Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

Mia Tangen

The Efficacy of Educational Apps in Second Language Learning

A comprehensive qualitative review

Master's thesis in Language Studies with Teacher Education
Supervisor: Mila Dimitrova Vulchanova
Co-supervisor: Giosuè Baggio
November 2021

NTNU
Norwegian University of
Science and Technology

Mia Tangen

The Efficacy of Educational Apps in Second Language Learning

A comprehensive qualitative review

Master's thesis in Language Studies with Teacher Education
Supervisor: Mila Dimitrova Vulchanova
Co-supervisor: Giosuè Baggio
November 2021

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

NTNU
Norwegian University of
Science and Technology

Abstract

The aim of this review is to investigate the efficacy of the use of apps in second language learning, specifically English as a second language. Theoretical background about how children learn, and how exactly they learn a second language is included as it is crucial information necessary for the analysis of how to design, plan, and implement digital technologies such as apps to facilitate second language learning, according to children's needs.

The material of this comprehensive qualitative review are relevant articles on the topic that have been found by a method of a systematic review. The studies were language related and involved interaction with an educational app. Both studies involving first and second language learning were included due to a limited number of studies concerning educational apps intended for second language learning.

The results indicated that educational apps can be effective for learning language, especially for the development of emergent literacy skills and vocabulary. Moreover, the discussion of these results suggests that these apps can be used effectively for second language learning as well as language learning in general. However, certain conditions must be met.

In conclusion, educational apps developed with children's learning in focus can potentially contribute to effective second language learning. Nevertheless, the teacher or parent must always pay attention to children's learning outcomes from these apps as children are quickly developing, and learning needs and outcomes will change accordingly.

Preface

The inspiration for this master's thesis was the discovery of "iPad schools". I was looking for teaching positions in a certain area when I discovered that all schools in that municipality used iPads in their teachings. This was relatively new to me, and I have limited knowledge of how such schools operate. Thus, I wanted to investigate whether the use of iPads and apps are efficient in their purpose. As an English teacher, I had special focus on language learning, especially second language learning.

Through cooperation with my supervisors, Mila Dimitrova Vulchanova and Giosuè Baggio, we decided that a comprehensive qualitative review would be best suited to investigate the topic considering the situation of the Covid-19 pandemic. However, the last couple of months have been very insightful as I have been working as a 6th grade English teacher in an "iPad school". Writing this master's thesis simultaneously as I have been working has had serious consequences for my own use of the iPad when teaching. The process of reviewing all articles on the subject has been very instructive, and I have become very aware of the tendency we have of adapting to apps, rather than the other way around. We want the apps to be educational so dearly that we give them educational goals without them actually being designed with that intention.

Finally, I would like to give a huge 'thank you' to my principal and secondary supervisors, Mila Dimitrova Vulchanova and Giosuè Baggio, for the good and constructive feedback, as well as the support in all stages of the writing process - especially in the final stages. It was very inspiring and helpful. A special thanks to Karoline Sønvisen Vargdal and Anja Tangen for proofreading the thesis.

Mia Tangen
Bærum, November 2021

Table of content

1 INTRODUCTION 9

1.1 Educational apps 10

2 THEORETICAL BACKGROUNDS 11

2.1 The Displacement Hypothesis 11
2.2 The Four Pillars of Learning 11
2.2.1 Active Learning 11
2.2.2 Engaged Learning 12
2.2.3 Meaningful Learning 13
2.2.4 Socially Interactive Learning 13
2.3 Second Language Learning 14

3 METHODS 16

3.1 Data Collection 16
3.2 Study Selection 17
3.3 Data Extraction and Data Analysis 17
3.4 Quality assessment 18

4 RESULTS 19

4.1 Description of Studies 19
4.1.1 Exclusion and Inclusion of Studies 23
4.1.2 Narrative Synthesis of Studies 23

5 DISCUSSION 26

5.1 Main findings 26
5.1.1 Vocabulary Learning 26
5.1.2 Literacy Skills 32
5.1.3 Engagement 36
5.2 What Are the Costs of Learning? 37
5.2.1 Apps Replacing Teachers? 37
5.2.2 Apps Replacing Children? 37
5.2.3 The Displacement Hypothesis 39
5.3 Limitations 40

6 CONCLUSION AND IMPLICATIONS 41

6.1 Professional relevance 41

WORKS CITED 42

1 Introduction

The last decade has been characterized by smartphones and apps. According to Statistics Norway (SSB, 2021), 96% of the Norwegian population aged 9-79 is in possession of a smartphone. Arguably, having a smartphone is quite common, independent on income (Arnold, Chary, Gair, Helm, Herman, Kang, & Lokhandwala, 2021). Hirsh-Pasek, Zosh, Golinkoff, Gray, Robb, & Kaufman (2015) state that there has been an app explosion after the release of the iPad in 2010, and people are still producing an endless number of apps with varying functions and designs. While some apps are designed for taking pictures, calling, or texting, a lot of apps have several functions embedded into one. According to function, apps can be classified as picture editing, workout apps or apps for children, for instance. Of particular interest for the current project is the category labelled 'educational' apps.

The iPad, as other modern digital technology, is multimodal by nature. "It allows for multisensory interaction and provides rich input in the form of visual, auditory, and haptic stimuli" (Vulchanova, Baggio, Cangelosi, & Smith, 2017, p. 3), and thus provide opportunities to make high-quality educational apps. However, only a limited number of apps are designed in a way that is considerate of how children actually learn (Hirsh-Pasek et al., 2015). The term and category 'educational' suggests that these apps could be used for teaching and learning, thus they could be implemented in school. Some of these apps have found their way into the classroom, via tablets or other devices, and are purposely used as tools for education and learning. Interestingly, there is limited research to support this use of tablets and apps (Hirsh-Pasek et al., 2015). Furthermore, even though the apps have been categorized as 'educational', these apps are being produced and made available for the masses at such a rate that it is impossible to evaluate every app as to whether it is in fact educational (Hirsh-Pasek et al., 2015). It is also possible that these apps, even those that are only remotely relevant for education, are changing our notion of an 'educational app'. We are adapting to the available apps rather than the other way around. Teachers make use of apps for educational purposes, even though they are not designed with that intent. Hence, how are these apps educational, and what kind of learning do they promote? Designers of these educational, children-focused apps are influenced by current trends in the field, developers own technological interactions and experiences, and their intuition about how learning happens. However, this understanding of learning and education is often full of misconceptions (Hirsh-Pasek et al., 2015). Thus, the question becomes: how do people learn, and are these apps efficient for that purpose?

This study will focus on apps and second language (L2) learning, because "there exist many forms of digital technologies for […] tablet computers that support second language learning, although there is little evidence about their efficacy" (Vogt, de Haas, de Jong, Baxter, & Krahmer, 2017, p. 1). While there is a difference between learning a first language (L1) and an L2, learning a language demands basic literacy skills. Apps that promote such literacy skills would naturally be relevant for language learning. Moreover, certain apps have more specific focus, such as vocabulary learning. Other apps focus on engagement to be effective for learning. Yet, there is a danger in using apps that are not efficient for the purposes of teaching skills and promoting language skill development.

This review will be limited to educational apps for children in preschool, elementary school, and middle school. There are several reasons for this. First, according to Arnold et al. (2021, p. 3) "we know far too little about apps that aim to teach children". There is still a lot to be learnt about educational apps for children. Additionally, the use of apps seems to be effective from a young age and may also have a potential long-term impact (Arnold et al., 2021). Educational apps could be considered a practical tool for "fostering academic success and narrowing the SES opportunity gap" (Arnold et al., 2012, p. 1). Importantly, the age gap from preschool to middle school is large, and it is not expected that children in preschool learn the same way as children in middle school. This age gap makes for an interesting comparison of the children. Furthermore, when apps are being used in the classroom, they may perform some of the functions that the teacher would normally do – providing feedback for instance - but the question remains whether and how efficient the different apps are for that purpose. It is also possible to argue that apps also perform some of the functions that kids used to perform, for instance preparing a visual context of learning by drawing which is now provided by the app.

