Computers and Education Open 3 (2022) 100102

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

Assessing the educational potential and language content of touchscreen apps for preschool children

Gemma Taylor [a,*], Joanna Kolak [a], Sarah H. Norgate [a,1], Padraic Monaghan [b,c]

[a] Department of Psychology, School of Health Sciences, University of Salford, Salford, M5 4WT, UK
[b] Department of Psychology, Lancaster University, Lancaster, UK
[c] Faculty of Humanities, University of Amsterdam, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Early years education
Cognitive development
Touchscreen apps
Digital media
Language development

ABSTRACT

Touchscreen apps have the potential to teach children important early skills including oral language. However, there is little empirical data assessing the educational potential of children's apps in the app market or how apps link to theories of cognitive development to support learning. We compared popular children's apps with a learning goal (N=18) and without (N=26) using systematic evaluation tools to assess the educational potential and app features that may support learning. We also transcribed all utterances in the apps that included language with a learning goal (N=18) and without (N=12) in order to compare a number of psycholinguistic measures relating to accessibility of the language. Apps with a learning goal had higher educational potential, more opportunities for feedback, a higher proportion of ostensive feedback, and age-appropriate language to support learning and language development. Thus, we argue that selecting children's apps based on the presence of a learning goal is a good first step for selecting an educational app for pre-school age children. Nevertheless, app developers could do more to promote exploratory app use, adjust content to a child's performance, and make use of social interactions with characters onscreen in their apps to enhance the educational potential. Children's apps could also make better use of feedback to ensure that it is specific, meaningful and constructive to better facilitate learning.

Introduction

The increase in preschool children's mobile device use internationally [1,2,3,4] combined with the educational potential of interactive mobile devices over and above traditional digital media [5,6,7,8] affords new opportunities for children's skill development and learning. In particular, as preschool language learning is a cornerstone to children's literacy development (e.g., [9,10]) and also serves as a critical gateway to the development of other social and cognitive skills [11,12] the value of interactive mobile devices could be used to enhance learning alongside caregiver-child interactions which do not involve technology. Reflecting this opportunity, in the UK the Department for Education [13] recommend six specific apps for preschool age children that encourage language and literacy learning to enable parents to support children's learning at home. This recommendation is a significant step, as identifying a set of apps perceived as quality provision for children's learning and enjoyment shows recognition of the challenges facing parents and education providers in choosing appropriate apps for their children [14].

Whilst we already know that children's interactive digital media use has the potential to teach children a wide range of vocabulary (see [15,16,17,18,19]) and support their communicative development, unfortunately, research to date has not yet explored the quality of the language included in children's touchscreen apps more generally. Specifically, not all apps are created equal (e.g., [20,7]). Although tools already exist to support parents and early years educators to select high quality educational media for children (e.g., [20]; Department for Education, 2019; [6,21,22,23, 24,25]), these are limited by long lists of criteria, requiring an understanding of educational and developmental theory and terminology and are consequently time consuming to apply (see [5,26] for review). Prior work has also found that the cost of an app [5] and recommendations from websites [27] do not guarantee that an app will be educational for young children.

One alternative approach would be to determine the extent to which

* Corresponding author.
E-mail address: g.taylor4@salford.ac.uk (G. Taylor).

[1] At the time the research was undertaken Dr Norgate had a position at The University of Salford.

https://doi.org/10.1016/j.caeo.2022.100102
Received 13 May 2021; Received in revised form 9 August 2022; Accepted 9 August 2022
Available online 12 August 2022

a single criterion can highlight educational potential, which could substantially simplify the task for parents and practitioners in selecting and recommending apps to support children's development. One promising contender for such a criterion is whether or not an app contains an identifiable learning goal which can then tap into potential for supporting children's development [20,28]. A learning goal is critically important in that it provides the context in which effective learning can take place [29,6]. In this paper, we ask whether an app containing a learning goal is sufficient to identify its educational potential, or if further detailed investigation using an app evaluation tool is necessary to identify educational apps for pre-school age children.

App design features for teaching early school readiness skills

App design has a significant role to play in promoting learning through interactive features [30]. Apps in the app marketplace vary drastically in terms of their content and consequently their possible potential to be educational. Theoretical and experimental research to date suggests that for an app to be educational, the learning within the app should be guided by a specific and well-defined learning goal (e.g., [20,6,28]), hence our focus on learning goal as a key criterion. However, to maximise their potential for learning, educational apps should also promote meaningful and authentic learning experiences rather than rote learning (e.g., [6,21,23]), and ideally also involve problem solving (e.g., [6,21,23]). To increase children's engagement with the content, a good educational app should also implement several other features. It should include meaningful, specific and constructive feedback (e.g., [31,20,6, 32]), give the child a sense of autonomy and agency (e.g., [30,28,23]), for example by promoting exploratory use (e.g., [6]), and adjust the level of difficulty to the child's performance (e.g., [20,33]). Children also benefit from educational apps when the content is embedded in an entertaining narrative, and when apps involve "parasocial" interactions with animated characters presented onscreen (e.g., [34,35]), as well as appropriate and good quality language (e.g., [36]).

There have been several tools developed to evaluate apps (e.g., [26]). However, these are limited by long lists of criteria (18-70+ items), requiring an understanding of educational and developmental theory and terminology and are consequently time consuming to apply (see [5, 26] for review). A recent, more user-friendly tool to assess these educational app features proposed by Kolak et al. [5] addressed the limitations of previously developed app evaluation tools (see [5,26]). Kolak et al. [5] developed two comprehensive tools to measure the educational potential of children's touchscreen apps. Importantly, the tools were informed by theories and research on children's learning and cognitive development in general as well as in the context of digital media (e.g., [37,38,39, 40,41,42,43]) and draw links to a number of existing evaluation tools (e.g., [20,6,21, 22,24]). The first tool – the questionnaire for evaluating the educational potential of children's touchscreen apps – is a user-friendly tool with high reliability (Cronbach's alpha = 0.81), and high content validity [5]. It is designed for use by a general audience and embraces each of the app design features for teaching early school readiness skills discussed above, including the extent to which an app manifests the following: learning goal, meaningful learning, solving problems, feedback, social interactions, opportunities for exploration, plotline, quality of language, adjustable content, and app design [5].

The second tool, providing coding criteria for quantifying app features, is designed for researchers and goes beyond previous app evaluation tools to quantify in detail app features for learning. It includes measurement of: touch gestures, active learning, complexity of the learning environment, feedback, and app design sophistication [5]. Quantifying app features for learning is insightful for determining how frequently an app includes educational features. Touch gestures as defined by Kolak et al. [5] include the four main touch gestures used in the apps: tap, swipe, drag, trace. Russo-Johnson et al. [18] found that 2-4-year-old children from low income and resource (low SES) families learned more object labels for novel objects when they were asked to drag objects versus tap on them, perhaps because dragging involves more active attention compared to tapping. Kolak et al. [5] coding criteria quantify active learning based on activity type (cognitive versus stimulus-reaction activity) and the number of activity goals. Apps facilitate learning when they promote opportunities for active cognition, such as answering questions, making cognitively challenging decisions, and solving problems (e.g., [6, 21]) and variability across learning experiences can support learning [44]. Complexity of the learning environment includes background sound, background visuals, number of screen elements and other app interactions. According to the Cognitive Theory of Multimedia Learning [38,45,46], children's learning might be disrupted if the software requires the child to engage in too much extraneous cognitive processing or includes too much material, such as unnecessary animation, text, graphics, or music. Feedback includes the presence of feedback, feedback delivery method (audio, onscreen) and its content (ostensive/referential vs other feedback). Feedback can promote engagement with the app and scaffold learning by providing an explanation for why the response is correct or incorrect (e.g., [31]) and by providing contingent responses or guide visual attention to relevant information on the screen (e.g., [30]). Finally, app design sophistication assesses the presence of animation and static objects in apps. Research suggests that animation and sound effects can disrupt learning story comprehension from ebooks [47]. Thus, Kolal et al. [5] tools can be used to assess the educational potential of apps for preschool children via the educational index derived from the questionnaire and by quantifying the educational app features. To date, the tools have been used to compare the educational potential of free vs paid apps [5] and to investigate the quality of apps recommended by app rating websites [27].

