ORIGINAL ARTICLE

British Journal of Educational Technology

BERA

Selecting educational apps for preschool children: How useful are website app rating systems?

Gemma Taylor[1] | **Joanna Kolak[1]** | **Eve M. Bent[1]** |
Padraic Monaghan[2]

[1]Department of Psychology, School of Health and Society, University of Salford, Salford, UK

[2]Department of Psychology, Lancaster University, Lancaster, UK

Correspondence
Gemma Taylor, Department of Psychology, School of Health and Society, University of Salford, Salford M5 4WT, UK.
Email: g.taylor4@salford.ac.uk

Funding information
Economic and Social Research Council, Grant/Award Number: ES/R004129/1

Abstract

In the present paper, we assess whether website rating systems are useful for selecting educational apps for preschool age children. We selected the 10 highest scoring and 10 lowest scoring apps for 2–4-year-olds from two widely used websites (Good App Guide; Common Sense Media). Apps rated highly by the two websites had a higher educational potential as assessed by a validated questionnaire for evaluating the educational potential of apps and were more likely to include a learning goal and feedback compared to low scoring apps. However, high scoring apps scored on average just 9/20 for indicators of educational potential, and both high and low scoring apps had poor language quality as determined by psycholinguistic and construction type analyses. We argue that website rating systems should also include quality of feedback, adjustable content, social interactions, storyline and a more fine-grained analysis of language in their assessments.

KEYWORDS
app features, children, educational technology, language, touchscreen apps

Practitioner notes

What is already known about this topic
- Appropriately designed apps for preschool age children have the potential to teach early school readiness skills.
- Selecting high quality educational apps for preschool age children is challenging.
- The children's app marketplace is currently unregulated.

What this paper adds
- We assess whether two leading app rating websites are useful for selecting educational apps for preschool age children.
- Children's apps rated highly by two app website rating systems had a higher educational potential than low rated apps as measured by a research informed app evaluation tool.
- In-depth analysis of the language in apps shows that highly rated children's apps on app rating websites may not enrich a child's early language environment.

Implications for practice and/or policy
- Children's app rating website assessments should include potential for feedback, language, adjustable content, storyline and social interactions.
- Policy should be implemented for app ratings in the app stores or on website app rating systems.

INTRODUCTION

Children's touchscreen apps are increasingly used in preschools (eg, Neumann & Neumann, 2017), and research recognises the potential of touchscreen apps for preschool age children as a valuable educational tool (Griffith et al., 2020). Children aged 0–5 in the UK use a variety of apps, some of which can support play and creativity. However, some of the apps most widely used by this group are not aimed at their age range (Marsh et al., 2018). Behnamnia et al. (2020) found that digital game-based learning can help children aged 3–6 years develop creative skills and critical thinking, and can facilitate knowledge transfer. In the UK, the Department for Education (2020) recommend six apps for literacy and language learning for preschool age children to support the development of early school readiness skills. However, not all touchscreen apps are created equal. The app marketplace is currently unregulated and as a result, app developers can choose to market their app as educational rather than needing to meet key criteria. We argue that policy needs to be implemented for educational app ratings, specifically for children's apps, similar to ratings for the age of computer games (eg, Pan European Game Information (PEGI)), movies and websites to be accessed via mobile devices (eg, British Board of Film Classification (BBFC)). At present, caregivers could select an educational app for their children based on a number of issues including whether the app has an educational aim (eg, to teach children colours), marketing materials which label apps as educational or websites which offer independent app ratings. Understanding how to select good quality educational apps is essential to ensure that the time children spend using touchscreens is constructive.

Crucially, research by Dardanou et al. (2020) shows that parents of children aged 0–3 from different cultures (Norway, Portugal, Japan) express the need for further guidance on touchscreen technology use at home. Furthermore, another cross-cultural study (England, Norway, Greece) investigating educators' approach and experiences around the use of technology with children aged 0–3 years has shown that educators are not confident when

integrating touchscreen technology for very young children (Fotakopoulou et al., 2020). Educators from that study also indicated that they were not satisfied with their training regarding the use of technology. Those two studies show that parents and educators alike need guidance in selecting and using technology with their children, including help in understanding what constitutes good quality touchscreen media for young children.

First and foremost, however, understanding what constitutes an educational app is an ongoing challenge and debate in the literature with many app rating systems proposed to assess the educational features of touchscreen apps for young children (see Kolak et al., 2021 Table 1 for review). However, these app rating systems are subject to limitations including long lists of criteria and jargon, making them inappropriate for use by caregivers and early years' practitioners (Kolak et al., 2021; Papadakis & Kalogiannakis, 2020).

Two tools to assess educational potential have recently been proposed, the four-pillar framework (see Hirsh-Pasek et al., 2015; Meyer et al., 2021) and the questionnaire for evaluating the educational potential of apps for preschool age children (Kolak et al., 2021). The four-pillar framework requires apps to have active learning, encourage engagement in the learning process, meaningful learning and social interaction (Hirsh-Pasek et al., 2015), apps can score 0–3 points on each of the four pillars with an overall app score of 4 or less indicating an app with low educational potential (Meyer et al., 2021). A more comprehensive measure of the educational potential of apps for preschool age children was proposed by Kolak et al. (2021). The questionnaire for evaluating the educational potential of apps for preschool age children is a brief user-friendly tool for general audience (caregivers, educators, app developers), which can be used alongside coding criteria for quantifying app features, a more thorough tool for researchers (Kolak et al., 2021). The questionnaire assesses 10 key educational features of touchscreen apps: learning goal, meaningful learning, solving problems, feedback, social interactions, opportunities for exploration, plotline, quality of language, adjustable content and app design (see Kolak et al., 2021 for justification of these features). These two tools have been used to assess whether apps with a learning goal or apps marketed as educational are indicators of an educational app for preschool age children.

Research suggests that children's touchscreen apps have room for improvement when it comes to educational potential. Taylor et al. (submitted) found that apps with a learning goal from the top 10 lists in the Amazon, Google and Apple app stores have a higher educational potential and provide more feedback than apps with no learning goal. However, apps with a learning goal rarely promoted exploratory use, adjusted content to a child's performance or made use of parasocial interactions with onscreen characters, which can all promote children's engagement and learning. In addition, analyses of apps labelled as educational for preschoolers found that math and literacy apps did not scaffold children's learning by

TABLE 1 Definition of construction type and examples from the apps in the present study

Construction type	Definition	Apps example
Fragments	Utterances without subject and predicate	*Almost*
Copulas	Utterances in which the main verb is some form of *to be*	*That's a big yawn, sleepyhead!*
Imperatives	Subjectless requests for child's action	*Catch another fish*
Questions	Utterances with question syntax in the main clause	*Can you help them?*
Subject–Predicate	Utterances with both a subject and single lexical predicate	*Look the first flower has opened!*
Complex	Utterances with two lexical verbs	*When I'm scared I close my eyes*

adjusting the level of difficulty to their performance and feedback could be improved by providing information to correct learning (Papadakis et al., 2018). Using the four-pillar framework, Meyer et al. (2021) found that the majority of apps labelled as educational via app marketing materials scored poorly for promoting active, engaging and meaningful learning or promoting social interactions, with approximately 58% of apps considered to be low quality in terms of educational potential. Thus, apps with a learning goal or marketed as educational do not guarantee that they will be educational for young children. Other methods for caregivers and early years' educators for selecting educational apps for preschool age children are website app rating systems, which provide expert ratings for children's apps based on specific criteria (eg, the Good App Guide assesses fun, ease of use and skills developed).