The aim of this review is to investigate the efficacy of using apps for language learning, with extra focus on learning English as a second language. Thus, it is important not only to understand how children learn, but how exactly they learn a second language and how this knowledge can be implemented and used to design effective learning apps. In other words, this review will also investigate how to design, plan, and implement digital technologies such as apps to facilitate second language learning, according to children's needs. The remaining question is whether a good design, according to available knowledge on how children learn, result in effective educational apps that succeed in their purpose.

1.1 Educational apps

It is necessary to clarify what is meant by the term 'educational'. The term is fluid in the context of apps. Available apps, even those only remotely relevant for education, are changing our notion of 'educational apps'. Interestingly, we adapt to these apps, not the other way around: Many apps are not designed with educational intentions, still, we are using some of these apps as if they were indeed educational.

It is important to be aware of the fluid use of the term 'educational'. It suggests a risk of losing actual educational apps, as developers find that we can make anything educational. Although this can be seen as a tribute to good teachers and parents, it will potentially put more responsibility on them if children are to interact with such apps. The more we use non-educational apps in an educational way, the less educational app developers will spend effort and resources on making them truly educational. To discuss this potential issue, non-educational apps used in an educational way that support a learning goal was included in this review. Consequently, in this thesis, educational apps are apps that can be used to support a learning goal, be it learning new vocabulary, pronunciation, writing, and more. Having a learning goal is suggested to be an important condition for effective learning (Hirsh-Pasek et al., 2015, p. 4).

2 Theoretical backgrounds

2.1 The Displacement Hypothesis

From the 1950s, television has been an undeniable part of everyday life, along with the computers from the 1980s and 1990s. These technological developments led to concerns about screen time and the effects of screen time on children's development. The drastic technological advancements we have witnessed the last 15 years have not made the concerns less prominent. According to Hassinger-Das, Brennan, Dore, Michnick Golinkoff and Hirsh-Pasek (2020), smartphones and tablets have completely changed the landscape of screen time. In the 1950s, screen time was limited to a specific room in the house and to set airing times. Today, children and adults can look at a screen wherever and whenever.

In the early phases of technological development, *the displacement hypothesis* was developed (Hassinger-Das et al., 2020). The hypothesis claims that television displaces other, much more enriching activities, such as reading books and socializing. While studies at the time did not find any evidence for such displacement, the displacement hypothesis has once again gained prominence due to the increased use of media the last 15 years: "We now have screens that we carry everywhere, and they may be more likely to displace parent-child interaction or simply children's time being bored and finding ways to entertain themselves" (Hassinger-Das et al., 2020, p. 3). Findings, however, are mixed regarding the negative effects of screen time. Interestingly, Guernsey and Levine (2016 in Hassinger-Das et al., 2020) suggest that the problem is not digital media per se but the way it is being used. There is a mismatch between the findings of child development researchers and new digital medias of app developers (Hassinger-Das et al., 2020): While developers prioritize marketing, researchers prioritize quality development (Hiniker et al., 2019 in Hassinger-Das et al., 2020).

2.2 The Four Pillars of Learning

A new field of study, the Science of Learning, has knitted together subfields of psychology, computer science, linguistics, animal behavior, neurobiology, machine learning, brain imaging, and other areas. They ask not only *what* children should be learning, but also *how* they should be learning to better learn necessary strategies to cope in a 21st century world (Hirsh-Pasek et al., 2015). The Science of Learning further examines how this knowledge can be used in classroom practices as well as at home, and how to incorporate digital media (Hassinger-Das et al., 2020). Importantly, the Science of Learning has introduced four pillars of learning which will be further explained in the following sections. This is not to say that learning cannot happen if we exclude on or more of these pillars, however, literature suggest that these pillars are important conditions for more effective learning (Hirsh-Pasek et al., 2015).

2.2.1 Active Learning

Children learn best when they are actively involved in their own learning (Hirsh-Pasek et al., 2015). When it comes to apps, it is necessary to draw a distinction between *physically* active and *mentally* active. Every app demands some physical activity, like swiping and tapping on the touch screen. However, this does not necessarily make children mentally active. Hirsh-Pasek et al. (2015) argue that children must be 'minds-on' to qualify as active learners in accordance with the active learning pillar. Minds-on apps require thinking and intellectual manipulation from the user. It is not enough to

simply swipe or tap. Tapping to make something disappear simply for the fun of it is 'minds-off', while tapping to make a word or a picture disappears because it does not belong, is 'minds-on'.

The level of 'control' is also an important factor for active learning. There must be an *appropriate* level of control depending on factors such as age and experience. Children must be allowed to proceed at their own pace to sustain interest. This is one of the many advantages with tablets contra computers: touch-screen apps are more controllable by children of almost any age – depending on design – compared to a computer mouse or a keyboard (Hirsh-Pasek et al., 2015).

2.2.2 Engaged Learning

Children learn best when they are engaged in the learning material. Engagement is crucial for learning as it predicates an individual's ability to stay 'on task' (Hirsh-Pasek et al., 2015). If children get distracted, they cannot stay on task and thus lose their engagement, which results in lower learning outcome. Hence, engagement and distraction are closely linked.

Children's engagement during learning can be disrupted in several ways. Tare, Chiong, Ganea and DeLoache (2010 in Hirsh-Pasek et al., 2015) found that children were distracted when reading a pop-up book. "Even when extra features were designed to call attention to a specific learning goal (e.g., letters in an alphabet book), children learned best when they were able to stay on task using a simpler version of the book" (Chiong & DeLoache, 2012 in Hirsh-Pasek et al., 2015, p. 11). Other distracting elements can be instrumental background music (Barr, Shuck, Salerno, Atkinson & Linebarger, 2010 in Hirsh-Pasek et al., 2015) and irrelevant entertaining content, known as seductive details (Garner, Brown, Sanders & Menke, 1992 in Hirsh-Pasek et al., 2015).

The younger the children the more important it is to avoid distractions. Younger children do not have the ability to inhibit attention to extraneous information. However, the danger of distraction is apparent throughout adulthood as well. Research suggests that multitasking, such as texting during class, results in decreased performance and learning outcomes (Hirsh-Pasek et al., 2015). It is necessary to design an app where the environment can be seen as a helping hand rather than a distracting obstacle. It is all about hitting the 'sweet spot' between being accessible and challenging, to promote engagement and help the learner stay on task and reduce distractions. The content cannot be too familiar or too challenging. This is known as the *traveling lens model* of viewing (Wright & Huston, 1983 in Hirsh-Pasek et al., 2015).

Contingent interactions, extrinsic motivation and feedback, and intrinsic motivation are deeply connected to engagement. Contingent interactions are immediate responses to a child's swipe or touch, which make the children feel in control and help maintain their focus, and thus the interaction continues (Hirsh-Pasek et al., 2015). Extrinsic motivation - like avoiding punishment or receiving a reward – and feedback needs to be meaningful to foster engagement and learning. According to Kolak, Norgate, Monaghan, & Taylor (2020), feedback is important and should be meaningful, specific, timely, and structured in order to engage. "By carefully structuring the feedback as well as allowing progressive access to content […], apps can focus children's attention on the app experience and extend engagement for a long time" (Hirsh-Pasek et al., 2015, p. 12). Importantly,

Hirsh-Pasek et al. (2015) explain that it is crucial to praise effort over intelligence, since praising children's intelligence has shown to lead them to avoid some risks of learning in fear of appearing stupid. For instance, praising intelligence may lead them to stop asking questions in fear of losing face in front of others. In contrast, praising children for their efforts makes them understand that hard work pays off and that learning is not something that happens in an instant.

Finally, intrinsic motivation is important for children's long-term development. Intrinsic motivation is about awaken sincere interest for a subject. This kind of motivation is user-driven and deeply engaging for both children and adults. Intrinsic motivation may give the sense of 'flow', the experience where a person loses his or hers sense of time while being engaged in an activity (Hirsh-Pasek et al., 2015). Kolak, Norgate, Monaghan, & Taylor (2020) state that the structure of the activity is important for intrinsic motivation and engagement: "Apps which give the opportunity for exploratory use alongside structured activities, might increase children's intrinsic motivation and engagement" (Kolak et al., 2020, p. 6).