There are thus a range of tools available to assess the potential educational value of apps, which are beneficial particularly with respect to informing app design of apps for preschool age children. However, these detailed tools may be less helpful for caregivers who typically select apps based on readily accessible information, such as recommendations and internet reviews [48] or whether the app targets learning topics typically covered in the school curriculum [49]. As Hirsh-Pasek et al. [6] argue, the definitive feature of an educational app is that it has a learning goal. However, clearly what constitutes an appropriate learning goal may differ according to whether the perspective is from that of app developers or researchers. For instance, while researchers may place more emphasis on learning goals driven by research evidence in developmental science, in contrast the app developers may place more emphasis on identifying suitable learning goals driven largely by 'end' functionalities or actions like 'play' or 'read' or 'draw'. While this example may oversimplify the difference in perspective, in reality, it uncovers the need to take both into account at the point of designing and applying research.

In any case, as of yet there has been no direct comparison of apps with and without a learning goal to determine whether this stands as a useful single criterion for parents and early years educators to use in selecting high quality educational apps for children. Prior work has made substantial progress in investigating the educational features available in apps marketed as educational or teaching children specific skills including language, literacy and math (e.g., [20,7,50]). These studies demonstrate that apps marketed as educational typically do not promote active and meaningful learning, nor do they adjust content to the user's performance, promote social interactions, or provide scaffolded feedback [7,20,50], which are critical features in supporting children's learning. Callaghan and Reich [20] did also assess whether apps had a clear learning goal, and 79.5% of their app sample had a learning goal, but they did not then compare the educational potential of apps with and without a learning goal, so the use of this as a single criterion is still unexamined.

Furthermore, by examining only apps marketed as educational and reporting solely on the presence or absence of a design feature, the variety and greater nuance of design features (e.g., in addition to a learning goal, how many activities support this learning goal) included in apps currently available for children are not captured by Callaghan and Reich's [20] analysis. For app developers, it is in addition important to reflect on whether the development of an educational app goes beyond the presence or absence of a learning goal to include app features that are established as valuable for learning. Our study thus appraises the quality and quantity of these features that occur in apps that are identified as containing a learning goal or not. A further feature of apps that can have a profound impact on children's learning is the extent to which apps can promote communicative development, and a further aim of our study was to investigate how properties of apps that can relate to language development are currently implemented in apps that have or do not have a learning goal.

Language development and interactive digital media

For language development, a recent meta-analysis of 42 studies found a moderate negative effect of children's overall screen use on children's language development but a positive effect of educational content on children's language development [51]. However, just 3 studies in Madigan et al. [51] meta-analysis measured mobile device use separately to other forms of screen time [52, 53, 54]. van den Heuvel et al. [54] found that mobile device use was associated with expressive language delays at 18-months of age and Moon et al. [53] found that mobile device use was negatively related to expressive language in children aged 3 years but there was no relationship in children aged 4 or 5 years. In contrast, Taylor et al. [52] found that mobile device use was not predictive of language comprehension or production at 6-36 months of age. Importantly, across all 3 studies, mobile device use was limited to time spent using mobile devices rather than assessing the role of different activities on mobile devices [52–54].

The interactivity afforded by touchscreens can help to support learning from touchscreens under some circumstances in a similar way to a social partner (see [30] for review). Indeed, experimental studies demonstrate that young children can learn new words from interactive touchscreen apps [15,16,17,18,19], indicating that apps could have a place for promoting language acquisition in addition to other forms of interaction with children. However, the language included in apps also needs to be appropriate to support language development and provide an enriched source of language input, and the language needs to be predominantly audio rather than onscreen to account for the limited reading abilities of preschool age children.

Research suggests that mean length of utterance (MLU) increases in both child-directed speech and children's own productions according to age [55], and longer utterances are useful for supporting vocabulary development as well as acquisition of grammar (e.g., [56,57]). At the same time, in child-directed speech, single-word utterances are frequent and known to support early stages of language acquisition [58]. In terms of vocabulary, higher-frequency words in children's environments are more likely to be acquired by children [59], concrete words are more likely to be directed to children than more abstract words [60] and words that are more commonly acquired early in children's development are more likely to be accessible to children [58]. Thus, measuring these psycholinguistic variables in children's apps is an important first step for understanding whether children's apps could enrich a child's language environment.

Current study

We discussed earlier how perspectives on what constitutes a learning goal are likely to vary between researchers, app designers and parents. Some apps may communicate the learning goal explicitly in an app description while other apps may contain activities which informally give rise to children's learning or enjoyment, but may not be explicitly advertising a learning focus or goal. Despite this potential for subjectivity in what constitutes a learning goal, we found a high degree of consistency in raters determining whether an app contains a learning goal or not (see Method, below, for more details) from a perspective of educational potential. This distinction is then tested in the present study in our focus on the extent to which each app contains certain activities indicative of a presence of a learning goal based on systematic criteria developed by the researchers rather than app designers or parents. Three research questions are investigated: 1) are there qualitative and quantitative differences in the educational potential of app design according to the presence or absence of a learning goal as determined by the researchers based on the identification of activities involved in using the app, 2) do apps with a learning goal have more age-appropriate language than apps that do not have a learning goal, and 3) are apps with a learning goal likely to be educational for young children in absolute terms, and not only relatively in comparison to apps without a learning goal? We investigated apps targeting preschool age children to enable an in-depth analysis of the features used in children's apps that may contribute to children's early learning and language development. We hypothesise that if learning is taken into consideration when apps have a learning goal, these apps should have a higher educational potential and will contain more features that should facilitate learning, including: engaging touch gestures, active learning, a simpler learning environment, and more and better quality feedback than apps that do not have a learning goal. Furthermore, if language content is effectively tuned to support children's language development, we hypothesise that the language used in apps with a learning goal will be more age appropriate and more varied as assessed by the presentation of words (audio or onscreen) and a number of psycholinguistic measures (mean length of utterance, single/multiword utterances, frequency, concreteness and age of acquisition) than the language used in apps that do not have a learning goal.

Through undertaking a systematic evaluation of the popular app marketplace, our study has two key practical consequences. First, we determine whether selecting apps based solely on the presence or absence of a learning goal alone is sufficient for caregivers and educators to effectively identify educational apps which promote children's early skills development, including language development and school readiness. Second, these findings can enable app developers and stakeholders to plan the types of collaboration needed for both commercial and educational viability. For instance, in the sphere of children's gesture use with touchscreens, there have already been initiatives set up which seek to bridge the gap between research evidence and design practice through the establishment of a framework for evidence-based touchscreen interaction design recommendations for children [61]. Similarly, our research will enable new frameworks for establishing evidence-based design recommendations for children's educational touchscreen apps.