To date, there is no research assessing whether apps labelled as educational via generally used website rating systems are able to promote children's learning. Common Sense Media (US) provides ratings for movies, television programmes, books, apps and games. Their website describes using a panel of unnamed experts in child development and children's media literacy to create their rating systems. Common Sense Media rates apps using a 5-point system based on educational value, positive models and messages, violence, scariness, sexy stuff, language, consumerism and commercialism, and risky and unhealthy behaviour for apps targeting 2–4-year-olds. A comparable website in the UK, the Good App Guide rates toys, baby products and apps submitted by developers for a fee. The Good App Guide asks their panel of experts, who are not named, as well as a selection of children, to assess each app. Ratings are given by both the children and experts to create a total score for each app. The Good App Guide rates apps using 3 5-point scales for fun, ease of use and skills' development. Given that these website app rating systems are not transparent, there are important implications for parents. In the present paper, we investigate whether website rating systems assess apps based on principles of children's cognitive development and learning in the digital age to ensure that apps can effectively promote children's learning using the questionnaire for evaluating the educational potential of apps for preschool age children and coding criteria to quantify app features (Kolak et al., 2021).

Until recently, app rating systems have largely ignored the potential for apps to enrich a child's early language environment by providing an additional source of language input (see Kolak et al., 2021) and this important feature appears to be currently missing from website rating systems. Common Sense Media limits language assessments to whether an app contains swearing or not and the Good App Guide does not mention language at all. In terms of providing an additional and enriching source of language input for young children, compared to child directed speech (CDS), apps with a learning goal contain lower frequency words similar to books, and repeat those words more often than CDS, which potentially makes them an enriched form of input for young children (Kolak et al., submitted). Research suggests that longer length utterances can support children's vocabulary development (Rowe, 2012), children tend to acquire higher frequency words earlier than low frequency words (eg, Swingley & Humphrey, 2018), concrete words are learnt earlier than abstract words (eg, Gleitman et al., 2005) and words with an earlier age of acquisition (age at which the word was learnt) may be more accessible to young children (eg, Monaghan & Ellis, 2010). Furthermore, with respect to utterance types, research suggests that language input sources that contain a lower proportion of fragments, copulas and imperatives, and a higher proportion of questions, subject–predicate and complex utterances (see Table 1, for definitions and examples) provide an enriched source of language input for preschool age children (see Cameron-Faulkner & Noble, 2013, for a similar argument about language in books and CDS). In the present paper, we therefore assess whether apps rated highly by website rating systems also contain good quality and age-appropriate language.

The aim of the present study was to determine whether apps labelled as educational by website rating systems have educational potential when assessed against principles of

children's cognitive development and learning in the digital age. 39 high scoring and low scoring apps labelled as educational for 2–5-year-old children by two websites: the Good App Guide (2020) and Common Sense Media (2020) were coded using a systematic coding scheme to assess educational potential (see Kolak et al., 2021). The Good App Guide and Common Sense Media were chosen rather than other available website rating systems such as Parents Choice Awards, which only includes apps with their seal of approval, and Children's Technology Review, which requires a subscription, because they are both free to parents and include a range of apps using their rating systems. In addition, we also conducted language analyses to assess language quality (similar to Kolak et al., submitted). App content was analysed and the following research questions were addressed in the analysis:

1. Can website rating systems accurately distinguish between apps with high educational potential and those with lower educational potential?
2. Does the app content recommended by the website rating systems have the potential to support children's early skill development?

METHOD

App sample

Apps were identified from two websites with media ratings for parents: Common Sense Media and the Good App Guide on 08/01/2020. The 10 highest scoring apps and 10 lowest scoring apps on each website (just 9 apps were identified as low scoring for the Good App Guide) were selected using the following searches.

1. For the Good App Guide, we selected "Educational Apps" for 2-, 3- and 4-year-olds. The apps that were identified were then sorted by "Skills rating" from highest to lowest.
2. For Common Sense Media, we went onto "App reviews" and selected "Education" and "Educational Games" for 2-, 3- and 4-year-olds. The apps that were identified were then sorted by the "Stars" rating from highest to lowest.

Importantly, Common Sense Media split gaming apps into multiple categories; we therefore included "Educational Games" for our search on Common Sense Media to ensure we included all educational apps. In contrast, the Good App Guide did not split gaming apps into categories.

No apps were found to overlap between the lists of apps from our searches on Common Sense Media and the Good App Guide. However, there were some apps (Good App Guide, $N = 3$ low scoring apps; Common Sense Media, $N = 6$ low scoring apps; $N = 7$ high scoring apps) that were no longer available on the app stores (Apple and Google Play). We explored whether these apps could be accessed in an alternative way; once we confirmed that the app had been taken off the app marketplace, we moved to the next app on the list. For details of the apps in our sample, see Table 2.

The apps selected within the Good App Guide top 10 apps, consisted of a selection of maths-focused ($N = 1$; eg, "EduGuru Maths"), literacy-focused ($N = 7$; eg, "Mario's Alphabet" and "Writing Wizzard"), and general play ($N = 2$; eg, "Get Well Soon Hospital with Dr Ranj" and "Dr Panda's House"). The Good App guide bottom 10 consisted of no maths or literacy-focused apps, but did include general play apps ($N = 10$; eg, "Tee and Mo bathtime" and "Create a car"). Similarly, the top 10 apps selected within Common Sense Media also consisted of a selection of maths-focused ($N = 4$; eg, "My First puzzles" and "Intro to maths by

TABLE 2 Details of the high and low scoring apps from Common Sense Media and the Good App Guide websites

	Number of apps			
	Common Sense Media		Good App Guide	
	High scoring	Low scoring	High scoring	Low scoring
5-point scale rating	10 (5*)	10 (2*)	10 (5/5)	7 (2/5), 2 (3/5)
Marketed for Age 2+	0	1	0	5
Marketed for Age 3+	5	5	8	1
Marketed for Age 4+	5	4	2	3
Android (Google play store)	1	2	3	2
Apple (Apple Store)	9	8	7	7
Free apps (/10)	2	6	3	3
Price range of paid apps	£1.99–£4.99	£1.99–£3.99	£1.99–£19.99	£1.99–£3.99

Montessori"), literacy-focused ($N = 4$; eg, "ABC Go" and "Intro to letters by Montessori") and general play ($N = 2$; eg, "First school" and "The monster at the end...") apps. Within the bottom 10 apps rated by Common Sense Media, there were math-focused apps ($N = 3$; eg, "Vegetable Maths Masters" and "Fiete Puzzles"), literacy-based apps ($N = 1$; eg, Juno lamb learn letters") and general play apps ($N = 6$; eg, "Octonauts" and "BubbleBud Kids Universe").