2.2.3 Meaningful Learning

Children learn best when the content is presented in a way that connects to existing knowledge the children might have and relates to their lived experiences. According to Hirsh-Pasek et al. (2015, p. 13), "meaningful learning takes many forms, including learning with a purpose, learning new material that is personally relevant, and linking new learning to preexisting knowledge". It is very important to distinguish meaningful learning from rote learning, where new information does not link to existing knowledge, and thus often fades from memory (Hirsh-Pasek et al., 2015).

A knowledge base where more meaningful learning is established, needs to be built. Acquisition of such a base often happens due to drill and practice. Existing work on middle school children suggest that this is exactly where apps will shine. Furthermore, promoting meaningful learning might sometimes depend on a context that stimulate greater motivation (Hirsh-Pasek et al., 2015). For example, a child that is interested in football might be more motivated to learn fractions by dividing game time among players than simply solving problems on a worksheet. When applying meaningful learning to apps it is important to consider the quality and quantity of connections between the app experience and the preexisting knowledge of a child.

Some apps are considered more meaningful than others. For instance, apps that require children to engage and solve problems as part of a larger game narrative may be more successful than other game apps where the challenges are not integrated into the game's narrative or larger context (Hirsh-Pasek et al., 2015). An entertaining narrative itself can be beneficial for children's learning (Kolak et al., 2020).

2.2.4 Socially Interactive Learning

Children learn best when they interact with other people, such as educators, peers, and caregivers. However, having a social partner alone is not enough. The interaction must be of high enough quality and cannot detract from the learning situation (Hirsh-Pasek et al., 2015). Social interaction allows young children to observe and imitate others, which may promote learning in how events typically unfold. Furthermore, social interaction impacts children's understanding in school. For instance, Johnson, Maruyama, Johnson,

Nelson, and Skon (1981 in Hirsh-Pasek et al., 2015) found that collaborative learning seems to be beneficial for critical thinking skills.

Hirsh-Pasek et al. (2015) has proposed three ways in which app design can incorporate the potential educational benefits of social interactions: First, there should be more real life, face-to-face interactions with multiple children around (and further away from) the screen. The second proposal is to socially engage children in mediated interactions through technologies such as screen-sharing apps or video-apps like FaceTime. And third, the design should develop a parasocial relation between the child and the on-screen characters. This would imply a one-sided relationship where the child becomes attached and invested in the on-screen character, much like the relationship many would have with celebrities. This can be done by designing the characters to be more realistic for a two-way interaction. For instance, Dora the Explorer from the TV series could potentially be a great character and would have a larger potential with apps than TV where the children can respond to her questions.

2.3 Second Language Learning

Krashen's Input Hypothesis says that input needs to be comprehensible to be learnt, but comprehensible input alone is not sufficient for L2 learning (Phakiti & Plonsky, 2018). Ideally, the input should be comprehensible, but still include some new input to move on to a higher level. There is also another related second language acquisition (SLA) theory, *Pienemann's Processibility Theory*, which states that "L2 learners can only produce and comprehend language that they can currently handle" (Phakiti & Plonsky, 2018, p. 225). Furthermore, the *Cognitive Load Theory* may help explain whether an L2 learner can process language, or not. Humans have limited information processing due to limited working memory and attention. This working memory can easily be overloaded. Thus, to maximize L2 learning, a set of schemas need to be generated in the long-term memory to help learners learn new content (Phakiti & Plonsky, 2018). For instance, activating prior knowledge. This automation of retrieval of information from long-term memory may help reduce cognitive load in the working memory, and such help the L2 learner in language processing.

Phakiti and Plonsky (2018) state that language exposure and language interaction are important factors for L2 learning, however, it depends on the quality of the exposure and interaction. Other critical factors are motivation and self-regulation (Phakiti & Plonsky, 2018). Motivation has a strong impact on short- and long-term L2 attainment and can be affected by teacher intervention. Furthermore, self-regulation and the use of language learning strategies may also be improved by teacher intervention.

Explicit grammar instruction has also proven effective in providing learners with an understanding of the L2 structure. Nonetheless, Phakiti and Plonsky (2018) state that explicit grammar instruction alone is not sufficient for L2 learning. Explicit instruction must be accompanied by opportunities to use the L2 in a communicative setting with a purpose and a goal.

The Zone of Proximal Development (ZDP), proposed by Lev Vygotsky, claims that it is important with social interaction and scaffolding from a teacher. "It is essential that teachers understand the critical roles of the many social factors at play in the course of L2 development, both inside and outside of the classroom" (Phakiti & Plonsky, 2018, p.

233). Children alone can only do so much, but they need help from someone more knowledgeable than themselves in order to progress. According to Vogt et al. (2017, p. 5) digital technologies designed to teach children must be "sufficiently challenging and varied so that the child has a target to learn from, but at the same time interactions should not be too difficult" as that may be frustrating to the child and may lead to loss of interest.

The field of SLA has discussed how and if the native language plays a role in the learning of L2, especially in areas such as vocabulary, grammar, and pragmatics (Phakiti & Plonsky, 2018). Ellis (2015 in Phakiti & Plonsky, 2018) suggests that there are five different levels at which L1 influence the L2: linguistically, psycholinguistically, contextually, developmentally, and individually. At the linguistic level, L1 may influence L2 at both the lexical level and at structure level. The psycholinguistic level "is related to L2 learners' self-perceptions about L1 influences and what they believe can or cannot be transferred to L2" (Phakiti & Plonsky, 2018, p. 221). Furthermore, at the contextual level, L2 learners are influenced by the context of L2 learning: "L1 transfer may be positive or negative depending on variables such as time pressure and whether the context is formal or informal" (Phakiti & Plonsky, 2018, p. 221). Developmentally, L1 transfer is related to developmental and universal features, such as simplification and overgeneralizations (Phakiti & Plonsky, 2018). Lastly, L1 transfer is also deeply connected to individual differences (Phakiti & Plonksy, 2018).

Vogt et al. (2017) wrote an article about social robots for second language learning. They proposed different design features that would develop a child-friendly social robot, effective for second language learning for children. These features are also relevant for tablets and apps, as they are simply underdeveloped robots without the human-like features. One of the features they proposed was that of scaffolding. By this it is understood that the app should be more knowledgeable than the child using it in order to help him or her progress. Furthermore, the app should "adapt to the language proficiency level of the individual child" (Vulchanova et al., 2017, p. 4). In other words, the apps should be designed to personalize the learning experience and individually adapt to children's learning needs. When it comes to design, the app design should be "simple and consistent, style of letters should be clear, and the arrangement of operating buttons should be appropriate" (Kolak et al., 2020, p. 7). Furthermore, the app should be easy to use, as well as being response to touch screen interactions.

Another important feature, proposed by Vogt et al. (2017), is that of contingently response. The interactions with the app should be contingent and multimodal, meaning that the app should be responsive to the child's actions using different approaches. The apps should also provide effective feedback, according to Vogt et al. (2017), as feedback is known to help language learning. Finally, a well-designed app should establish joint attention: "Joint attention, where interlocutor attend on the same referent, is a form of social interaction that has been shown to support children's language learning" (Tomasello & Farrar, 1986 in Vogt et al., 2017, p. 4).

3 Methods

3.1 Data Collection

An electronic database search was conducted in seven databases in the fields of education psychology, biotechnology, and other multidisciplinary databases. The seven databases first included were *Google scholar*, *ScienceDirect*, *PsycINFO*, *Web of Science*, *PubMed*, *Scopus*, and *ERIC*.

The following keywords were used: 'apps', 'English', 'second language', 'child', 'evidence', and 'educational'. I used the following search phrase: apps AND English AND second language AND child AND evidence AND educational. These keywords led to unique search results in the different databases. Due to an overwhelming result in *Google Scholar*, with over 14.000 results, and not a single finding in *ERIC* and *PubMed*, these three databases were excluded. Thereby, the data was collected from a total of four databases: *ScienceDirect*, *PsycINFO*, *Web of Science*, and *Scopus*.