Method

App sample

We used the same sample of 44 apps in the present study as those in Kolal et al. [5]. Apps were identified from the top 10 free and paid charts for ages 5 and under in the Apple, Google and Amazon app stores on 7th June 2018. We initially selected 10 free and paid apps from each app store (i.e., 20 apps per app store), which resulted in 60 apps in total. Our final sample of 44 apps was established after removing duplicates (apps that repeated across the app stores); 19 of those apps were free and 25 were paid. The top 10 charts are based on a number of factors including: number of downloads, uninstalls, user ratings, keywords and updates, however the rankings for the top charts in the app stores are not transparent. Apps were selected from the app stores according to ratings because this is one of the methods parents use when searching for apps for their children [48]. Video-based apps (e.g., YouTube Kids) were excluded from our sample because they primarily promote passive use. Apps were categorised as 'learning goal' ($N = 18$) or 'no learning goal'

($N = 26$)[2] based on the 'Learning goal' item in the questionnaire for evaluating the educational potential of apps [5]. Apps which scored 2 points ("There is a clear overall learning goal(s) targeting early skills development, e.g. linking sounds and letters, counting, learning shapes and colours, teaching about people, places and environment (relevant to each age/stage") or 1 point ("There is no clear overall learning goal but some or all activities within the app teach early skills relevant to each age/stage e.g., selecting objects in a particular colour, matching shapes, selecting ingredients to bake a cake") were categorised as 'learning goal', and apps which scored 0 points ("There is no clear learning goal, e. g. child is avoiding obstacles in a race") were categorised as 'no learning goal'.

Of the 18 apps that included a categorisation of educational or non-educational in the marketplace, 7 were originally labelled as educational but did not have a learning goal in our evaluation (see Table 1). For 4 of the apps, this was because they only involved an adventure plotline with famous characters from children's TV (e.g., dog characters from Paw Patrol on a mission), but did not provide any learning opportunities. For the remaining 3 apps, this was because they were structured as free play with no learning goal, no language and no activities planned within the apps (e.g., drawing). For the apps that were categorised on the marketplace as not educational, 2 were classified as having a learning goal in our study. One of those apps was Peppa Pig Party time. The reason we classified it as having a learning goal was the inclusion of a task instructing the child to count the objects, hence the app provided some training on numeracy. The other app - Tiny Tap - included tasks focusing on letters, numbers and shapes.

Based on this sample size, in the following analyses we had power of 0.8 to detect effect sizes of f = 0.25 or greater ($\eta^2 = .20$) for the main effect of presence of a learning goal, assuming conservatively that there was no correlation among the within subject measures. The sample also meant that we had power of 0.8 to detect effect sizes for main effects or interactions of the within subject measures of f = 0.28 ($\eta^2 = .22$). Cohen described f = 0.10 as a small effect size, f = 0.50 as a medium effect size, and f = 0.80 as a large effect size [62]. Hence, the study was sufficiently powered to provide a good chance of finding any effects that were small, according to Cohen's conventions.

Measures

Apps were coded using two complementary tools: (1) a questionnaire for evaluating the educational potential of apps and (2) coding criteria for quantifying the app features, developed by Kolak et al. [5]. The app evaluation tools measured features of apps that can contribute to children's learning (see Table 2 for summary of the measures).

Questionnaire for evaluating the educational potential of apps for pre-school age children [5]. The questionnaire included 10 items: learning

Table 1
Number of apps described versus classified as having learning goal (out of 44 apps).

	Learning goal in app description		
Study classification	Yes	No	Not available
N of apps classified as Learning goal	15	2	1
N of apps classified as No learning goal	7	18	1

Table 2
List of DVs from each measure and their associated coding.

Measure	App feature	Variable	Coding type
Questionnaire		Educational potential index (sum of the 10 questionnaire items)	Rate each item (learning goal, meaningful learning, solving problems, feedback, social interactions, opportunities for exploration, plotline, quality of language, adjustable content, and app design) on a scale 0-2 based on the whole app use
Coding criteria for app features	Touch gestures	Touch gestures	Code the frequency of each gesture (tap, swipe, drag, trace) during app use
	Active learning	Activity type	Code the frequency of stimulus reactions and cognitive reactions during app use
		Activity goal	Code the number of different activity goals during app use
	Complexity of the learning environment	Screen elements	Code the number of interactive screen elements on screen during app use
		Background visual	Code the frequency of simple (plain/one colour) and complex (multiple colours, features e. g., farm) backgrounds during app use, then calculate the proportion of complex background during app use
		Background sound	Code the frequency of no sound, simple sound, music and complex sound (more than one sound at a time) during app use
		Other app interactions	Calculate the mean number of other app interactions (available beyond the target app gesture) across all screens during app use
	Feedback	Proportion of feedback	Code the number of times app provided feedback when there was an opportunity for feedback, then calculate the proportion of feedback during app use
		Feedback delivery method	Calculate the frequency of audio, onscreen, audio + onscreen feedback delivery method during app use
		Feedback content	

(*continued on next page*)

[2] The imbalance in the number of apps in the learning goal vs no learning goal category is motivated by our aim to focus on the most popular apps and the features relating to educational potential that they contain. If we had instead selected an equal set of apps with and without a learning goal, then there would have been an imbalance in the popularity (because we would be selecting learning goal apps from outside the top 10 apps, which could potentially include a confound).

Table 2 (*continued*)

			Code the frequency of ostensive/referential and non-specific feedback during app use, then calculate the proportion of ostensive feedback
	App design sophistication	Object property	Code the frequency of static, static movement, animation and mixed property during app use
Language	Psycholinguistic variables	Utterance presentation MLU Single/multi-word utterances Frequency Concreteness Age of acquisition	Code the frequency of utterances provided onscreen/audio during app use Code the mean length of each utterance during app use Code the frequency of single and multi-word utterances during app use Assign frequency, concreteness and age of acquisition to each word

goal, meaningful learning, solving problems, feedback, social interactions, opportunities for exploration, plotline, quality of language, adjustable content, and app design. Each app could score 0-2 points for each of the 10 items, 20 points in total, with higher scores indicating higher educational potential. The evaluation questionnaire can be found in the supplemental materials (see Table A1 in Supplemental material available from OSF: https://osf.io/8bdf3/?view_only=bfab6375669f4740b889160af673db19); descriptors for scoring the app as 0, 1 or 2 are given in detail using examples from apps on the app market. Kolak et al. [5] specified a 0-2 scoring system on the basis that the tool could be used by caregivers, educators and app developers as well as researchers, and we followed the scoring outlined in Kolak et al. [5]. As part of the Kolak et al. [5] study in which we described the development of the tools, two independent raters scored the 5-minute recordings of the apps using the evaluation questionnaire. We included all the app ratings for all the questionnaire items from both raters in the inter-rater reliability analysis. A single IRR was calculated for the total educational potential index. Based on this analysis, and according to McHugh's [63] six category classification of levels of agreement, the inter-rater reliability was interpreted to be 'almost perfect' (κ = .912, $p < .001$).