Data collection

Each app was downloaded and used for 5 min by the third author. The apps were used in a systematic way by using all available features on the screen just once and completing activities in the order suggested by the app, ensuring that the main features of the app were explored during the five-minute period (see Kolak et al., 2021 for justification). The in-device screen recorder was used to capture the 5-min app use.

Coding

Apps' educational potential was assessed using the questionnaire for evaluating the educational potential of apps, and apps' features were coded using criteria for quantifying app features developed by Kolak et al. (2021). Additionally, an in-depth language analysis was performed on the language present during app use.

For the questionnaire for evaluating the educational potential of apps, the first and second authors scored the apps using the 10 questionnaire items: learning goal, meaningful learning, solving problems, feedback, social interactions, opportunities for exploration, plotline, quality of language, adjustable content, and app design. Each item can score 0–2 points giving an educational potential index of between 0–20 for each app. The 5-min app screen recordings were assessed individually by the first and second author using the tool. The discrepancies were discussed and resolved between the coders. Inter-rater reliability was high ($\kappa = 0.911$, $p < 0.001$).

For the coding criteria for quantifying app features, app recordings were coded in ELAN v5.9 using the template by Kolak et al. (2021). Coding commenced at the first point each app had fully loaded and in-app play was available to commence. Each screen shown during app use was coded for 11 categories using the coding criteria for quantifying app features to assess the frequency of: touch gestures, activity type, activity goal, screen elements, background complexity, background sound, other app interactions, presence of feedback, feedback delivery method, content of the feedback and object property. 15% of apps ($N = 6$) were double coded by the second author of the manuscript. The inter-rater reliability was high ($\kappa = 0.880$, $p < 0.001$).

To measure language potential for learning in our app sample (ie, its age-appropriateness and variety and depth of language usage), we focussed on five psycholinguistic measures:

1. The proportion of different utterance types (fragments, imperatives, questions, copulas, subject-predicate, complex). All the utterances from the apps were transcribed by the third author of the manuscript, and each utterance was coded for its grammatical construction type by the second author according to the taxonomy used in Cameron-Faulkner and Noble (2013), which is based on standard linguistic criteria. For the construction types and their definitions, as well as examples of these constructions, see Table 1.
2. Mean length of utterance (MLU), which measures the number of words the utterance contains and is an indicative measure to characterise quality of language input (Dickinson & Porche, 2011).
3. Word frequency, which was taken from a corpus of child-appropriate speech from television programmes aimed at children aged up to 6 years. This corpus was derived from transcripts of 5,848,083 words from a UK public broadcast television channel Cbeebies (van Heuven et al., 2014).
4. Age of acquisition (AoA), the ratings were obtained from Kuperman et al. (2012).
5. Concreteness ratings of words were obtained from Brysbaert et al. (2014) on a scale from 1 (abstract word) to 5 (concrete word).

Frequency, AoA and concreteness were assigned separately for each word in each utterance.

RESULTS

We analysed whether high rated apps differed from low rated apps in their potential to support children's early skill development, and whether this was true for both rating systems. In this section, we first present analyses regarding the educational potential of apps, which were based on the questionnaire for evaluating the educational potential of apps (Kolak et al., 2021). We then proceed to the analyses of app features, which were based on the coding criteria for quantifying app features (Kolak et al., 2021). Finally, we present the analyses of language in the apps, focussing on the five psycholinguistic measures outlined in the Coding section above.

Educational potential

In order to test whether there was a correspondence between a high rating and educational potential and whether that was true for both rating systems, we ran a 2 (Website) $\times$ 2 (Rating) ANOVA with educational potential as a dependent variable. We found no main effect of

website ($F < 1$) and no interaction between website and rating ($F < 1$). The main effect of rating was significant, $F(1,35) = 13.151$, $p = 0.001$, $\eta_p^2 = 0.273$. High-rated apps across the two websites had higher educational potential ($M = 9.20$, $SD = 3.07$) than low-rated apps ($M = 5.53$, $SD = 3.36$). See Figure 1, for a breakdown of the app scores for high and low rated apps on each website for each questionnaire item.

We also performed chi-square analysis to investigate whether the two websites differed in the number of high- versus low-rated apps that contained a learning goal (learning goal was one of the items in the questionnaire for evaluating the educational potential of apps). We found that among the apps that contained a learning goal, there were more high-rated apps in general (18/20) than low-rated apps (9/19) $\chi^2(2) = 38.028$, $p < 0.001$.

App features

We analysed whether high rated apps differed from low rated apps in 11 app features from the coding criteria for quantifying app features (Kolak et al., 2021), and whether this was true for both rating systems. We present the summary of the results in Table 3 and descriptive statistics for each app feature in Table 4.

In summary, we found a main effect of touch gesture with tapping required by apps more often than dragging, tracing and swiping, and dragging more often than tracing (see Tables 3 and 4). In addition, there was a main effect of object property with static
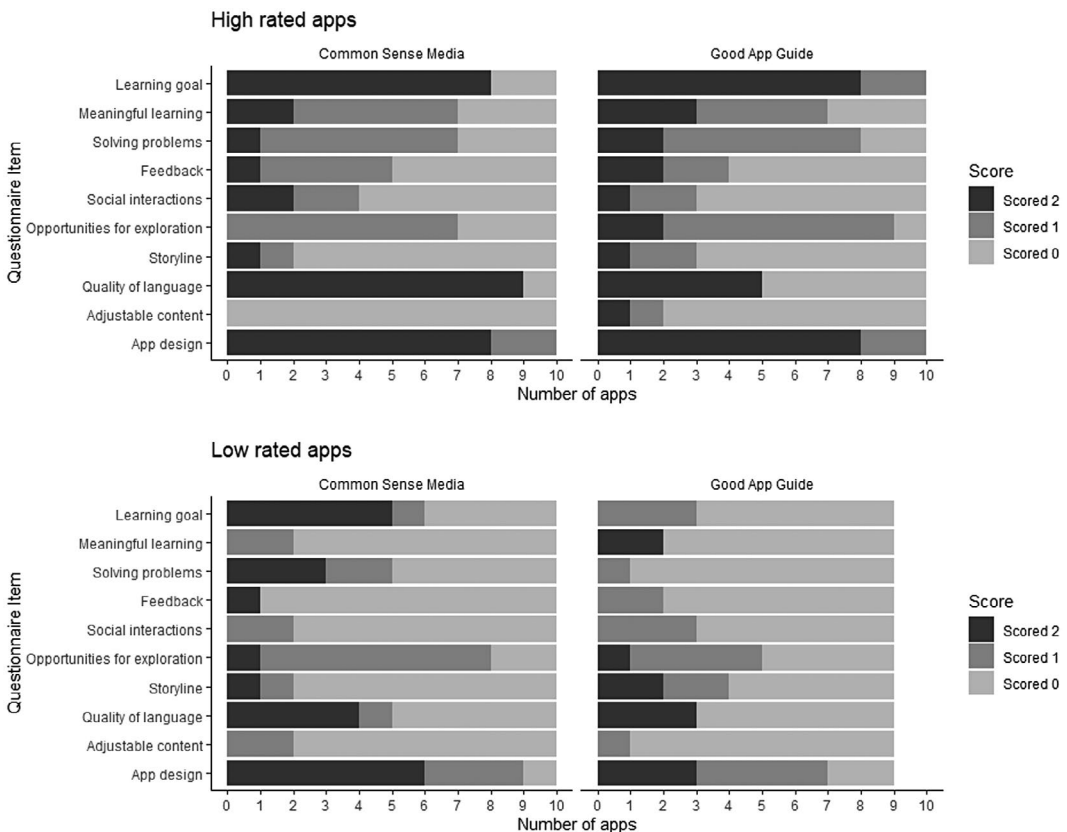