Articles published before 2014 and written in other languages than English were also excluded from the results. 2014 was set as cut-off year, because, while the iPad was introduced as early as in 2010, it only took three years before they had developed so much that they released the first iPad Air in 2013. The following year, in 2014, they had once again updated their iPads and released the iPad Air 2 (One World Rental, 2019). The rapid technological development reflects the need for updated and relevant research on the use of iPads and other tablets.

The first database search was conducted the 22nd of April 2021. However, due to the first attempt not being planned well enough, I had to conduct another search for relevant articles and thus restart the entire process. I had not kept any records on how many results each database got, and I did not have any fixed criteria for inclusion or exclusion of records. Before conducting the second search on October 7th, 2021, I made a list of inclusion criteria presented in table 1 in chapter 3.2. Results that did not fit these criteria were excluded from the research.

An initial screening of article titles was conducted before the screening of article abstracts. The initial screening of titles involved looking for words in the title that suggested the article was relevant. Words indicating that the article had something to do with tablets or mobiles devices, apps, education, or language learning, meant the article was included for abstract screening. Moreover, the screening of abstracts looked for inclusion or exclusion criteria presented in table 1. If all the criteria were present, the article was selected for full-text assessment. However, if one or more of the exclusion criteria were present, the article was excluded from full-text assessment. This is to say that if all other criteria were present, but the study sample included college students and not children under the age of 14, the study was excluded from further assessment. Furthermore, if the abstract did not say anything specifically about the study that would either include or exclude it, it was included for further assessment. Such articles would be either included or excluded in the full-text assessment.

Additionally, the reference sections of included articles and other review articles, as well as suggested articles, were examined to identify other studies that met the inclusion criteria. While review articles are equally valuable as other research articles, I aimed to include only primary literature in this review, and thus other review articles were not

included. Review articles were only accepted to find relevant articles in the reference section and to provide useful context for the research paper.

3.2 Study Selection

The search initially identified 513 unique records. Additionally, 12 records from the reference lists and suggested literature were included in the screening sample. After initial screening, 30 records were subject to full text review. In total, 13 articles, describing 15 studies, were included in this study**.** The inclusion and exclusion criteria are presented in table 1.

Category	**Inclusion Criteria**
Population	Children in preschool and elementary school, aged between 3-13 years. Studies conducted in any country were included, but the article had to be published in English.
Intervention	Intervention involves use of educational apps. 'Educational' suggests that the apps have been categorized as educational either in Apple's App Store or Android's Google Play Store, or that they have been used with an educational purpose. Studies were excluded if they involved only interventions with apps belonging to other categories, without any learning intentions.
Comparison	Any comparison groups.
Outcome	Language related academic achievement, task-specific learning (e.g., learning new vocabulary), socio-emotional skills, literacy skills, and engagement. Studies were included when the learning outcome was language related.
Study design	Randomized controlled design, case studies, quasi-experimental designs, and observational studies. Descriptive studies and reviews were excluded.

Table 1

3.3 Data Extraction and Data Analysis

The following information was extracted from each study: design, study sample, length of intervention, name of app(s), control condition, what was examined and whether it was an L1 or L2, and results.

Due to the number of articles (n = 13), which would not have been sufficient to calculate an effect size (and the clinical, methodological, and statistical heterogeneity of the included studies in terms of design, outcomes measured, and comparison group used), formal meta-analytic strategies were not used to summarize overall effects. Instead, the data will be described and summarized as a narrative synthesis.

3.4 Quality assessment

All articles were peer-reviewed journal articles, and all journals were controlled for not being predatory journals. Furthermore, the quality of the data selected for this study was assessed in several stages of the process. The initial search was limited to a set of pre-decided keywords. The choice of keywords is important as it will decide which articles and studies will be available for screening and analyzing. The chosen keywords for this study would naturally limit the search results, and it might have excluded some relevant articles due to the choice of terms. Moreover, the initial screening of titles might also have excluded a few relevant articles due to a misleading title, however, this method helped limit the search result a great deal, and the screening of abstract could thereby be done in further detail without spending too much time on this process.

4 Results

Figure 1 is a graphical description of the steps of this data extraction process, presented in a flowchart.

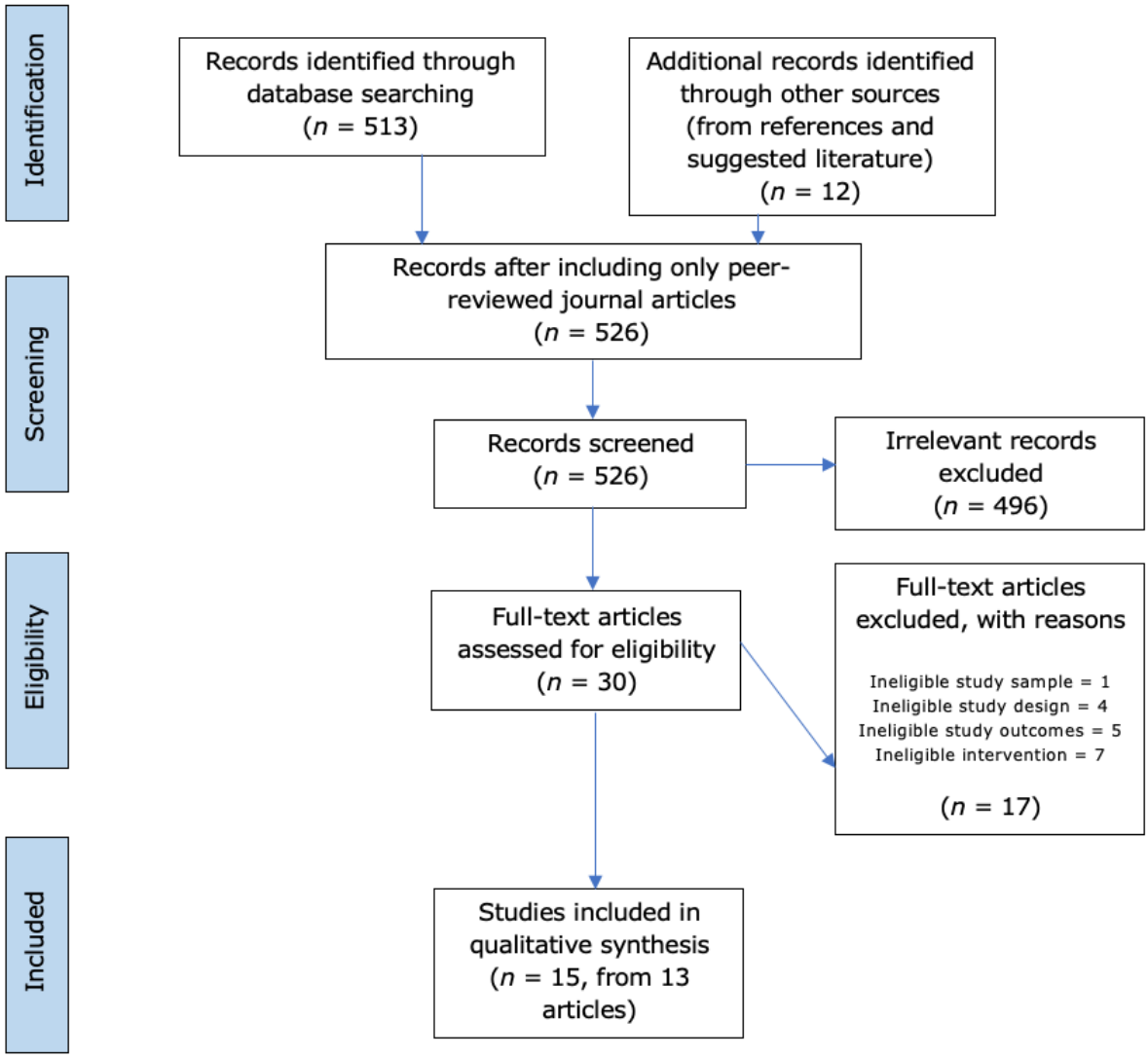


Figure 1

4.1 Description of Studies

Description of studies is presented in table 2. From 13 articles included in qualitative synthesis, there were 15 studies in total. The total numbers of participants across all 15 studies were 1084. Out of these studies, 10 focused on children aged 5 or younger. The remaining studies focused on children aged 5 or older. Three of the studies focused on children between the age of 5 and 8, while two studies focused on middle school children aged 12-13 years old.