Coding criteria for quantifying app features [5]. This was the second tool used to quantify app features, where 5 minute recordings of app use (taken while the second author used the apps) were coded in ELAN 5.2 to assess: touch gestures, active learning, complexity of the learning environment, feedback, and app design sophistication (see Table B1 in Supplemental material available from OSF which includes detailed coding instructions: https://osf.io/8bdf3/?view_only=bfab6375669f4740b889160af673db19).

Following the instructions for coding, each screen of the app use video (lasting few seconds) was separated and prepared for the annotations using ELAN 5.2. The frequency of each app feature was coded separately for each screen, For example, for a touch gesture feature for a given screen, the coder would code 'tap x 5' if 5 taps were required. Once the coding was completed, the coder would calculate the total frequency of all touch gestures (tap, swipe, drag, trace) across the app use. Coding instruction specifies that researchers can calculate frequencies or proportions of app features for their analyses. For example, we recommend calculating the proportion (rather than the frequency) of complex background during app use. To calculate it, the coder needs to perform the following calculation: the frequency of complex background on all the screens of the app use video / the frequency of complex background + the frequency of simple background

A second independent rater coded 5 apps from the sample, using all 11 coding categories; inter-rater reliability for this quantification of app features measure was high (κ = .889, $p < .001$). We included the app ratings for all the coded categories by the two raters in the inter-rater reliability analysis.

Psycholinguistic variables. To determine the age-appropriateness of language used in the two app categories, we looked at five psycholinguistic measures. First, we examined mean length of utterance (MLU), which measures the length of language productions in terms of the number of words the utterance contains and is an indicative measure to characterise quality of language input. An utterance was defined as a sequence of words that were separated, either auditorily by a pause of two seconds or more, or visually by spatial or temporal distance (e.g., the sequence of words occurred at different parts of the screen, or occurred on the screen at an interval of two seconds or more. Second, we measured the frequency of multi-word utterances and words presented on their own for each app. Third, we measured mean frequency of the words within the app indexed from a corpus of child-appropriate speech from television programmes. The corpus was derived from programme transcripts from a UK public broadcast television channel – cbeebies – directed to children aged up to 6 years [64] which contains 5,848,083 tokens, values were log-compressed. Fourth, we measured concreteness of words in the apps. Fourth, we took measures of concreteness ratings for each word in the app, obtained from Brysbaert et al. [65] on a scale from 1 (abstract word, e.g., accidents, echo, gross) to 5 (concrete word, e.g., apple, finger, dinner). Finally, we assessed age of acquisition. Age of acquisition ratings for words were obtained from Kuperman et al. [66], where participants rated words according to when they were acquired in years, with ratings greater than 25 years reduced to a ceiling of 25 years, thus age of acquisition was in the range 0 to 25. Frequency, concreteness and age of acquisition were calculated separately for each word in each utterance, then the means of each of those psycholinguistic variables were estimated for each utterance and a grand mean from all utterances was calculated for each app.

A second independent rater, who had coded the 5 apps from the sample for the app features inter-rater reliability, transcribed all the utterances from the apps as well. The inter-rater reliability was high, κ = .857, $p < .001$.

Procedure

Each app was downloaded and a five-minute screen recording was taken while the second author used the app. During the 5-minute app use, each available feature on the screen was used once and activities were completed in the order suggested by the app to ensure that the main features of the app were explored during use (see [5], for justification). Inter-user reliability was determined by comparing coded app use data (touch gesture- drag, tap, swipe, trace, number of different possible activity goals, mean number of possible app interactions available on a screen, object property - static and animated screen elements; see [5] for details) for 5 apps from the sample that were also used by a second independent user. Overall, inter-user reliability for the agreement with respect to transcription of sentences from apps (i.e., whether the raters have identified the same sentences from the stream of speech within apps) was high (κ = .872, p < .001). Each utterance was compared and the raters had to agree on the utterance completely for agreement to be counted for that utterance. The main reason for discrepancies was when one of the coders failed to hear the utterance correctly. In the case of discrepancy, the main coder (the second author who transcribed the whole sample of apps) listened to the recording again (in slower pace, if necessary) and decided which utterance was correct.

In addition, all the utterances that were presented either as (a) audio, (b) onscreen, or (c) audio & onscreen simultaneously during app use were transcribed. The language presented as audio included any audio instructions or feedback within the app (e.g., “Trace the letter”, “Well done!”) or any conversation between the characters (e.g., “Are they ready, Flop?”). The language presented onscreen included titles of the activities to choose from on the screen (e.g., “Puzzles”, “Songs”), any information about the app or the activities (e.g., “Level 1”, “Change the language”), or words presented onscreen as part of the literacy activity (e.g., “cat”, “car”). The language presented as audio & onscreen included instances when the instruction was simultaneously given audio by the narrator/character and presented onscreen as text, or when the songs within the app were sung aloud by the narrator/characters and the lyrics were displayed onscreen.

Results

Educational potential

To test whether there is a difference in the educational potential between apps with and without a learning goal (as measured with the questionnaire for evaluating the educational potential of apps), and whether that was a similar effect for all the app stores, we ran a 2 (Presence of a learning goal) x 3 (App store) ANOVA with educational potential as a dependent variable. App store was included in order to account for any variance associated with different ranking systems to provide a clearer assessment of the variance associated with presence of a learning goal. The main effect of learning goal was significant ($F(1,38)=41.260$, $p<0.001$, $\eta_p^2=0.521$). Apps with a learning goal had higher educational potential (M=10.83, SD=2.50) than apps without a learning goal (M=4.04, SD=2.55). As anticipated, there was no main effect of App store ($F(2,38)=2.448$, $p=0.100$, $\eta_p^2=0.114$) and no interaction between learning goal and app store ($F<1$).[3] Fig. 1 shows the breakdown of number of apps against each educational potential property, divided by those with and without a learning goal, in order to give an impression of where the effect of learning goal was observed. Key contributors were a greater number of apps with learning goals containing meaningful learning, solving problems, feedback, quality of language, and adjustable content.

Interestingly, only 50% of apps with a learning goal scored 2 points for the learning goal item on the questionnaire (the remaining ones scored 1 point). An app could score 2 points if it included a clear overall learning goal(s) targeting early skills development, e.g., linking sounds and letters, counting. However, it could score only 1 point if there was no clear overall learning goal but some or all activities within the app taught early skills relevant to each age/stage e.g., selecting objects in a particular colour, matching shapes. For example, an app called ABC Tracing scored 2 points for learning goal because the app’s overall learning goal was linking sounds and letters, while an app called Peppa Pig Holidays scored 1 point for learning goal because it did not include an overall learning goal but involved one activity targeting counting and one activity targeting reading, among various other activities that did not target early skills development.

App features

To test whether quantity and type of features differed for apps with or without a learning goal, we conducted ANOVAs with learning goal, and feature as factors and frequency of occurrence as dependent variable. App store was also included as a factor to determine if patterns were similar across the stores. As shown in Tables C1 and C2 in Supplemental materials, we found a main effect of touch gesture overall, tapping occurred more frequently than swiping and tracing, and dragging was more frequent than swiping and tracing. There was also a main effect of activity type, the apps in our sample had a higher frequency of cognitive activities (e.g., solving a problem, finding correct answer, making decision) than stimulus-reaction activities (e.g., avoiding obstacles during a car ride). We also found a main effect of object property across the apps, static object property was more frequent than static movement and mixed.