FIGURE 1 The number of high-rated and low-rated apps from the two websites that scored 2, 1 or 0 points on each item of the questionnaire for evaluating the educational potential of apps

TABLE 3 Quantifying app features: summary of the analyses (n.s. means non-significant)

Area of features	Compared app feature	Aim of analysis	Type of analysis	Main results/effects	Interactions
Touch gestures	Touch gestures	Comparing the frequency of the four touch gestures: • tapping • swiping • dragging • tracing	2 (Website) × 2 (Rating) × 4 (Touch gesture) ANOVA	Main effect of website n.s. Main effect of rating n.s. Main effect of Touch gesture $F(3,35) = 22.325$, $p < 0.001$, $\eta_p^2 = 0.659$ tapping > dragging $p = 0.001$ tapping > swiping $p = 0.001$ tapping > tracing $p < 0.001$ draging > tracing $p = 0.013$	Website × Touch gesture n.s. Rating × Touch gesture n.s. Website × Rating n.s. Touch × Website × Rating n.s.
Active learning	Activity type	Comparing the frequency of cognitive activities and stimulus-reaction activities	2 (Website) × 2 (Rating) × 2 (Activity type) ANOVA	Main effect of website n.s. Main effect of rating n.s. Main effect of Activity type n.s.	Website × Activity type n.s. Rating × Activity type n.s. Website × Rating n.s. Website × Rating × Activity type, n.s.
	Activity goal	Comparing the number of unique activity goals	2 (Website) × 2 (Rating) ANOVA	Main effect of website n.s. Main effect of rating n.s.	Website × Rating n.s.

(Continues)

TABLE 3 (Continued)

Area of features	Compared app feature	Aim of analysis	Type of analysis	Main results/effects	Interactions
Complexity of the learning environment	Screen elements	Comparing the mean number of screen elements	2 (Website) × 2 (Rating) ANOVA	Main effect of website n.s. Main effect of rating n.s.	Website × Rating n.s.
	Background complexity	Comparing the proportion of complex background	2 (Website) × 2 (Rating) ANOVA	Main effect of website n.s. Main effect of rating n.s.	Website × rating n.s.
	Background sound	Comparing the frequency of different sound types: • no background sound • simple background sound • music • complex background sound	2 (Website) × 2 (Rating) × 4 (Background sound) ANOVA	Main effect of website n.s. Main effect of rating n.s. Main effect of background sound n.s.	Website × background sound n.s. Rating × Background sound $F(3,33) = 3.808$, $p = 0.013$, $\eta_p^2 = 0.098$ Website × rating n.s. Website × rating × background sound n.s.
	Other app interactions	Comparing the mean number of other interactions available on a screen	2 (Website) × 2 (Rating) ANOVA	Main effect of website n.s. Main effect of rating n.s.	Website × rating n.s.
Feedback	Opportunities for feedback	Comparing the number of apps that include opportunities for feedback	Chi-square	No difference between the websites Significant difference between the high versus low-rated apps: $\chi^2(2) = 7.248$, $p < 0.018$ High-rated apps ($N = 17$), low-rated apps ($N = 10$)	
	Presence of feedback	Comparing the proportion of feedback	2 (Website) × 2 (Rating) ANOVA	Main effect of Website n.s. Main effect of Rating n.s.	Website × Rating n.s.

TABLE 3 (Continued)

Area of features	Compared app feature	Aim of analysis	Type of analysis	Main results/effects	Interactions
	Feedback delivery method	Comparing the frequency of feedback delivered: • via audio, • onscreen • simultaneously via audio & onscreen	2 (Website) × 2 (Rating) × 3 (Feedback delivery method) ANOVA	Main effect of website n.s. Main effect of rating n.s. Main effect of Feedback delivery method $F(1,23) = 3.538$, $p = 0.037, \eta_p^2 = 0.133$ Audio & onscreen > onscreen, $p = 0.003$	Website × Feedback delivery method n.s. Rating × Feedback delivery method n.s. Website × Rating n.s. Website × Rating × Feedback delivery method n.s.
	Content of the feedback	Comparing the proportion of ostensive feedback	Website (2) × Rating (2) ANOVA	Main effect of website n.s. Main effect of rating n.s.	Website × rating n.s.
App design sophistication	Object property	Comparing the frequency of different object property: • static • static movement • animation • mixed	2 (Website) × 2 (Rating) × 4 (Object property) ANOVA	Main effect of website n.s. Main effect of rating n.s. Main effect of Object property $F(3,35) = 7.159$, $p = 0.001, \eta^2 = 0.388$, static < static movement, $p < 0.001$; static movement > mixed, $p = 0.033$	Website × Object property n.s. Rating × Object property $F(3,33) = 3.727$, $p < 0.001, \eta_p^2 = 0.242$ Website × Rating n.s. Website × Rating × Object property n.s.

movement occurring more often than static and mixed (see Tables 3 and 4). For feedback, high-rated apps were significantly more likely to contain opportunities for feedback (see Table 3). There was also a main effect of feedback delivery method, with apps using more audio plus onscreen feedback compared to onscreen only feedback (see Tables 3 and 4).

TABLE 4 Descriptive statistics for each app feature for high- and low-rated apps, and for the whole sample