Author (Year)	Study design	Study sample	Length of inter-vention	App	Control condition	What was examined (L1/L2)	Result of intervention
Arnold et al. (2021)	RCT	49 children, 4-5 y/o	10 weeks	*Khan Kids app*	Control group interacting with apps not targeting emergent literacy skills: *Bord* and *MiniPiano*	Emergent literacy skills: Phonological awareness, print knowledge, and vocabulary knowledge. (English, L1)	Increased emergent literacy skills (phonological awareness) compared to control condition
Patel et al. (2021)	RCT	136 children, 5-7 y/o	8 weeks: 20 minutes sessions, 5 days per week.	*GraphoLearn* (GL) English Rime	A control group playing a math game	Letter sounds, word recognition, transfer of learning to out-of-game context (English, L2)	Effectively learning of critical sub-skills for reading: faster development on in-game measures of letter sound knowledge, word recognition, and rime unit recognition. No transfer of learning to out-of-game contexts.
Rowe et al. (2021)	RCT	76 children, 3-4 y/o	3 weeks	*Animal Antics, Story Mixer, Photo Play*	Apps-only group (Intervention: apps-with-info group)	Syntax, interactions, vocabulary (English, L1)	No main effect between intervention group and comparison group. Most measures increased over time in both groups: syntactic complexity and vocabulary improved. A significant positive association between the number of videos accessed and the increased number of parent-child conversations and parent vocabulary use. Children who engaged more apps with in-between visits had greater MLU (length of morphemes in utterances) at visit 2.
Dore et al. (2019)	**Study 1:** RCT	57 children, 4 y/o	1 session, 10-12 minutes	*The Quicksand Rescue Mission*	No-interaction control group	Receptive vocabulary (English, L1)	Increased receptive vocabulary compared to control group
	Study 2: Experimental	33 children, 3-4 y/o	4 weeks (once a week)	*The Quicksand Rescue Mission* and *the Golden Eggs Mission*	Within-subjects design: No exposure control words.	Receptive and expressive vocabulary (English, L1)	Larger gain for target words than no-exposure words in both receptive and expressive test.
O'Brien et al. (2019)	RCT	148 children, 6-8 y/o	2x7 weeks (14 weeks in total): 10 minutes per day, 5 days	*SeeWord Reading, GraphoLearn Phoneme, GraphoLearn Rime,* In-house developed	Three intervention groups: phoneme level, rime level, and word level.	Reading and decoding accuracy, fluency, and spelling (English, L1)	The phoneme-level intervention yielded better growth in decoding accuracy than the rime-level intervention, while word-level did not differ. Reading and spelling

			per week.	app designed to develop reading skills.	The word-level group serves as comparison between lexical level compared with the sub-lexical level.		outcomes increased in all groups. Different results at pretest were significant to different results at posttest, depending on the app.
Neumann (2018)	RCT	48 children, 2-5 y/o	9 weeks: 30 minutes sessions per week.	*Endless Alphabet, Letter School, Draw Buddy*	Classroom instruction as usual, without iPads	Emergent literacy skills: Letter name, sound knowledge, and name writing (English, L1)	Intervention group performed significantly better than the control group on log letter name knowledge, log letter sound knowledge, and name writing. No significant differences were found for letter writing and numeral name knowledge.
Teepe et al., (2017)	RCT	71 children, 3 y/o	2x10 minutes session over 2 weeks	*Jeffy's Journey* (TES)	No-treatment control group	Expressive and receptive vocabulary (Dutch, L1)	Intervention group learned more words compared to the control group. Intervention and control group made similar growth on receptive vocabulary knowledge.
Cavus & Ibrahim (2017)	RCT	37 children, 12-13 y/o	4 weeks: 0.5 hours daily	*Near East University Children's Story Teller* (NEU-CST)	Traditional learning: Paper book	Vocabulary, comprehension, pronunciation, and listening (English, L2)	Intervention group outperformed the control group on all skills.
Patchan & Puranik (2016)	RCT	46 children, 3-5 y/o	8 weeks: 20 minutes session, 3 times per week.	*Writing Wizard*	Control group using pen and paper	Emergent literacy skills: Alphabet knowledge (English, L1)	Significant difference between the condition for posttest letter writing: Children in the iPad-finger condition wrote more letters correctly than children in the iPad-stylus condition and the paper-pencil condition. No significant differences between conditions for letter naming.
Al-Bogami & Elyas (2020)	Observational	20 female middle school students, 12-13 y/o	5 weeks: 10 sessions, 40 minutes each	*Quizlet, iBook, Popplet Lite, Pollevery-where, Pixton Comic Maker*	No control group. Comparisons of the different apps.	Helpfulness, ease of use, and engagement (English, L2)	Students found that iPads were motivating and boosted their engagement and learning in EFL/ESL reading classes. Observations found that gamification of apps were more engaging. Different apps engage at different levels.
Christ et al. (2019)	Observational	53 children, 5-6 y/o	1 year. Six sessions, ranging from 7-	*Frog, But not Hippo, Gustav the Goldfish, Troop, Fox*	No control group. Comparisons of the	Relations between reader characteristics, app book characteristics,	Mode selection, sequence, hotspot use, layered use of modalities, and use of comprehension

			32 mintues.	*in Socks, Penguin*	different app books	transactions between readers, and comprehension outcomes (English, L1)	monitoring strategies were linked to inference/critical reading and thinking, vocabulary, and prompted/unprompted retelling.
Russo-Johnson et al. (2017)	**Study 1:** Experimental	77 children, 2-5 y/o	6 1-minute trials	In-house word learning app.	4/5-year-olds	Self-regulation, word learning, and interactions with touchscreen apps (English, L1)	4- and 5-year-olds performed better on word learning and self-regulation.
	Study 2: Experimental	170 children, 2-4 y/o	6 1-minute trials	In-house word learning app (Not the same app as in study 1)	Non-interactive video of gameplay in word learning app	Self-regulation, word learning, and interactions with touchscreen apps (English, L1)	No overall main effects for which behavior (dragging, tapping, or watching) led to best word learning. Only the older children were proficient at transferring novel object labels to actual 3D objects.
Vatalaro et al. (2018)	Quasi-experimental	63 children, 3-5 y/o	8 weeks	<u>Scaffolding-like (SL) apps</u>: *Endless Alphabet, Noodle Words HD-Action Set 1, Goodnight ABC, ABC Go* <u>Open-ended (OE) apps:</u> *Beck and Bo, Draw and Tell HD, Don't Let the Pigeon Run This App! Alien Assignment*	<u>Head Start-chosen apps:</u> *Letter School, Gazzili Science, Yumiloo Rainbow Power, Faces iMake ABC, Counting Bear*	Receptive and expressive vocabulary (English L1)	<u>PPVT-4:</u> Children in the I-SL condition performed statistically significantly higher than children in the I-OE condition. No significant differences between the other groups. No statistically significant interaction effect. <u>EVT-2:</u> No statistically significant within-between subjects interaction effect between scores and group status. Large increase in EVT scores across time for the I-SL group and a slight increase for the I-OE group. <u>iPad Receptive Vocabulary:</u> The I-SL group performed significantly higher than the comparison groups. No statistically significant differences between the other groups. A significant within-between subjects interaction effect between scores and group status: larger increase across time for the I-SL and the I-OE groups.

							iPad Expressive Vocabulary Correct Word: No statistically significant between-subjects main effect between the groups. A statistically significant within-subjects interaction effect between scores and group status: larger increase across time for the I-SL and I-OE groups. iPad Expressive Vocabulary Correct Word or Description: The I-SL and I-OE groups had the largest increases across time. Slight increases in the comparison groups.

Table 2

4.1.1 Exclusion and Inclusion of Studies

The studies in this comprehensive qualitative review were included if the sample mean age was between three and 13 years old, and the study design was randomized, experimental, quasi-experimental, or observational. The intervention in all studies involved playing or interacting with an app with an educational intention. Studies where literacy skills or more specific language related skills were measured, were included. The target language did not necessarily need to be English as a first- or second language (L1 or L2), but the language learning in one way or the other had to be one of the outcomes.