The only differences we found between apps with and without a learning goal were to do with feedback; apps with a learning goal provided more activities that require feedback than apps with no learning goal. To assess the role of different types of feedback on the educational potential of the apps, we categorised feedback as ostensive and/or referential, or non-specific. We categorised feedback as ostensive if it included any parasocial, motivational or neutral audio feedback (e.g. “Correct”) or visual cues to indicate the correct answer (e.g. the box with the correct answer is shaken or highlighted). Apps which did not provide activities requiring feedback were excluded from the analysis. The final sample for the feedback analyses consisted of 16 apps with a learning goal and 16 apps with no learning goal. Apps with a learning goal had higher proportion of ostensive feedback (M=0.84, SD=0.21) than apps with no learning goal (M=0.66, SD=0.43). Crucially, none of the other 9 features differed between apps with and without a learning goal.

Language analysis

To test whether apps with and without a learning goal differed in the proportion of apps which included either audio or onscreen language, a chi-square test was performed. Overall, there was a significant association between the presence of a learning goal and presence of audio or onscreen language during app use ($\chi2(1)=7.700$, $p=0.001$), with 100% of apps with a learning goal including some form of language, compared to only 46% of apps with no learning goal. The final sample for the more detailed language analyses consisted of 18 apps with a learning goal and 12 apps with no learning goal. The 14 apps with no learning goal which had no language content were omitted from these analyses. Notably 13/18 apps with a learning goal either explicitly targeted language development or contained language related activities during app use, whereas none of the apps without a learning goal targeted language development.

For each language property as dependent variable, we conducted ANOVAs with learning goal or no learning goal, and app store as factors. For the analysis of utterance presentation, we also included whether this was audio, onscreen, or both as a factor. As shown in Tables C3 and C4 in Supplemental materials, we found a main effect of utterance presentation, the frequency of language presented as audio (M=34.60, SD=4.90) was greater than the frequency of language presented as both audio and onscreen (M=8.37, SD=2.70). Also, the frequency of language presented onscreen (M=20.60, SD=4.35) was greater than the frequency of language presented as both audio and onscreen. We also found a main effect of utterance context, overall, the apps in our sample contained more multi-word (M=43.10, SD=4.37) than single word utterances (M=20.63, SD=4.03). The only difference we found between apps with

[3] To confirm that having a very similar variable as the independent variable (presence of a learning goal) that is also part of the scoring for the dependent variable (scoring for the learning goal item constitutes part of the combined score on the educational potential index) did not make the analysis circular, we repeated the analysis after omitting learning goal and meaningful learning (as apps that do not have a learning goal also do not include meaningful learning) on the questionnaire ensuring that any possible circularity between independent variable and dependent variable did not contribute to the overall results. The 3 (App store) x 2 (Presence of learning goal) ANOVA showed no main effect of App store ($F(2,38)=2.822$, $p=0.072$, $\eta_p^2=0.129$) and no interaction between App store and Presence of learning goal ($F<1$). The main effect of Presence of learning goal was significant, $F(2,38)=19.170$, $p<0.001$, $\eta_p^2=0.335$, revealing that apps with learning goal had higher educational potential than apps without learning goal, even after omitting learning goal and meaningful learning from analyses.

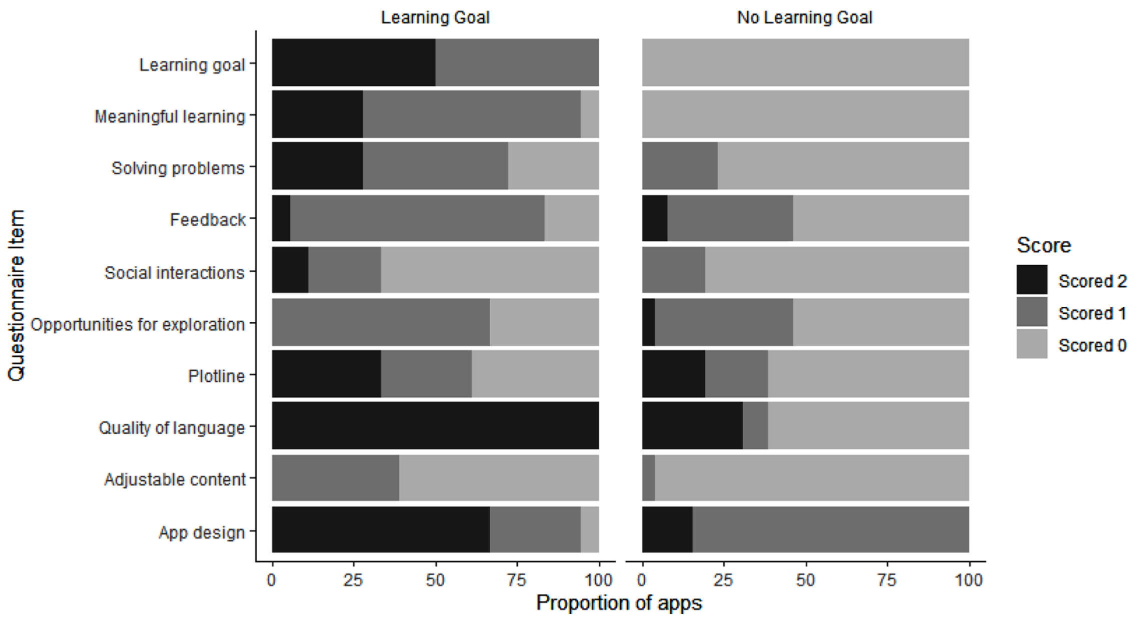

Fig. 1. The proportion of apps with (N=18) and without a learning goal (N=26) that scored 2, 1 or 0 points on each item of the questionnaire for evaluating the educational potential of apps.

and without a learning goal was for AoA, which is a key property for effective communication with children: words included in apps with a learning goal tending to have earlier AoA (M=4.58, SD=0.28) than words included in apps with no learning goal (M=5.22, SD=0.24).

Discussion

The present study investigated three research questions: 1) are there qualitative and quantitative differences in the educational potential of app design according to the presence or absence of a learning goal), 2) do apps with a learning goal have more age-appropriate language than apps that do not have a learning goal, and 3) are apps with a learning goal likely to be educational for young children in absolute terms, and not only relatively in comparison to apps without a learning goal? To address these three questions, we took a representative set of popular children's apps and identified whether or not these apps contained a learning goal (where this was determined by the authors based on the implicit activities within the app). We then compared the features available to support learning in these apps, as well as assessing the language content of those apps, for their educational potential.

Regarding the first research question, as demonstrated by our qualitative analysis, apps with a learning goal scored higher on the educational potential index, as hypothesised. Our quantitative analysis provided further insight into the differences between apps with and without a learning goal. Specifically, apps with a learning goal also included feedback to support learning consistent with prior work [20, 32]. Apps with a learning goal provided more activities requiring feedback than apps with no learning goal. Overall, a mixture of audio and onscreen feedback was included in apps that do provide feedback. Hirsh-Pasek et al. [6] argue that feedback will likely facilitate a child's engagement with an app which can subsequently promote learning. Apps with a learning goal also provided a higher proportion of ostensive feedback. Ostensive feedback can increase attention to the target to-be-learned material and facilitate learning as a result [67]. Consistent with prior work, however, apps with a learning goal need to ensure that feedback is specific, meaningful and constructive to scaffold learning [20], rather than simply motivational feedback (e.g., "Well done") typically used by apps in our sample, to better support children's learning during app use. Thus, there are qualitative and quantitative differences in the educational potential of app design according to the presence or absence of a learning goal. However, it is worth noting that in the quantitative analyses, the differences between apps with and without a learning goal were only present in 2 out of 9 app features. According to the "active learning" pillar (c.f. [6]) we would expect apps with a learning goal to include more cognitively active activities, whereas in the present sample we did not see the difference between the apps with and without a learning goal when it comes to the frequency of cognitively active activities. We would also expect apps with a learning goal to include fewer sound effects and animations than apps without a learning goal, as previous literature suggests that sound effects and animations can interfere with pre-schoolers' comprehension of content in digital media [47]. Taylor et al. [27] reported that apps rated high by app rating websites were more likely to contain no sound rather than complex sound (two sounds playing simultaneously) compared to low-rated apps and were more likely to include static objects rather than animations compared to low-rated apps.

