–68. <https://doi.org/10.1016/j.compedu.2017.07.012>
- Hoffmann, J. D., & Russ, S. W. (2016). Fostering pretend play skills and creativity in elementary school girls: A group play intervention. *Psychology of Aesthetics, Creativity, and the Arts*, 10(1), 114–125. <https://doi.org/10.1037/aca0000039>
- Hoicka, E., Powell, S., Knight, J., & Norwood, M. (2018). Two-year-olds can socially learn to think divergently. *British Journal of Developmental Psychology*, 36(1), 22–36. <https://doi.org/10.1111/bjdp.12199>
- Hong, E., & Milgram, R. M. (2010). Creative thinking ability: Domain generality and specificity. *Creativity Research Journal*, 22(3), 272–287. <https://doi.org/10.1080/10400419.2010.503535>
- International Data Corporation (Idc). (2022). *Smartphone market share, 2022* <https://www.idc.com/promo/smartphone-market-share>.
- Kabali, H. K., Irigoyen, M. M., Nunez-Davis, R., Budacki, J. G., Mohanty, S. H., Leister, K. P., et al. (2015). Exposure and use of mobile media devices by young children. *Pediatrics*, 136(6), 1044–1050. <https://doi.org/10.1542/peds.2015-2151>
- van de Kamp, M. T., Admiraal, W., van Drie, J., & Rijlaarsdam, G. (2015). Enhancing divergent thinking in visual arts education: Effects of explicit instruction of meta-cognition. *British Journal of Educational Psychology*, 85(1), 47–58. <https://doi.org/10.1111/bjep.12061>
- Kang, J., & Liu, M. (2022). Investigating navigational behavior patterns of students across at-risk categories within an open-ended serious game. *Technology, Knowledge and Learning*, 27(1), 183–205. <https://doi.org/10.1007/s10758-020-09462-6>
- Karwowski, M., Lebuda, I., & Beghetto, R. A. (2019). Creative Self-Beliefs. In *The Cambridge Handbook of Creativity* (pp. 396–418).
- Kaufman, J. C., & Glăveanu, V. P. (2019). A review of creativity theories. In J. C. Kaufman, & R. J. Sternberg (Eds.), *The cambridge handbook of creativity* (pp. 27–43). Cambridge University Press. <https://doi.org/10.1017/9781316979839.004>
- Khatena, J. (1971). Teaching disadvantaged preschool children to think creatively with pictures. *Journal of Educational Psychology*, 62(5), 384–386. <https://doi.org/10.1037/h0031634>
- Kirsh, D. (2014). The importance of chance and interactivity in creativity. *Pragmatics and Cognition*, 22(1), 5–26. <https://doi.org/10.1075/pc.22.1.01kir>
- Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6), 1534–1555. <https://doi.org/10.1037/0278-7393.25.6.1534>
- Kolak, J., Norgate, S., Monaghan, P., & Taylor, G. (2021). 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. <https://doi.org/10.1080/17482798.2020.1844776>
- Kucirkova, N., & Sakr, M. (2015). Child-father creative text-making at home with crayons, iPad collage & PC. *Thinking Skills and Creativity*, 17, 59–73. <https://doi.org/10.1016/j.tsc.2015.05.003>
- Lennon, M., Pila, S., Flynn, R., & Wartella, E. A. (2022). Young children's social and independent behavior during play with a coding app: Digital game features matter in a 1:1 child to tablet setting. *Computers & Education*, 190, Article 104608. <https://doi.org/10.1016/j.compedu.2022.104608>
- Leong, T., Howard, S., & Vetere, F. (2008). Take a chance on me: Using randomness for the design of digital devices. *Interactions*, 15(3), 16–19. <https://doi.org/10.1145/1353782.1353787>

- Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J., Lahmar, J., & Scott, F. (2018). Play and creativity in young children's use of apps. *British Journal of Educational Technology*, 49(5), 870–882. <https://doi.org/10.1111/bjet.12622>
- Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J. C., Lahmar, J., Scott, F., et al. (2015). *Exploring play and creativity in pre-schoolers' use of apps: Final project report*. Medeiros, K. E., Partlow, P. J., & Mumford, M. D. (2014). Not too much, not too little: The influence of constraints on creative problem solving. *Psychology of Aesthetics, Creativity, and the Arts*, 8(2), 198–210. <https://doi.org/10.1037/a0036210>
- Meyer, M., Zosh, J. M., McLaren, C., Robb, M., McCaffery, H., Golinkoff, R. M., et al. (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*, 1. <https://doi.org/10.1080/17482798.2021.1882516>. –23.