Articles were included or excluded based on the criteria listed in table 1. Importantly, all studies had to include applications used with an educational purpose, and more specifically, the educational purpose had to be related to language learning. Furthermore, five of the 13 articles included in this comprehensive review were suggested literature, or literature taken from the reference list of another systematic review on a similar topic. One article was suggested to me by my supervisor, while the remaining four articles were found in a systematic review written by Griffith, Hagan, Heymann, Heflin, and Bagner (2020). They examined whether children under the age of six could learn from interactive apps by reviewing 35 articles on the subject. After full examination, four of these were also relevant for this study and was therefore included. I chose to include these articles due to a limited number of relevant articles in my own database search.

4.1.2 Narrative Synthesis of Studies

The 15 studies reviewed here had one of four different study designs: randomized controlled trial (RCT), experimental, quasi-experimental, or observational. Of these, nine were randomized controlled trial (RCT) studies, three were other experimental studies, two were observational studies, and finally, one was quasi-experimental. All studies examined some aspect of language learning, whether it was emergent literacy skills, more specific language skills, or important conditions for language learning (e.g., engagement).

The studies were either conducted at home, in school, or in labs. Two of the studies were unclear of exactly where the intervention happened: Teepe, Molenaar, & Verhoeven (2017) conducted testing in unspecified quiet rooms while Patchan & Puranic (2016) conducted their intervention in unspecified "small groups". Furthermore, eight of the studies were conducted at school. Half of the at-school interventions happened in the classroom, while the other half were done in separate rooms at the school. It is worth mentioning that one of the studies involved classroom intervention, but the students involved in the study had to face away from the rest of the class and wear headphones while intervening with the app. Moreover, only two studies were conducted in a lab. The remaining two studies involved intervention with apps at home.

The duration of intervention varied a great deal across all studies. Three studies involved eight weeks of intervention with apps, while two studies intervened with educational apps for four weeks. Other studies involved three-, five-, nine-, and 10-weeks of intervention. One of the studies in this sample made their participants do two intervention periods of seven weeks each. Three studies operated with sessions rather than weeks: One study only examined children for one session of intervention. Another study conducted two intervention sessions in a two-week period. The third study had six sessions of intervention in a year. Finally, two of the studies examined how children interacted with the apps in six trials, where each trial lasted for about one minute.

12 studies focused on L1 learning, while the remaining three studies were looking at L2 learning. Of the studies involving L1, all but one had English as the L1: The study by Teepe et al. (2017) was looking at the expressive and receptive vocabulary skills of Dutch L1 children. The three L2 studies examining the efficacy of apps on second language learning were all focusing on English as a second language. These were also the studies involving middle school children, except for the study by Patel, Torppa, Aro, Richardson, & Lyytinen (2021) who studied L2 learning in 5-7-year-olds.

Most of the studies in this review reported that apps could be effective for learning some aspects of language, however, a few were only effective for certain age groups, specific modes, etc. There were a lot of conditions that would have to be met for the apps to be effective. One of the studies reported that the app was not effective for teaching reading skills due to the lack of effect on out-of-game measures (Patel et al., 2021). Another study found that the app under investigation could boost engagement, but there was uncertainty whether the app was effective for actual language learning (Al-Bogami & Elyas, 2020).

Finally, the studies in this review examined different aspects of language learning. Eight studies were examining the efficacy of using educational apps to learn vocabulary. Specifically, expressive and receptive vocabulary knowledge and growth was tested in half of them. One of the vocabulary studies was also looking at the effect that self-regulation and interactions with the touchscreen would have on word learning. Another of these studies was also looking at comprehension, pronunciation and listening. A third one also studied the effect apps could have on parent-child interactions and the development of syntactically complex utterances. Six studies were looking into the possibility of developing emergent literacy skills through educational apps. Three of these had special focus on phonological awareness. Other focus areas were print knowledge, word recognition, name writing, and letter knowledge. Furthermore, two of the vocabulary

studies focused on reading comprehension. One of these was an observational study which measured prompted and unprompted retelling, inference/critical thinking, as well as vocabulary and connection (Christ, Wang, Chiu, & Cho, 2019). The other study examined reading and decoding accuracy, reading and decoding fluency, as well as spelling. Finally, the remaining study of this review focused on the degree of helpfulness, ease of use, and engagement the use of tablets and apps could provide, and how this could benefit learning.

5 Discussion

5.1 Main findings

There were only three articles that involved studies on educational apps and second language learning in this review. Nonetheless, whether it is L1 or L2 learning, language learning in young children is dependent on basic literacy skills, and there were several studies focusing on this. Furthermore, whether one is learning an L1 or an L2, learning vocabulary is always important. Eight of the included studies were looking at vocabulary learning through apps and are therefore important for this review. The following sections will discuss the findings in these studies in relation to the four pillars of learning and second language learning.

5.1.1 Vocabulary Learning

Hirsh-Pasek et al. (2015) claimed that learning needs to be meaningful to the child. For young children, this includes building a knowledge base where more learning can be established. This often happens through drill and practice, which is especially useful for vocabulary learning. For instance, when learning a language, children need a vocabulary base to have more meaningful learning. "Sometimes apps that invite drill and practice and are instantiated in a game-like framework can be educational and effective for building up the base on which meaningful learning rests" (Hirsh-Pasek et al., 2015, p. 14). More exposure to language input is also a form of drill and practice. The eight studies that examined some aspect of vocabulary learning were the two studies in Dore, Shirilla, Hopkins, Collins, Scott, Schatz, Lawson-Adams, Valladares, Foster, Puttre, Toub, Hadler, Golinkoff, Dickinson, & Hirsh-Pasek (2021), and the two experimental studies in Russo-Johnson, Troseth, Cuncan, & Mesghina (2017), as well as the studies conducted by Rowe, Turco, & Blatt (2021), Teepe et al. (2017), Cavus & Ibrahim (2017), and the quasi-experimental study by Vatalaro, Culp, Hahs-Vaughn, & Barnes (2018). Learning vocabulary is not much different in L1 than L2. The main difference is that learning vocabulary in an L2 usually involves the knowledge of the same vocabulary in the L1. As mentioned in chapter 2.3, Ellis (2015, in Phakiti and Plonsky, 2018) noted that there are five different levels at which L1 may influence the L2: linguistically, psycholinguistically, contextually, developmentally, and individually. Linguistically, the L1 may influence L2 at both a lexical and structural level, meaning that the L1 can influence L2 vocabulary learning. This influence can both lead to errors as well as it can be a positive thing. If the target L2 word is similar to that of the L1, it might be easier to acquire. However, if the target word resembles another word in the learner's L1 which does not convey the same meaning, error may occur. This is something that teachers need to be aware of when implementing vocabulary learning apps in the second language classroom.

The RCT study by Dore et al. (2021) was examining app use and receptive vocabulary learning, while the second, experimental study in the article examined both receptive and expressive vocabulary knowledge. The first study concluded that children can in fact learn new vocabulary from a single bout of playing a mobile game. They also learned how to generalize beyond the game context, which suggests that they learned more than just the game. The game in question was an in-house developed app, *the Quicksand Rescue Mission*, which included 10 target words that the children were supposed to learn. The app was designed to align with the four pillars of learning introduced by Hirsh-Pasek et al. (2015). The app used second-person narration and a gender-neutral duck as the protagonist. It was designed to be attractive by using colorful images. Importantly, they

made sure not to use extraneous and unrelated hotspots or sound effects that had nothing to do with the target words. Moreover, the app was meaningful by providing the new vocabulary as meaningful parts of the story rather than it being presented isolated from the text, for instance on flashcards. Finally, the app was socially interactive in that it gave the children a chance to provide a response, and it also provided feedback to these responses. The feedback was supposed to scaffold their learning.

Dore et al. (2021) found that the intervention group answered a significantly higher proportion of test questions correctly compared to the control group, and thus proved that vocabulary can be learnt from the *Quicksand Rescue Mission*. As this app was designed to activate and engage children, as well as being meaningful and socially interactive, this study could provide further evidence for Hirsh-Pasek et al. (2015) that educational apps should be designed accordingly to be effective for learning.