App feature	Measure	Low-rated apps (*N* = 19)	High-rated apps (*N* = 20)	All apps (*N* = 39)
		Mean (*SD*)		
Touch gestures				
Touch gestures	Freq of tapping	28.68 (19.08)	22.20 (18.82)	28.68 (19.08)
	Freq of swiping	10.32 (19.93)	4.90 (12,33)	7.54 (16.48)
	Freq of dragging	9.58 (15.64)	8.70 (10.18)	9.13 (12.96)
	Freq of tracing	0.89 (2.92)	2.55 (5.93)	1.74 (4.72)
Active learning				
Activity type	Freq of cognitive activities	17.26 (17.97)	22.90 (12.32)	20.15 (15.40)
	Freq of stimulus-reaction activities	14.84 (21.26)	10.30 (20.77)	12.51 (21.03)
Activity goal	Number of different goals	10.80 (8.23)	11.00 (6.63)	10.90 (7.38)
Complexity of the learning environment				
Screen elements	Mean number on the screen	7.40 (9.16)	6.89 (2.14)	7.14 (6.57)
Background complexity	Proportion of complex background to simple background	0.62 (0.36)	0.39 (0.39)	0.50 (0.39)
Background sound	Freq of no sound	2.15 (7.09)	11.70 (17.51)	6.93 (14.04)
	Freq of simple sound	6.10 (10.79)	13.20 (17.33)	9.65 (14.70)
	Freq of music	8.10 (10.75)	9.35 (16.06)	8.73 (13.50)
	Freq of complex sound	20.30 (24.74)	6.60 (9.33)	13.45 (19.72)
Other app interactions	Mean number on the screen	1.39 (0.85)	2.11 (2.34)	1.75 (1.78)
Feedback				
Presence of feedback	Proportion of feedback to no feedback	1.00 (0.00)	0.97 (0.12)	0.98 (0.09)
Feedback delivery method	Freq of audio	1.20 (2.90)	3.06 (8.81)	2.37 (7.18)
	Freq of onscreen	1.10 (3.14)	1.59 (3.02)	1.41 (3.02)
	Freq of audio & onscreen	7.90 (6.84)	10.71 (11.17)	9.67 (9.74)
Content of the feedback	Proportion of ostensive feedback compared to other feedback	0.82 (0.37)	0.91 (0.26)	0.88 (0.31)
App design sophistication				
Object property	Freq of static	2.84 (5.69)	5.75 (8.42)	4.33 (7.27)
	Freq of static movement	9.84 (8.40)	22.90 (20.96)	16.54 (17.58)
	Freq of mixed	5.05 (7.54)	9.25 (9.85)	7.21 (8.94)
	Freq of animation	19.16 (25.66)	2.80 (5.28)	10.77 (19.86)

Abbreviation: Freq, frequency.

Finally, there were two significant interactions, website rating by background sound and website rating by object property (see Table 3). Specifically, high-rated apps had higher frequency of no sound ($p = 0.033$) and lower frequency of complex sound ($p = 0.029$) than low-rated apps (see Figure 2). High-rated apps also had a higher frequency of static movement ($p = 0.016$) and lower frequency of animation ($p = 0.009$) than low-rated apps rated (see Figure 3).

Language analysis

In order to present an in-depth assessment of the features of language in the apps, we present the utterance types' analysis followed by the analysis of MLU, frequency, AoA and concreteness. Out of 39 apps, 23 contained language: 11 from Common Sense Media (5 high-, 6 low-rated) and 12 from Good App Guide (8 high-, 4 low-rated). Chi-square analysis showed that the two websites did not differ in the number of high-rated versus low-rated apps that contained language, $\chi^2(1) = 1.051$, $p < 0.305$.

We first compared the proportion of different utterance types between the websites and high- versus low-rated apps. We ran a series of 2 (Website) × 2 (Rating) ANOVAs separately for each construction type. Summary of the results is presented in Table 5. Apps
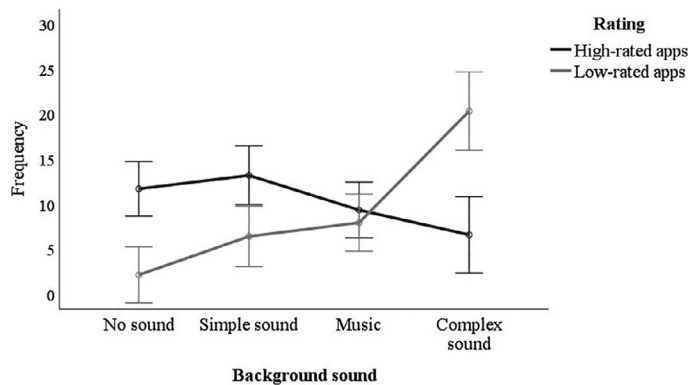

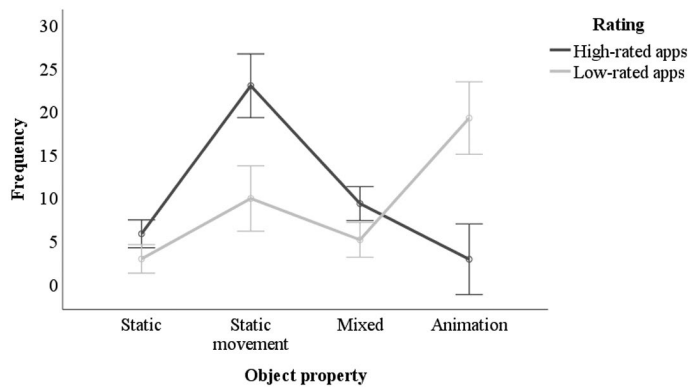**FIGURE 2** The interaction between high- versus low-rated apps and background sound

FIGURE 3 The interaction between high/low-rated apps and app design sophistication

TABLE 5 Utterance types analysis: summary of the analyses (n.s. means non-significant)

	Main effect of website	Main effect of rating	Interaction between website × rating
Fragments	n.s.	n.s.	n.s.
Copulas	n.s.	$F(1,19) = 4.886$, $p = 0.040$, $\eta_p^2 = 0.205$	n.s.
Questions	$F(1,19) = 7.184$, $p = 0.015$, $\eta_p^2 = 0.274$	n.s.	n.s.
Imperatives	$F(1,19) = 4.039$, $p = 0.050$, $\eta_p^2 = 0.175$	n.s.	n.s.
Subject-predicate	n.s.	n.s.	n.s.
Complex	n.s.	n.s.	n.s.

from the two websites differed in the proportion of questions, with apps recommended by Good App Guide ($M = 0.07$, $SD = 0.08$) having a higher proportion of questions than apps recommended by Common Sense Media ($M = 0.01$, $SD = 0.03$). They also differed in the proportion of imperatives, with apps recommended by Good App Guide having a lower proportion of imperatives ($M = 0.10$, $SD = 0.12$) than apps recommended by Common Sense Media ($M = 0.26$, $SD = 0.21$). In terms of the differences between the high- and low-rated apps, the two groups of apps differed only in the proportion of copulas, with low-rated apps having a higher proportion of those constructions ($M = 12.77$, $SD = 3.19$) than high-rated apps ($M = 3.37$, $SD = 2.81$). For the proportional frequency of each utterance type in the two websites, separately for high- and low-rated apps, see Figure 4.

For the descriptive statistics regarding the frequency, AoA and concreteness, see Table 6.

In order to investigate whether high- versus low-rated apps differ in the MLU, and whether this is true for both rating systems, we ran a mixed-effects model with MLU as the dependent variable, a fixed effect of rating (high- vs. low-rated apps) and website (Common Sense Media, Good App Guide), an interaction between website and rating, and a random intercept for individual app. For this analysis, each utterance was counted as a separate observation. The model was fitted with inverse-Gaussian function. As can be seen in Table 7, no significant main effects or interactions were observed.