- Moore, H. C. (2017). Look what I made!: Open-ended apps that spark creativity. *YC Young Children*, 72(5), 21–27.
- Nasuwt. (2009). Creativity and the arts in the primary school. *April Discussion Document and Proceedings of the Consultative Conference on Education*, 2009, 1–142.
- Nielson, K. A., & Powless, M. (2007). Positive and negative sources of emotional arousal enhance long-term word-list retention when induced as long as 30 min after learning. *Neurobiology of Learning and Memory*, 88(1), 40–47. <https://doi.org/10.1016/j.nlm.2007.03.005>
- Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence*, 39(1), 36–45. <https://doi.org/10.1016/j.intell.2010.11.002>.
- Ofcom. (2021). *Children and parents: Media use and attitudes report 2020/21*.
- Osborne, J. W. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research and Evaluation*, 15(12), 1–9.
- Pachler, N., Bachmair, B., & Cook, J. (2010). *Mobile learning: Structures, agency, practices*. Springer. <https://doi.org/10.1080/10494820.2015.1113705>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 1–11. <https://doi.org/10.1186/s13643-021-01626-4>
- Papadakis, S. (2022). Apps to promote computational thinking and coding skills to young age children: A pedagogical challenge for the 21st century learners. *Education Process: International Journal*, 11(1). <https://doi.org/10.22521/edupij.2022.11.1>
- 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. <https://doi.org/10.1007/s10639-017-9579-0>
- Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2018). Educational apps from the android Google play for Greek preschoolers: A systematic review. *Computers in Education*, 116, 139–160. <https://doi.org/10.1016/j.compedu.2017.09.007>
- Papert, S. (1980). Computers for children. In *Children, computers, and powerful ideas* (2nd ed., pp. 3–18).
- Papert, S. (1996). A word for learning. In Y. Kafai, & M. Resnick (Eds.), *Constructionism in practice: Designing, thinking, and learning in a digital world* (pp. 9–24).
- Pearce, K. E., Yip, J. C., Lee, J. H., Martinez, J. J., Windleharth, T. W., Bhattacharya, A., et al. (2022). Families playing animal crossing together: Coping with video games during the COVID-19 pandemic. *Games and Culture*, 17(5), 773–794. <https://doi.org/10.1177/15554120211056125>
- Piotrowski, J. T., & Meester, L. (2018). Can apps support creativity in middle childhood? *Computers in Human Behavior*, 85, 23–33. <https://doi.org/10.1016/j.chb.2018.03.030>
- Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39(2), 83–96. https://doi.org/10.1207/s1532698Sep3902_1
- Puckett Cliatt, Shaw, J. M., Sherwood, J. M., Puckett Cliatt, ... Sherwood, J. M. (1980). Effects of Training on the Divergent-Thinking Abilities of Kindergarten Children. *Child Development*, 51(4), 1061–1064.
- Richardson, C., & Mishra, P. (2018). Learning environments that support student creativity: Developing the SCALE. *Thinking Skills and Creativity*, 27(February 2017), 45–54. <https://doi.org/10.1016/j.tsc.2017.11.004>
- Robson, S. (2014). The Analysing Children's Creative Thinking framework: Development of an observation-led approach to identifying and analysing young children's creative thinking. *British Educational Research Journal*, 40(1), 121–134. <https://doi.org/10.1002/berj.3033>
- Sakr, M., Connelly, V., & Wild, M. (2016). 'Evil cats' and 'jelly floods': Young children's collective constructions of digital art making in the early years classroom. *Journal of Research in Childhood Education*, 30(1), 128–141. <https://doi.org/10.1080/02568543.2015.1107156>
- Sari, B., Takacs, Z. K., & Bus, A. G. (2019). What are we downloading for our children? Best-selling children's apps in four European countries. *Journal of Early Childhood Literacy*, 19(4), 515–532. <https://doi.org/10.1177/1468798417744057>
- Segundo Marcos, R. I., López Fernández, V., Daza González, M. T., & Phillips-Silver, J. (2020). Promoting children's creative thinking through reading and writing in a cooperative learning classroom. *Thinking Skills and Creativity*, 36, Article 100663. <https://doi.org/10.1016/j.tsc.2020.100663>
- Sheehan, K. J., Pila, S., Lauricella, A. R., & Wartella, E. A. (2019). Parent-child interaction and children's learning from a coding application. *Computers in Education*, 140, Article 103601. <https://doi.org/10.1016/j.compedu.2019.103601>
- Shuler, C. (2009). *iLearn: A content analysis of the iTunes app store's education section*. November 2009. The Joan Ganz Cooney Center at Sesame Workshop.