Like the first study, the second study by Dore et al. (2021) applied the *Quicksand Rescue Mission* app, but it also looked at another in-house app called the *Golden Eggs Mission*. The apps were the same in principle, but the *Golden Eggs Mission* included a different storyline. Both apps included the same five target words for the study. By conducting a receptive and expressive test which included both target words and no-exposure control words, Dore et al. (2021) found that there was a significantly larger gain from pretest to posttest for target words as compared to the no-exposure control words. Focusing on target words only, it was also found that there was a significant gain for target words in the posttest as compared to the pretest. The expressive test also found gains from pretest to posttest on target words. No such improvement was found for the control words. These differences were found even with a week-long delay between gameplay with the apps and the posttest. The results suggest that learning from a mobile game that aligns with the four pillars of learning, could also be positive for long term memorization of vocabulary. Furthermore, Dore et al. (2021 p. 464-465) noted that "children liked the game used in the current studies, appeared highly engaged during game play, and often asked to play it again". This quote underlines the importance of engagement for learning.

Regarding second language learning, it is important that the stories in these apps are comprehensible enough for the learner to understand. The Krashen's Input Hypothesis state that input needs to be comprehensible to be learnt, but comprehensible input alone is not sufficient for L2 learning. New input is necessary to move on to a higher level (Phakiti & Plonsky, 2018). In other words, storytelling apps like the *Quicksand Rescue Mission* and the *Golden Eggs Mission* could also be beneficial for second language learning, but it is crucial that the story is at a comprehensible level for the child. A story filled with new input would be ineffective. For the new vocabulary to make sense, and therefore be learnt, the rest of the story must be understood. That is the only way the new vocabulary will makes sense.

Two studies, Teepe et al. (2017) and Rowe et al. (2021), focused on parent-child interactions and vocabulary learning. Parent-child interactions are important for learning, because: "Before children enter formal schooling, their vocabulary develops mainly through verbal interaction with parents and other family members" (Teepe et al., 2017, p. 123). Teepe et al. (2017) examined the efficacy of a technology-enhanced storytelling (TES), *Jeffy's Journey*, on stimulating parent-child interaction and vocabulary knowledge development. *Jeffy's Journey* was designed following principles derived from research

into children's vocabulary learning and parent-child interaction, and involved three phases: the introductory phase, the avatar selection phase, and the story creation phase. The introductory phase was where the parent and child were introduced to six wordless images which demonstrated the different elements of the storyline: the main character, problem and setting of the story. The second phase of *Jeffy's Journey* asked the parent and child to choose an avatar each. Hirsh-Pasek et al. (2015) stated that such involvement promotes a minds-on mind-set and makes the children actively engage with the app. Objects to further the story line would also promote such a mind-set, which is what the final and third phase of Jeffy's Journey was about. The story creation phase was the final phase where parent and child interacted and took turns in the creation of Jeffy's story. Jeffy came to life supported by visual, auditory, and textual prompts chosen by the parent and child. Once again, this would mentally involve the child and parent. These features would also be engaging for the child, according to Hirsh-Pasek et al. (2015). The prompts to choose from were emotion, word, swop, and question prompts. These were expected to help build elaborate character descriptions and a more in-depth storytelling.

Teepe et al. (2017) found that only half of the dyads in the study used the storytelling prompts. Of the prompts used, the emotion prompt was applied most often, and the question prompt the least. However, the emotion and question prompts were the prompts that were most positively associated with story-related utterances for both parent and child, when used. These prompts were also positively associated with both decontextualized and contextualized use of language. In other words, these prompts supported high-quality interaction. Contrary to their believes, Teepe et al. (2017) found that the swop and word prompts were not associated with story-related interaction. The swop prompt was only found to associate with interaction-related utterances for the child, while the word prompts was only related to decontextualized language for both parent and child. Teepe et al. (2017) summarized it by stating that the prompts that aligned with the story were the ones that encouraged active and verbal participation and thus facilitated high-quality story-related language. "Prompts that were more difficult to connect to the storyline did not result in story-related language" (Takacks et al., 2015 in Teepe et al., 2017, p. 134). In other words, the prompts had to be meaningful to the story. This is in accordance with Hirsh-Pasek et al. (2015) and their four pillars of learning stating that the learning must be meaningful to the child. Also, along with the four pillars of learning, Teepe et al. (2017) found that the children actively participated during TES.

In Rowe et al. (2021), three apps were used to develop vocabulary: *Photo Play*, *Story Mixer*, and *Animal Antics*. All three apps were designed to initiate parent-child interactions. The *Photo Play* app centers around the parents and the photos of their families which they have on their phones. The app contains prompts which aims to lead the parents to discuss these photos with their child. This would help establish joint attention (Vogt et al., 2017). Additionally, the app incorporated features that allowed the parent-child dyads to decorate and play games with the photos. The second app, the *Story Mixer* was a story game that allowed the parent-child dyads to create their own unique stories by changing the objects in popular nursery rhymes. By using popular nursery rhymes, they make the task at hand more familiar, which is important for second language learning. However, it cannot be too familiar either: the task must be sufficiently challenging (Vogt et al., 2017). The final app, *Animal Antics*, was a performance-based app where the parent and child were supposed to create and perform their own stories as voices of characters having a conversation in different settings. The app was designed to

promote turn-taking in conversations. The app allowed the parent and child to record their own voices and perform their story. The study found a significant positive association between the production of longer utterances and time spent engaging with the apps. It was also found that the apps did indeed elicit linguistically rich conversations between parent and child.

Such apps as those in Rowe et al. (2021) and Teepe et al. (2017), where parent-child interaction is encouraged, could be very useful for L2 learning as well as L1 learning. However, as with the storytelling apps in Dore et al. (2021), the input from the parents must be at a comprehensible level. If parents interact with children in a second language which is too complex for the child at that exact moment, the learning will be inefficient. Phakiti & Plonsky (2018) suggest that activating prior knowledge may help reducing the cognitive load in working memory. For instance, the *Photo Play* app could be useful as it involves talking about pictures of something that is familiar to the child. The *Animal Antics* app could easily be used to create a story about a subject the child already has some vocabulary about, for instance the name of animals, or food. Furthermore, input in the target language is critical for second language learning. Both exposure to the language and interaction with the language is important to learn a second language, and thus parent-child interaction apps could potentially be very efficient. However, the key concern regarding language interaction and language exposure is the quality of that exposure and interaction, and the stages and contexts of the learning (Phakiti & Plonsky, 2018). In parent-child interactions where the parents themselves have limited knowledge of the L2, the parents should be careful as the quality of the exposure to a language is critical.

Vatalaro et al. (2018) examined the efficacy of different apps on expressive and receptive vocabulary. They looked at scaffolding-like (SL) apps and open-ended (OE) apps and compared them with other 'control' apps. The SL-apps were *Endless Alphabet*, *Noodle Words HD-Action Set 1*, *Goodnight ABC*, and *ABC Go*. These apps provided the children with the target vocabulary and more detailed definitions and meanings of the words, while the OE-apps more often demanded the children to produce the vocabulary. The OE-apps were *Beck and Bo, Draw and Tell HD, Don't Let the Pigeon Run This App!* and *Alien Assignment*. The control apps of the study were chosen by the Head Start program and included *Letter School*, *Gazzili Science*, *Yumiloo Rainbow Power*, *Faces iMake ABC*, and *Counting Bear*. These apps varied in their content and structure, as well as in their learning goals.

One of the more important findings in Vatalaro et al. (2018) was that the OE-apps more often than the SL-apps required help from the teacher to produce the vocabulary. The SL-apps did not require the teacher as often, because the vocabulary was provided by the apps. This is in accordance with literature on digital media and second language learning. Second language learning apps should indeed provide scaffolding (Vogt et al., 2017). In general, Vatalaro et al. (2018) saw that SL-apps worked well in a classroom setting, because the children could use them on their own, or with assistance from a peer, at their own pace. All that was required from the teacher was to introduce the apps to the children. In other words, the primary vehicle for scaffolding vocabulary knowledge in the SL-apps was the apps themselves, and not the teacher. Furthermore, while these apps were not considered creative or open-ended by experts, the SL-apps were still found to be highly engaging. This might suggest that for specific tasks, such as vocabulary learning, "apps which label and scaffold vocabulary have a stronger impact on

receptive vocabulary than open-ended apps, which require the child or teacher to produce the vocabulary word" (Vatalaro et al., 2018, p. 463).