In answer to the second research question, we found that apps with a learning goal were more likely to contain language and included words with an earlier age of acquisition compared to apps with no learning goal. All apps with a learning goal contained language compared to less than half of the apps with no learning goal. For all apps, a higher proportion of utterances were audio or onscreen only compared to both audio and onscreen. It is encouraging that apps with a learning goal are more likely to contain language and that audio language is included given the strong correlation between the number of words children hear and their language development [68,69,70,71]. The words used in apps with a learning goal tended to have earlier age of acquisition (age at which the word is typically learnt) which makes them more suitable for preschool age children (e.g., [58]). Thus, these findings suggest that apps with a learning goal may provide a valuable source of additional language input for young children (see [72], for related findings with storybooks). However, it is important to note that a large proportion of apps with a learning goal were specifically targeting language development. This means that the high quality of language that we found in the sample of apps with a learning goal, might be due to the fact that those apps specifically targeted language and thus the developers might have put more emphasis on good quality language in the process of app development. It cannot be concluded at this stage whether this finding could be extended to other apps, e.g., those focussing on maths.

Addressing our third research question, we found that selecting apps based on the single criterion of learning goal alone will yield an app that contains more features for supporting learning than an app without a learning goal. However, though the apps with a learning goal gained a higher educational potential score, there was still substantial space for increasing features to further promote learning. Even apps with a learning goal were only scoring around 10 (range: 7 - 15) out of a possible 20 on the questionnaire for assessing the educational potential of apps for pre-school age children [5]. There are two key practical consequences of these research questions. First, the analyses presented here test whether caregivers or practitioners identifying whether an app has a learning goal is a useful shortcut to finding apps with educational potential for children. The answer from our results is: yes. Yet, it is worth noting that there is a gap between marketplace categorisations and learning potential of apps. As mentioned in the Method, seven apps originally labelled as educational in app marketplaces did not appear to have a learning goal, at least according to our evaluation. That is, under our research-based criterion they did not appear to target early skills development (e.g., linking sounds and letters, counting, learning shapes and colours, teaching about people, places and environment). In this way, there are apps on the app market which make the claim to be educational which from a research perspective do not reach the threshold for being educational, by having the inclusion of a learning goal [6]. However, one limitation of our study is that we did not see how children would use the apps in context, so there might have been more impact on their skills and understanding than anticipated. In future, efforts to quantify the extent to which children develop skills and understanding before and after using specific apps would bring new insights about any changes in children's learning potentially attributable to the app design and surrounding learning context. In addition, researching children's perspectives about their experiences of using any target apps could generate new knowledge about children's perceptions of skills acquired as well as any impact on their enjoyment and/or emotional wellbeing.

The second consequence of our research is the implications of our analyses for app developers in maximising the educational potential of apps. We note above that there remains substantial room for improvement of apps with regard to features that can promote educational potential, even in apps that contain a learning goal. For instance, though apps can contain language that supports children's language learning, app developers should further consider the role of onscreen language in apps for pre-school age children and whether this might be better as audio rather than written language given the limited reading abilities of pre-school age children and the fact that children often use apps independently [73,74].

For other app design features that can increase children's engagement and learning, no apps with a learning goal scored the maximum 2 points for opportunities for exploration or adjustable content. Apps with a learning goal therefore rarely fully promoted exploratory use which can give a child a sense of autonomy and agency [30,28,23] or adjusted the level of difficulty within the app to the child's performance to support learning [20,33]. In addition, less than half of the apps with a learning goal facilitated social interactions which are important for children's engagement [34,35]. On the positive side, apps with a learning goal were more likely to include meaningful learning experiences (relevant to real life e.g., learning a bedtime routine) and involve solving problems (e.g., selecting items of a certain colour and shape; [6,21,23]). However, few apps scored 2 points on these items.

Furthermore, there was no difference between apps with and without a learning goal for touch gestures. Overall, apps required tapping and dragging touch gestures more often than swiping and tracing. Given that different touch gestures can facilitate learning in pre-school age children [18], app developers could consider customizing apps for use with age-appropriate touch gestures. For instance, Aziz, Sin, Batmaz and Chung [75] concluded from samples based in Malaysia and UK that by the age of four years children typically use a full repertoire of seven touchscreen gestures (tap, drag/slide, drag/drop, rotate, pinch, flick and spread) in contrast to three-year olds (tap, drag/slide, drag/drop, rotate, pinch, flick) and two-year olds (tap, drag/slide and flick) . Given this variation in gesture use characterizing the pre-school period, it is essential that future app design continues to be informed by evidence-based touchscreen interaction design recommendations for children [61]. A limitation of the present study is our relatively small sample of 44 apps. However, this is within a similar scale to comparable research (e.g., [24]) and besides, we selected these apps to be representative of the most popular apps available in the app marketplace. Our coding of the apps using the detailed coding scheme applied to actual usage of the apps provides one of the first thorough analysis of app features for children's apps in the app market, providing a link between the app features and whether the app has an explicit learning goal or not for the user. Nevertheless, analysing 44 apps meant that we were restricted to examining effect sizes that were medium or large, with the study not sufficiently powered to be likely to detect small effect sizes for differences between distribution of features. The post-hoc power analyses confirmed that our sample of apps was sufficient for detecting the key results that we intended to observe (with power exceeding 0.99 for relations between learning goal and educational potential, and the more fine-grained analyses of Age of Acquisition of language contained in apps with and without learning goals). Future studies that involve larger samples could feasibly detect a greater range of differences, and the current study can thus be taken as a first step in this field that highlights medium to large effect sizes that occur between apps with and without a learning goal.

A second limitation is that we selected apps from the top 10 lists in the Apple, Google and Amazon app stores and these lists are constantly changing. However, our analysis provides a snapshot of how apps within these top 10 lists relate to learning potential, regardless of whether they are educational or non-educationally oriented. A further limitation is that we have not explicitly linked app use to specific learning outcomes. However, our analysis of the apps is grounded in theoretical models of children's language and communicative development. The tool for educational potential is validated by expert users, practitioners and caregivers [5], nevertheless future work could more explicitly relate language and vocabulary development to children's use of particular apps, to make the link between theory and outcomes more concrete.

Finally, a third limitation of our study is that we did not take into account the children's perspective, e.g., the role of children's individual differences and engagement during app use in their learning from apps. Ongoing empirical work in our lab is investigating the role of the child and the wider context on children's ability to learn from educational apps as well as mapping out how the theoretically educational features and content of apps outlined in the present paper and in Kolak et al. [5] help children learn from apps in practice.