In order to explore whether high- versus low-rated apps differ with respect to the frequency, concreteness and AoA of words they include, and whether this is true for both rating systems, we ran three mixed-effects models with the frequency, AoA or concreteness as the dependent variable, fixed effects of website and rating, an interaction between website and rating, and a random intercept for app. Each word type in an individual app was counted as a separate observation, with words weighted according to the number of repetitions within the app. Frequency was log transformed. The models were fitted with an inverse-Gaussian function. The psycholinguistic variables analysed in this study are intercorrelated (eg, Balota et al., 2004); thus we initially aimed to include the remaining variables as predictors in a model when investigating one of the variables, to account for variance associated with other variables. However, the models with added variables did not converge, and thus, they had to be simplified as described above.

The model results for frequency are presented in Table 8. We found a significant main effect of rating, with low-rated apps including lower frequency words than high-rated apps. Interestingly, we also found an interaction between website and rating, which showed that the effect of frequency was driven by the apps rated by Common Sense Media website (see Figure 5).

The models for AoA and concreteness did not reveal any significant main effects or interactions (see Tables 9 and 10).
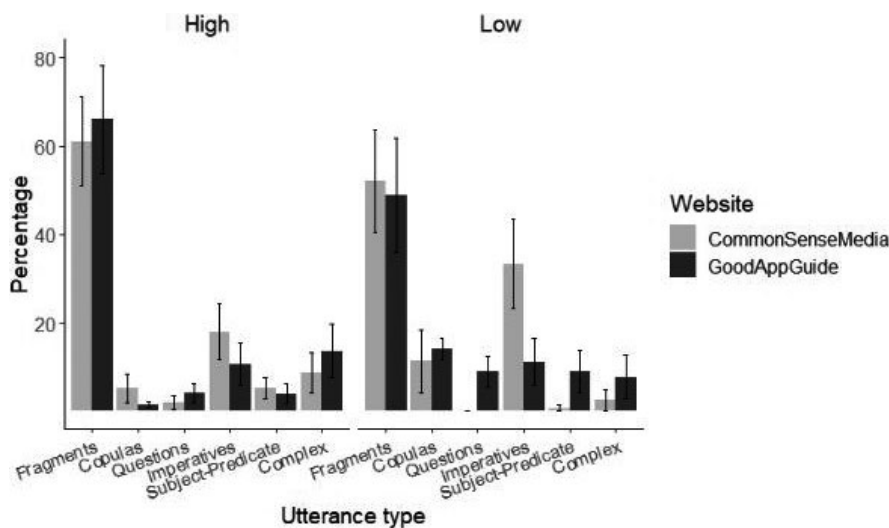
FIGURE 4 Percentage of utterance type in the two websites in high- versus low-rated apps (±1 *SE*)

TABLE 6 Descriptive statistics for frequency, AoA and concreteness for high-rated and low-rated apps in the two websites

	Common Sense Media		Good App Guide	
	High-rated	**Low-rated**	**High-rated**	**Low-rated**
	Mean (*SD*)			
MLU	3.76 (0.25)	4.30 (0.32)	5.98 (0.26)	5.02 (0.38)
Log frequency	2.89 (0.07)	2.46 (0.10)	3.00 (0.07)	3.01 (0.08)
Concreteness	2.60 (0.08)	2.80 (0.12)	2.63 (0.07)	2.56 (0.09)
AoA	4.67 (0.08)	4.63 (0.12)	4.52 (0.07)	4.61 (0.08)

TABLE 7 Mixed model results for the analysis of MLU

	Estimate	**SE**	**t**	**p**
Intercept	4.84	1.10	4.40	<0.001
Website	0.54	1.34	0.40	0.69
Rating	−0.45	1.44	−0.32	0.75
Website:Rating	0.07	1.87	0.04	0.97

Note: R syntax for the model:

glmer(MLU ~ (1 | app) + website*rating, data=data1, glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 10000)), family=inverse.gaussian(link = "identity")).

DISCUSSION

The aim of the present study was to determine whether apps labelled as educational by website rating systems (Good App Guide and Common Sense Media) relate to validated measures when assessed by research informed rating tools of educational potential (Kolak et al., 2021) and whether they have good quality, age-appropriate language. Crucially, the

TABLE 8   Mixed model results for the analysis of log frequency

	Estimate	SE	t	p
Intercept	403.91	11.44	35.30	<0.001
Website	3.96	14.74	0.27	0.79
Rating	−82.53	14.63	−5.64	<0.001
Website:Rating	95.83	22.56	4.25	<0.001

Note: R syntax for the model:

glmer(log_freq ~ rating*website + (1 | app), data=data1, weights = repetitions, glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 10000)), family=inverse.gaussian(link = "identity")).
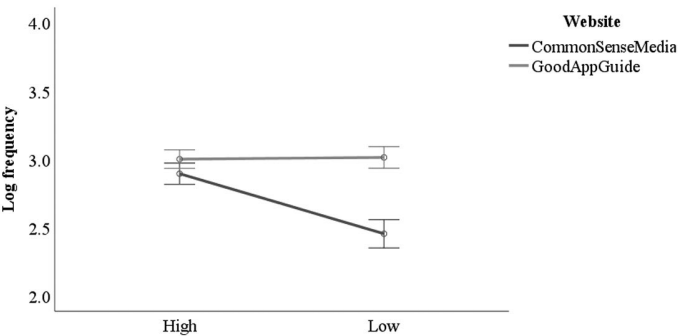

FIGURE 5   An interaction between website and rating for the log frequency analysis

TABLE 9   Mixed model results for the analysis of AoA

	Estimate	SE	t	p
Intercept	96.11	9.96	9.66	<0.001
Website	−18.49	13.00	−1.42	0.155
Rating	8.07	13.13	0.61	0.54
Website:Rating	8.70	18.30	0.46	0.64

Note: R syntax for the model:

glmer(aoa ~ rating*website + (1 | app), data=data1, weights = repetitions, glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 10000)), family=inverse.gaussian(link = "identity")).

app evaluation tools (Kolak et al., 2021) used in the present study were created based on the principles of EYFS framework and were previously assessed for both reliability and validity. Thus, the assessment of apps in the present study is closely aligned with the EYFS learning goals. For example, some of the prime areas of learning and development as outlined in the EYFS framework are 'communication and language' and 'literacy', which include providing children with opportunities to hear and use new words in different contexts. Therefore, the tools used in the present study put emphasis on the presence of high-quality language in the apps. Another prime area outlined in the EYFS framework is 'understanding the world'. Thus, one of the items in the questionnaire for evaluating the educational potential of apps (Kolak et al., 2021)—learning goal—includes in the descriptive criteria "learning about places and environment", and another item—meaningful learning—includes in the descriptive criteria "learning is meaningful and has a purpose (…), the content is relevant to real life". Therefore,

TABLE 10 Mixed model results for the analysis of concreteness

	Estimate	SE	t	p
Intercept	2.68	0.25	10.65	<0.001
Website	0.11	0.33	0.32	0.75
Rating	0.01	0.35	0.04	0.97
Website:Rating	−0.43	0.48	−0.90	0.37

Note: R syntax for the model:

glmer(concreteness ~ rating*website + (1 | app), data=data1, weights = repetitions, glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 10000)), family=inverse.gaussian(link = "identity")).

the assessment of apps in the present study is related to the EYFS framework due to the tools used for app evaluation.