- Sowden, P. T., Clements, L., Redlich, C., & Lewis, C. (2015). Improvisation facilitates divergent thinking and creativity: Realizing a benefit of primary school arts education. *Psychology of Aesthetics, Creativity, and the Arts*, 9(2), 128–138. <https://doi.org/10.1037/aca0000018>.
- Statista. (2022a). *Apple: Most popular app store categories 2022*. Statista. <https://www.statista.com/statistics/270291/popular-categories-in-the-app-store/>.
- Statista. (2022b). *Biggest app stores in the world 2022*. Statista. <https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/>.
- Statista. (2022c). *Google Play most popular app categories 2022*. Statista. <https://www.statista.com/statistics/279286/google-play-android-app-categories/>.
- Sun, M., Wang, M., & Wegerif, R. (2019). Using computer-based cognitive mapping to improve students' divergent thinking for creativity development. *British Journal of Educational Technology*, 50(5), 2217–2233. <https://doi.org/10.1111/bjet.12825>
- Taylor, G., Kolak, J., Bent, E. M., & Monaghan, P. (2022). Selecting educational apps for preschool children: How useful are website app rating systems? *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13199>. February, 1–21.
- Taylor, G., Kolak, J., Norgate, S. H., & Monaghan, P. (2022). Assessing the educational potential and language content of touchscreen apps for preschool children. *Computers in Education*, 3, Article 100102.
- Vaala, S., & Levine, M. H. (2015). *Getting a read on the app stores*.
- Vatavu, R. D., Cramariuc, G., & Chipor, D. M. (2015). Touch interaction for children aged 3 to 6 years: Experimental findings and relationship to motor skills. *International Journal of Human-Computer Studies*, 74(1), 54–76. <https://doi.org/10.1016/j.ijhcs.2014.10.007>
- Ward, T. B., & Kolomyts, Y. (2019). Creative cognition. In J. C. Kaufman, & R. J. Sternberg (Eds.), *The cambridge handbook of creativity* (2nd ed., pp. 175–199). Cambridge University Press.
- Warren, F., Mason-Apps, E., Hoskins, S., Azmi, Z., & Boyce, J. The role of implicit theories, age, and gender in the creative performance of children and adults. *Thinking Skills and Creativity*, 28(December 2017). <https://doi.org/10.1016/j.tsc.2018.03.010>.
- Weatherford, D. R., Esparza, L. V., Tedder, L. J., & Smith, O. K. H. (2020). Using a Fork as a Hairbrush: Investigating Dual Routes to Release from Functional Fixedness. *Journal of Creative Behavior*, 0, 1–14. <https://doi.org/10.1002/jocb.442>.
- Webb, A. N., & Rule, A. C. (2013). Effects of teacher lesson introduction on second graders' creativity in a science/literacy integrated unit on health and nutrition. *Early Childhood Education Journal*, 351–360. <https://doi.org/10.1007/s10643-013-0615-4>
- Wen, M.-C., Butler, L. T., & Koutstaal, W. (2013). Improving insight and non-insight problem solving with brief interventions. *British Journal of Psychology (London, England: 1953)*, 104(1), 97–118. <https://doi.org/10.1111/j.2044-8295.2012.02107.x>
- Winner, E., Hetland, L., Veenema, S., Sheridan, K., & Palmer, P. (2020). *Studio thinking: How visual arts teaching can promote disciplined habits of mind*. *New directions in aesthetics, creativity, and the arts*. April 2015 (pp. 189–206). <https://doi.org/10.4324/9781315224084-15>.
- Xiong, Z., Liu, Q., & Huang, X. (2022). The influence of digital educational games on preschool Children's creative thinking. *Computers & Education*, 189, Article 104578. <https://doi.org/10.1016/j.compedu.2022.104578>