Conclusion

The present study suggests that selecting apps to support the development of children's early skills based on the presence of an explicit learning goal alone is a good first step. That less than half of the apps that we sampled had a learning goal is surprising, particularly given that apps have the potential to be educational due to their interactive nature [6,30]. App developers should therefore be encouraged to design children's apps with a learning goal. To answer our key research questions, we found that apps with versus without a learning goal scored higher on the educational potential index, provided more opportunities for feedback, and contained a higher proportion of ostensive feedback. In addition, we found that apps with a learning goal were more likely to contain language and include words with an earlier AoA suggesting that apps could provide a valuable source of language input for young children. Our research highlights a number of underutilised features for designing educational apps for children. In particular, apps should do more to promote exploratory use, adjust content to a child's performance, and make use of social interactions to promote engagement with the app. Furthermore, feedback in apps should move away from simply motivational messages to meaningful, specific and constructive feedback. Including these features in apps, alongside a learning goal, could then increase the educational potential of apps in supporting children's early cognitive and social development.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Eve M. Bent, for help with coding. This work was supported by the Economic and Social Research Council [grant number ES/R004129/1].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.caeo.2022.100102.

References

[1] Kabali HK, Irigoyen MM, Nunez-Davis R, Budacki JG, Mohanty SH, Leister KP, Bonner RL. Exposure and use of mobile media devices by young children. Pediatrics 2015;136(6):1044–50.
[2] Konok V, Bunford N, Miklósi Á. Associations between child mobile use and digital parenting style in Hungarian families. J Children Media 2020;14(1):91–109.
[3] Ofcom. Children and parents media use and attitudes: annex 1. Children's research annex; 2019. Retrieved 13/08/20 from, https://www.ofcom.org.uk/_data/assets/pdf_file/0027/134892/Children-and-Parents-Media-Use-and-Attitudes-Annex-1.pdf.
[4] Rideout V. The Common Sense census: media use by kids age zero to eight. Common Sense Media. CA: San Francisco; 2017. p. 263–83.
[5] Kolak J, Norgate SH, Monaghan P, Taylor G. Developing evaluation tools for assessing the educational potential of apps for preschool children in the UK. J Child Media 2021;15(3):410–30.
[6] Hirsh-Pasek K, Zosh JM, Golinkoff RM, Gray JH, Robb MB, Kaufman J. Putting education in 'educational' apps: Lessons from the science of learning. Psychol Sci Public Interest 2015;16:3–34.
[7] Meyer M, Zosh JM, McLaren C, Robb M, McCaffery H, Golinkoff RM, Radesky J. How educational are "educational" apps for young children? App store content analysis using the Four Pillars of Learning framework. J Children Media 2021:1–23.
[8] Xie H, Peng J, Qin M, Huang X, Tian F, Zhou Z. Can touchscreen devices be used to facilitate young children's learning? A meta-analysis of touchscreen learning effect. Front Psychol 2018;9:2580.
[9] Dickinson DK, Nesbitt KT, Hofer KG. Effects of language on initial reading: Direct and indirect associations between code and language from preschool to first grade. Early Childhood Res Q 2019;49:122–37.
[10] Piasta SB, Logan JA, Thomas LJ, Zettler-Greeley CM, Bailet LL, Lewis K. Implementation of a small-group emergent literacy intervention by preschool teachers and community aides. Early Childhood Res Q 2021;54:31–43.
[11] Pakarinen E, Salminen J, Lerkkanen M-K, von Suchodoletz A. Reciprocal associations between social competence and language and pre-literacy skills in preschool. J Early Childhood Educ Res 2018;7(2):207–34.
[12] Soto-Calvo E, Simmons FR, Adams AM, Francis HN, Patel H, Giofrè D. Identifying the preschool home learning experiences that predict early number skills: Evidence from a longitudinal study. Early Childhood Res Q 2020;53:314–28.
[13] Department for Education. Early years apps approved to help families kick start learning at home. 2020. Retrieved 13/08/20 from, https://www.gov.uk/government/news/early-years-apps-approved-to-help-families-kick-start-learning-at-home.
[14] Guernsey L, Levine M, Chiong C, Severns M. Pioneering Literacy in the Digital Wild West: Empowering Parents and Educators. New York: The Joan Ganz Cooney Center at Sesame Workshop; 2012.
[15] Chiong C, Shuler C. Learning: Is there an app for that. In Investigations of young children's usage and learning with mobile devices and apps 2010:13–20.
[16] Dore RA, Shirilla M, Hopkins E, Collins M, Scott M, Schatz J, Toub TS. Education in the app store: using a mobile game to support US preschoolers' vocabulary learning. J Children Media 2019:1–20.
[17] Kirkorian HL, Choi K, Pempek TA. Toddlers' word learning from contingent and noncontingent video on touch screens. Child Dev 2016;87(2):405–13.
[18] Russo-Johnson C, Troseth G, Duncan C, Mesghina A. All tapped out: touchscreen interactivity and young children's word learning. Front Psychol 2017;8:578.
[19] Walter-Laager, Brandenberg K, Tinguely L, Schwarz J, Pfiffner MR, Moschner B. Media-assisted language learning for young children: Effects of a word-learning app on the vocabulary acquisition of two-year-olds. Br J Educ Technol 2017;48(4):1062–72.
[20] Callaghan MN, Reich SM. Are educational preschool apps designed to teach? An analysis of the app market. *Learning,*. Med Technol 2018;43(3):280–93.
[21] Lee C-Y, Cherner TS. A comprehensive evaluation rubric for assessing instructional apps. J Inf Technol Educ 2015;14:21–53.
[22] Lee JS, Kim SW. Validation of a tool evaluating educational apps for smart education. J Educ Comput Res 2015;52(3):435–50.
[23] Papadakis S, Kalogiannakis M. Mobile educational applications for children: what educators and parents need to know. Int J Mobile Learn Organ 2017;11(3):256–77.
[24] Papadakis S, Kalogiannakis M, Zaranis N. Educational apps from the android Google play for Greek preschoolers: a systematic review. Comput Educ 2018;116:139–60.
[25] Shoukry L, Sturm C, Galal-Edeen GH. Pre-MEGa: a proposed framework for the design and evaluation of preschoolers' mobile educational games. Innovations and advances in computing, informatics, systems sciences, networking and engineering. Cham: Springer; 2015. p. 385–90.
[26] Papadakis S, Kalogiannakis M. A research synthesis of the real value of self-proclaimed mobile educational applications for young children. Mobile Learn Applic Early Childhood Educ 2020:1–19.
[27] Taylor G, Kolak J, Bent EM, Monaghan P. Selecting educational apps for preschool children: How useful are website app rating systems? Br J Educ Technol 2022.
[28] Kucirkova N. iPads in early education: separating assumptions and evidence. Front Psychol 2014;5.
[29] Darling-Hammond L, editor. Powerful learning: What we know about teaching for understanding. San Francisco, CA: Jossey-Bass; 2008.
[30] Kirkorian HL. When and how do interactive digital media help children connect what they see on and off the screen? Child Dev Perspect 2018;12(3):210–4.
[31] Blair KP. Learning in critter corral: evaluating three kinds of feedback in a preschool math app. In: Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13; 2013. p. 372–5.
[32] Nikolayev M, Reich SM, Muskat T, Tadjbakhsh N, Callaghan MN. Review of feedback in edutainment games for preschoolers in the USA. J Children Media 2021;15(3):358–75.
[33] Rosell-Aguilar F. State of the app: a taxonomy and framework for evaluating language learning mobile applications. CALICO J 2017;34(2):243–58.
[34] Calvert SL. Children and digital media. In: Bornstein M, Leventhal T, Lerner R, editors. Handbook of child psychology and developmental science. 7th 2015. p. 375–415.
[35] Oxford Ed. Calvert SL, Richards MN. Children's parasocial relationships with media characters. In: Bossert J, Jordan A, Romer D, editors. Media and the well being of children and adolescents; 2014.
[36] Danielson K, Wong KM, Neuman SB. Vocabulary in educational media for preschoolers: a content analysis of word selection and screen-based pedagogical supports. J Children Media 2019;13(3):345–62.
[37] Aladé F, Lauricella AR, Beaudoin-Ryan L, Wartella E. Measuring with Murray: touchscreen technology and preschoolers' STEM learning. Comput Hum Behav 2016;62:433–41.
[38] Mayer RE. Incorporating motivation into multimedia learning. Learn Instr 2014;29:171–3.
[39] Mulliner E, Tucker M. Feedback on feedback practice: perceptions of students and academics. Assess Eval Higher Educ 2017;42(2):266–88.
[40] Nacher V, Jaen J, Navarro E, Catala A, González P. Multi-touch gestures for pre-kindergarten children. Int J Hum Comput Stud 2015;73:37–51.
[41] Schroeder EL, Kirkorian HL. When seeing is better than doing: preschoolers' Transfer of STEM skills using touchscreen games. Front Psychol 2016;7.
[42] Schwartz DL, Tsang JM, Blair KP. The ABCs of How We Learn: 26 Scientifically Proven Approaches, How They Work, and When to Use Them. W. W. Norton & Company; 2016.
[43] Vygotsky LS. Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press; 1978.
[44] Thiessen ED. When variability matters more than meaning: the effect of lexical forms on use of phonemic contrasts. Dev Psychol 2011;47(5):1448.
[45] Mayer RE. Cognitive Theory of Multimedia Learning. In: Mayer R, editor. The Cambridge Handbook of Multimedia Learning. Cambridge University Press; 2005. p. 31–48.
[46] Moreno R. Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. J Comput Assist Learn 2006;22(3):149–58.
[47] Reich SM, Yau JC, Warschauer M. Tablet-based ebooks for young children: what does the research say? J Dev Behav Pediatr 2016;37(7):585–91.
[48] Livingstone S, Blum-Ross A, Pavlick J, Ólafsson K. In the digital home, how do parents support their children and who supports them? Parenting for a Digital Future: Survey Report 1. 2018. Retrieved 13/08/20 from, https://www.lse.ac.uk/media-and-communications/assets/documents/research/preparing-for-a-digital-future/P4DF-Survey-Report-1-In-the-digital-home.pdf.
[49] Dias P, Brito R. Criteria for selecting apps: Debating the perceptions of young children, parents and industry stakeholders. Comput Educ 2021;165:104134.
[50] Vaala S, Ly A, Levine MH. Getting a Read on the App Stores: A Market Scan and Analysis of Children's Literacy Apps. Full Report. New York, NY: Joan Ganz Cooney Center at Sesame Workshop; 2015. p. 10023. Joan Ganz Cooney Center at Sesame Workshop. 1900 Broadway.
[51] Madigan S, McArthur BA, Anhorn C, Eirich R, Christakis DA. Associations Between Screen Use and Child Language Skills: a Systematic Review and Meta-analysis. JAMA Pediatr 2020.