First, we assessed the extent to which website ratings (high/low) relate to the educational potential of the apps. We found that high scoring apps had a higher educational potential and were more likely to contain a learning goal. We then appraised the extent to which the website ratings related to design features that can promote learning and found that all apps in the sample offered a high proportion of feedback with more highly rated apps containing opportunities for feedback. Given the importance of feedback for children's learning (eg, Callaghan & Reich, 2021) and engagement with apps, it is promising that website rating systems do differentiate apps that offer feedback. Thus, to answer our first research question, website rating systems can accurately distinguish between apps with high and low educational potential. However, the high scoring apps were scoring only around 9 out of a possible 20 on our app evaluation questionnaire. Thus, even high scoring apps on website rating systems have considerable room for development to increase their educational potential for preschool age children. Apps rated highly by the two websites scored particularly low on the quality of feedback, which should be specific and constructive to scaffold a child's learning; adjustable content, which should adapt the app according to the child's performance (eg, reducing the number of response options if the child enters an incorrect response); and social interactions and storyline, which can promote children's engagement with the app (see Kolak et al., 2021). These are therefore areas where even high scoring apps could be improved (also see Kolak et al., 2021; Meyer et al., 2021; Taylor et al., submitted). App rating systems therefore need to expand their assessments to incorporate specific features such as good quality feedback and adjustable content for children's apps.

High and low scoring apps did not differ in terms of the frequency of cognitive (eg, selecting the missing element to complete a pattern) versus stimulus reaction activities (eg, swiping to move across the screen). Cognitive activities are important for promoting children's active learning (see Hirsh-Pasek et al., 2015). For example, young toddlers learn better from a touchscreen app if they are required to touch the screen in a specific location rather than touching anywhere on the screen, though for older toddlers touching the screen in a specific location disrupted learning (Kirkorian et al., 2016). Nevertheless, website rating systems do not differentiate apps that use more cognitive activities rather than stimulus reaction activities.

There was no difference between apps in the complexity of the learning environment with the exception that high-rated apps were more likely to contain no sound and less likely to contain complex sound (two sounds playing simultaneously, eg, music and cheering). In addition, in terms of design sophistication, high-rated apps were more likely to contain objects with static movement (eg, a static character that moves across the screen) rather than animations. These findings are consistent with research suggesting that sound effects and animation can interfere with 3–6-year old's digital storybook comprehension (Reich

et al., 2016). Thus, high and low app ratings by website rating systems can capture some of the important features of app design and the learning environment important for facilitating children's learning.

We also assessed whether website ratings were related to the quality and age-appropriateness of language in apps for preschool age children. While high scoring apps performed well in our analysis of the appropriateness of language with reference to clear speech, child-appropriate language and good pace of speech with 70% of highly rated apps scoring two points for the item on our questionnaire, our fine-grained analysis of language revealed that there were no differences in terms of utterance types between high and low-rated apps with the exception of copulas. Low-rated apps had a higher proportion of copulas compared to highly rated apps, research suggests that a lower proportion of copulas, fragments and imperatives but a higher proportion of questions, subject-predicate and complex utterances will provide an enriched source of language input (Cameron-Faulkner & Noble, 2013). Apps rated by the Good App Guide had a higher proportion of questions and lower proportion of imperatives than apps rated by Common Sense Media, making those apps a potentially more enriched source of language input. However, in general, the apps in our sample had a relatively high proportion of fragments and imperatives compared to questions, subject-predicate and complex utterances suggesting that these apps do not provide an enriched source of language input compared to books or CDS (see Kolak et al., submitted for similar findings). Given that app developers submit their apps to the Good App Guide for review, it is possible that those app developers have given more consideration with respect to the educational value of their app and included better quality language. Nevertheless, high and low website ratings did not capture fine grained differences in the quality of the language of children's apps.

The apps rated as high and low on the two websites did not differ in terms of MLU, concreteness or AoA, but differed in terms of frequency. Specifically, for Common Sense Media ratings, low-rated apps rated contained lower frequency words compared to high-rated apps. This was driven by the topics of the activities within the low-rated apps; shapes (eg, trapezoid) and food (eg, aubergine). The psycholinguistic variables for both high and low scoring apps were broadly similar to previous analyses of children's apps, which showed that MLU, concreteness and AoA were similar to CDS (Kolak et al., submitted). App website rating systems therefore do not distinguish between those apps that have the potential to enrich a child's language environment. It is important that app website rating systems should therefore be encouraged to include language assessments in their app ratings and likewise, app developers should incorporate good quality language into children's app design.

To conclude, we argue that while highly rated apps on websites are more likely to be educational than low-rated apps, a highly rated app using a website rating system does not guarantee that the app has educational potential for young preschool age children. Thus, to answer our second research question—there is considerable room for improvement with respect to the potential of app content recommended by the website rating systems to support children's early skill development. While highly rated apps scored high for app design, and reasonably well for opportunities for exploration, solving problems and meaningful learning, highly rated apps did not score well for quality of feedback and adjustable content, which can support learning or social interactions and storyline, which have a potential to keep children motivated and engaged further supporting learning.

Furthermore, website rating systems do not assess the potential for apps to enrich a child's language environment. Preschoolers' language skills are predictive of their literacy development and later academic success (eg, Dickinson et al., 2019; Piasta et al., 2021), and are also crucial for the development of other skills (Pakarinen et al., 2018; Soto-Calvo et al., 2020). Given the considerable amount of time children spend using touchscreens (Rideout & Robb, 2020), we argue that apps should contain good quality language to

support children's language development (also see Kolak et al., submitted; Taylor et al., submitted). Website rating systems should therefore be encouraged to add a separate criterion for language as part of their rating system. An example of how this could be operationalised can be found in the questionnaire for evaluating the educational potential of apps (Kolak et al., 2021), which gives an app 2 points for the quality of language if "App always contains age-appropriate and child-directed language; speech is clear, its pace is slow or moderate and easy to follow. Sentences are not overly complex and not too long. Language is comprehensible." We therefore suggest that policy for app ratings in the app stores or on website rating systems should also include quality of feedback, adjustable content, social interactions, storyline and a more fine-grained analysis of language in their assessments. Furthermore, website rating systems should be more transparent in the criteria used to evaluate apps as well as remaining up to date with the latest research in this fast-developing area in order that the criteria used for assessing apps are openly available to users of the ratings systems.

The findings of this study are of use to parents and early years' practitioners who can be made aware of how useful website rating systems are when compared to the research-based tools for assessing apps. The findings of this paper can increase their awareness of the strengths and weaknesses of the criteria that website rating systems use.