[52] Taylor G, Monaghan P, Westermann G. Investigating the association between children's screen media exposure and vocabulary size in the UK. J Child Media 2018;12(1):51–65.

[53] Moon JH, Cho SY, Lim SM, Roh JH, Koh MS, Kim YJ, Nam E. Smart device usage in early childhood is differentially associated with fine motor and language development. Acta Paediatr 2019;108(5):903–10.

[54] van den Heuvel M, Ma J, Borkhoff CM, Koroshegyi C, Dai DW, Parkin PC, Birken CS. Mobile media device use is associated with expressive language delay in 18-month-old children. J Dev Behav Pediatr 2019;40(2):99.

[55] Snow CE. Mothers' speech to children learning For instance, one would expect such a floor if language. Child Dev 1972;43:549–65.

[56] Monaghan P, Mattock K. Integrating constraints for learning word–referent mappings. Cognition 2012;123(1):133–43.

[57] Rowe ML. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. Child Dev 2012;83(5):1762–74.

[58] Monaghan P, Christiansen MH. Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. J Child Lang 2010;37:545–64.

[59] McMurray B. Defusing the childhood vocabulary explosion. Science 2007;317 (5838):631.

[60] Phillips JR. Syntax and vocabulary of mothers' speech to young children: age and sex comparisons. Child Dev 1973:182–5.

[61] Soni N, Aloba A, Morga KS, Wisniewski PJ, Anthony L. A framework of touchscreen interaction design recommendations for children (TIDRC): Characterizing the gap between research evidence and design practice. In: Proceedings of the 18th ACM International Conference on Interaction Design and Children; 2019. p. 419–31.

[62] Brydges CR. Effect size guidelines, sample size calculations, and statistical power in gerontology. Innov Aging 2019;3(4):1–8.

[63] McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012; 22:276–82.

[64] Van Heuven WJ, Mandera P, Keuleers E, Brysbaert M. SUBTLEX-UK: A new and improved word frequency database for British English. Q J Exp Psychol 2014;67 (6):1176–90.

[65] Brysbaert M, Warriner AB, Kuperman V. Concreteness ratings for 40 thousand generally known English word lemmas. Behav Res Methods 2014;46(3):904–11.

[66] Kuperman V, Stadthagen-Gonzalez H, Brysbaert M. Age-of-acquisition ratings for 30,000 English words. Behav Res Methods 2012;44(4):978–90.

[67] Axelsson EL, Churchley K, Horst JS. The right thing at the right time: Why ostensive naming facilitates word learning. Front Psychol 2012;3:88.

[68] Chang YN, Monaghan P. Quantity and diversity of preliteracy language exposure both affect literacy development: Evidence from a computational model of reading. Sci Stud Read 2019;23(3):235–53.

[69] Hart B, Risly T. Meaningful differences. Baltimore, MD: P.H. Brookes; 1995.

[70] Huttenlocher J, Haight W, Bryk A, Seltzer M, Lyons T. Early vocabulary growth: relation to language input and gender. Dev Psychol 1991;27(2):236.

[71] Weisleder A, Fernald A. Talking to children matters: Early language experience strengthens processing and builds vocabulary. Psychol Sci 2013;24:2143–52.

[72] Cameron-Faulkner T, Noble C. A comparison of book text and child directed speech. First Lang 2013;33(3):268–79.

[73] Broekman FL, Piotrowski JT, Beentjes HW, Valkenburg PM. A parental perspective on apps for young children. Comput Hum Behav 2016;63:142–51.

[74] Rideout V, Robb MB. The common sense census: Media use by kids age zero to eight. San Francisco, CA: Common Sense Media; 2020.

[75] Aziz NAA, Sin NSM, Batmaz F, Stone R, Chung PWH. Selection of touch gestures for children's applications: repeated experiment to increase reliability. Int J Adv Comput Sci Applic 2014;5:97–102.