Limitations

We are aware that the attractiveness of apps for young children is not always in their educational value, but present in the extent to which they reflect children's hobbies and interests, such as including their toys, their favourite TV shows and relating to their individual sense of humour. As it was mentioned in the Introduction, previous studies acknowledged the child's perspective on app use (eg, Behnamnia et al., 2020; Marsh et al., 2018). In this paper, the child's perspective as the user of the apps was not taken into account because the aim of the study was to compare whether the criteria that website rating systems use to assess the apps allow to select high-quality apps when compared against principles of children's cognitive development and learning in the digital age. We suggest that educational potential and the child's own interest and engagement with an app are likely to be somewhat orthogonal issues; apps can have entertainment and educational quality—not either/or of the two. In the future studies, it would certainly we worth investigating whether preschool children's preference for apps aligns with apps favoured by website rating systems. However, we know that children are engaged and driven by learning goals (Hirsh-Pasek et al., 2015), which means that educational potential enhances children's engagement with apps as well as their educational development.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

ETHICS STATEMENT
Ethics was not required for this study. Human participants were not used in this research and our app sample was taken from the public domain.

DATA AVAILABILITY STATEMENT

ORCID

Gemma Taylor 🄳 https://orcid.org/0000-0002-0988-7168
Joanna Kolak 🄳 https://orcid.org/0000-0003-2689-7103
Padraic Monaghan 🄳 https://orcid.org/0000-0003-3965-2682

REFERENCES

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Behnamnia, N., Kamsin, A., Ismail, M. A. B., & Hayati, A. (2020). The effective components of creativity in digital game-based learning among young children: A case study. *Children and Youth Services Review*, *116*, 105227. https://doi.org/10.1016/j.childyouth.2020.105227

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Callaghan, M. N., & Reich, S. M. (2021). Mobile app features that scaffold pre-school learning: Verbal feedback and leveling designs. *British Journal of Educational Technology*, *52*(2), 785–806. https://doi.org/10.1111/bjet.13055

Cameron-Faulkner, T., & Noble, C. (2013). A comparison of book text and Child Directed Speech. *First Language*, *33*(3), 268–279. https://doi.org/10.1177/0142723713487613

Common Sense Media. (2020). *App reviews*. https://www.commonsensemedia.org/reviews/age/2/age/3/age/4/category/app/genre/education-58/genre/educational-games-60?sort=field_stars_rating&order=desc

Dardanou, M., Unstad, T., Brito, R., Dias, P., Fotakapoulou, O., Sakata, Y., & O'Connor, J. (2020). Use of touchscreen technology by 0-3-year-old children: Parents' practices and perspectives in Norway, Portugal and Japan. *Journal of Early Childhood Literacy*, *20*(3), 551–573. https://doi.org/10.1177/1468798420938445

Department for Education. (2020). *Early years apps approved to help families kick start learning at home*. https://www.gov.uk/government/news/early-years-apps-approved-to-help-families-kick-start-learning-at-home

Dickinson, D. K., Nesbitt, K. T., & Hofer, K. G. (2019). Effects of language on initial reading: Direct and indirect associations between code and language from preschool to first grade. *Early Childhood Research Quarterly*, *49*, 122–137. https://doi.org/10.1016/j.ecresq.2019.04.005

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, *82*(3), 870–886. https://doi.org/10.1111/j.1467-8624.2011.01576.x

Fotakopoulou, O., Hatzigianni, M., Dardanou, M., Unstad, T., & O'Connor, J. (2020). A cross-cultural exploration of early childhood educators' beliefs and experiences around the use of touchscreen technologies with children under three years of age. *European Early Childhood Education Research Journal*, *28*(2), 1–14.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Good App Guide. (2020). *Good app guide*. https://www.fundamentallychildren.com/good-app-guide/

Griffith, S. F., Hagan, M. B., Heymann, P., Heflin, B. H., & Bagner, D. M. (2020). Apps as learning tools: A systematic review. *Pediatrics*, *145*(1), e20191579. https://doi.org/10.1542/peds.2019-1579

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*, 3–34. https://doi.org/10.1177/1529100615569721

Kirkorian, H. L., Choi, K., & Pempek, T. A. (2016). Toddlers' word learning from contingent and noncontingent video on touch screens. *Child Development*, *87*(2), 405–413. https://doi.org/10.1111/cdev.12508

Kolak, J., Monaghan, P., & Taylor, G. (submitted). Language in educational apps for pre-schoolers. A comparison of grammatical constructions and psycholinguistic features in apps, books and child directed speech.

Kolak, J., Norgate, S. H., 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

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

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.

Meyer, M., Zosh, J. M., McLaren, C., Robb, M., McCaffery, H., Golinkoff, R. M., Hirsh-Pasek, K., & Radesky, J. (2021). How educational are "educational" apps for young children? App store content analysis using the Four Pillars of Learning framework. *Journal of Children and Media*, *15*(4), 526–548. https://doi.org/10.1080/17482798.2021.1882516

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, *63*(4), 506–525. https://doi.org/10.1016/j.jml.2010.08.003

Neumann, M. M., & Neumann, D. L. (2017). The use of touch-screen tablets at home and pre-school to foster emergent literacy. *Journal of Early Childhood Literacy*, *17*(2), 203–220. https://doi.org/10.1177/1468798415619773

Pakarinen, E., Salminen, J., Lerkkanen, M.-K., & von Suchodoletz, A. (2018). Reciprocal associations between social competence and language and pre-literacy skills in preschool. *Journal of Early Childhood Education Research*, *7*(2), 207–234.

Papadakis, S. T., & Kalogiannakis, M. (2020). A research synthesis of the educational value of self-proclaimed mobile educational applications for young age children. In S. Papadakis & M. Kalogiannakis (Eds.), *Mobile learning applications in early childhood education* (pp. 1–19). IGI Global.

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2018). Educational apps from the Android Google Play for Greek preschoolers: A systematic review. *Computers & Education*, *116*, 139–160. https://doi.org/10.1016/j.compedu.2017.09.007

Piasta, S. B., Logan, J. A., Thomas, L. J., Zettler-Greeley, C. M., Bailet, L. L., & Lewis, K. (2021). Implementation of a small-group emergent literacy intervention by preschool teachers and community aides. *Early Childhood Research Quarterly*, *54*, 31–43. https://doi.org/10.1016/j.ecresq.2020.08.002

Reich, S. M., Yau, J. C., & Warschauer, M. (2016). Tablet-based eBooks for young children: What does the research say? *Journal of Developmental & Behavioral Pediatrics*, *37*(7), 585–591. https://doi.org/10.1097/DBP.0000000000000335

Rideout, V., & Robb, M. B. (2020). *The Common Sense census: Media use by kids age zero to eight, 2020*. Common Sense Media.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Soto-Calvo, E., Simmons, F. R., Adams, A. M., Francis, H. N., Patel, H., & Giofrè, D. (2020). Identifying the preschool home learning experiences that predict early number skills: Evidence from a longitudinal study. *Early Childhood Research Quarterly*, *53*, 314–328. https://doi.org/10.1016/j.ecresq.2020.04.004

Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific English words. *Child Development*, *89*(4), 1247–1267. https://doi.org/10.1111/cdev.12731

Taylor, G., Kolak, J., Norgate, S. H., & Monaghan, P. (submitted). Assessing the educational potential and language content of touchscreen apps for preschool children.

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521