Interestingly, even though the OE-apps encouraged expressive production of new words by asking the children to talk out loud and using the iPad's microphone or camera functions, no improvement in expressive vocabulary was found in these children. This could potentially be due to the limited time of intervention. Eight weeks of intervention might not be enough time to make significant improvements in a child's expressive vocabulary. However, this could also be explained by the fact that the words were not comprehensible for them (Phakiti & Plonsky, 2018). If they needed to produce the vocabulary themselves, without knowing the meaning of the word, it would be difficult to learn. Often in vocabulary learning, retaining the phonological shape of the word happens first, whereas acquiring the meaning is more difficult. This might be the reason for failure on this task. In relation to second language learning, the point remains that "L2 learners can only produce and comprehend language that they can currently handle" (Phakiti & Plonsky, 2018, p. 225). However, with Vygotsky's theory of the Zone of Proximal Development in mind, it should have been possible to learn the new words if the teacher helped them, or the app itself provided some sort of scaffolding. Phakiti and Plonsky (2018) state that L2 teachers should provide guidance before they allow their students to become more and more independent in their learning. For instance, the OE-apps in Vatalaro et al. (2018) could potentially be used in an L2 classroom, but the teacher should provide guidance *before* letting the children play with the apps individually. A quick explanation of the target words would go a long way before the children go on and play with the apps individually. On the other hand, apps that have in-game scaffolding may reduce the need for a teacher, however, studies on L2 learning suggest it is important with social interaction and scaffolding from a teacher either way (Phakiti & Plonsky, 2018). For the OE-apps, help and guidance from the teacher would be crucial to progress and learn the new vocabulary, which is supported by the findings in Vatalaro et al. (2018). Vatalaro et al. (2018) conclude their study by saying that apps that have been carefully and intentionally selected for a learning purpose, can support and transform learning. However, "children should be observed using an app, and if it does not serve its desired purpose or does not seem to be working well with a specific group of children, it should be removed from the selection of apps" (Vatalaro et al., 2018, p. 464).

Moving on from parent-child interactions to touchscreen interactivity, Russo-Johnson et al. (2017) examined how interactivity with the touchscreen could influence young children's word learning in two studies. In the first study, they compared the results of 2-year-olds to those of 4-5-year-olds. While children under the age of three were not the target sample age of this review, the comparison with 4-5-year-olds made it relevant as it could potentially say something about the efficacy of language learning for children from the age of three. The second study also involved 2-year-old children, however, while the main purpose of the study was not to compare the age groups to each other, the age of the children was included as a variable and was analyzed in relation to the results.

Russo-Johnson et al. (2017) designed two different word learning apps for the two studies. The first study used a word learning app which included flashcards. First of all, this could potentially be a limitation as Hirsh-Pasek et al. (2015) argue that the learning activity must be meaningful. It could possibly be better to integrate the target words as

part of a larger game narrative. The narrative itself can be beneficial for children's learning if it is entertaining (Kolak et al., 2020). The app began with a brief narration which instructed the child not to tap the screen until the narration voice stopped. The app would not advance before the labeling was finished, independent on how frequent the child tapped the screen. The app was purposely designed with a hotspot – a tap-the-butterfly filler task - that was supposed to be unsupportive and irrelevant to the word-learning task. The second study, however, used another word learning app designed and programmed to automatically record the frequency, location, and the time of children's interactivity with the screen, like taps or drags. The app incorporated interactivity designed to be supportive and considerate, and thus support learning. They purposely excluded any irrelevant hotspots. Furthermore, the children were tasked with actively engaging with virtual novel objects on the screen.

The primary finding in the first study was that older children were significantly more successful at learning new words and inhibiting their tapping during instruction, which suggests higher self-regulation (Russo-Johnson et al., 2017). «These results highlight the particular struggle that very young children have in inhibiting their tendency to tap during moments when they are instructed to wait and listen, such as during teaching moments» (Russo-Johnson et al., 2017, p. 6). This is relative to the finding that younger children do not have the ability to inhibit attention to extraneous information (Hirsh-Pasek et al., 2015). The 4-5-year-olds were also tapping the screen when they were not supposed to, however, they were more successful than the 2-year-olds and did not tap quite as frequent. Either way, Phakiti and Plonsky (2018) note that self-regulation is an important factor for L2 learning. Additionally, they underline that self-regulation can be improved by teacher intervention. Teachers can provide different learning strategies to help the children better their self-regulation, and thus better their learning outcomes. In other words, if the L2 learning outcomes related to an educational app are only limited due to low self-regulation, this could be improved and thus making the apps more efficient.

The second study of Russo-Johnson et al. (2017) concluded that all age groups could learn new words from using the app. This could be due to the less distracting app design, where self-regulation was easier. Sometimes the newly learned words were also seen to be transferred to the real objects. However, there was a significant age difference in the number of words learned, and especially on number of words correctly transferred to real objects. Comparison analysis found that 2-year-olds learned significantly fewer words that the 3-and 4-year-olds. Word learning for the older children was equivalent, but a paired sample t-test used to compare the children's transfer of labels to real objects suggested that only the 4-year-olds were proficient enough in their word learning to transfer their newly acquired vocabulary to real objects.

The two studies conducted by Russo-Johnson et al. (2017) reveal that children learn vocabulary better when the interactivity with the app is meaningful, and not distracting. This suggests that apps that are designed in line with research on how children learn, are more efficient for learning than other apps. Meaningful in that there are no distracting hotspots that are irrelevant for the word learning task, which is distracting and inefficient for learning. The tap-the-butterfly filler task for the in-house designed app in the first study illustrates this. Furthermore, since younger children are more sensitive to distractions than older children (Hirsh-Pasek et al., 2015), the importance of settings in apps becomes apparent. Children and their parents should have the possibility to turn on

and off distracting options as it suits them so that engagement during learning is not disrupted.

> Extraneous animations, sound effects, and tangential games might be appealing to a child when activated but not add to the child's understanding of the primary content because they disrupt the coherence of the learning experience and the child's engagement (Hirsh-Pasek et al., 2015, p. 12)

The final vocabulary study, by Cavus and Ibrahim (2017), focused on actual L2 word learning for young children between the age of 12 and 13 years old. Cavus and Ibrahim (2017) designed their own interactive book app, the *Near East University Children's Story Teller* (NEU-CST), and compared the vocabulary-, comprehension-, pronunciation-, and listening posttest results of the interactive group with a control group that only interacted with a traditional paper book of the same children's story as the intervention group. The NEU-CST was designed in accordance with Mayer's Multimedia Learning Principles to improve vocabulary, comprehension, pronunciation and listening skills of young, middle school students. From pretest to posttest, the intervention group outperformed the control group in all the skills. Particularly, students in the intervention groups performed significantly better on pronunciation skills compared to those in the control group. This could be due to the incorporation of a novel speech recognition engine in the app. The students used a 'text-to-speech' application which they read to, and which further checked their pronunciation by using the speech recognition engine. The students were then provided instant feedback on their pronunciation and were asked to repeat the word if the pronunciation was incorrect (Cavus & Ibrahim, 2017).

Hirsh-Pasek et al. (2015) noted that this element of app design can lead to deep engagement. When the interactions are contingent and the feedback is understood to be meaningful, it can focus children's attention on the app experiment and extent engagement for a longer time. It can also be argued that speaking and producing words requires thinking from the user and is such a more minds-on activity than simply tapping or hearing the words. Consequently, it can be argued that the speech-recognition engine in an educational app could contribute to active learning. Moreover, as Cavus and Ibrahim (2017) were in fact studying L2 learning, the result in this study is of outmost importance. They found that the students could use this app without the help of a teacher, and still improve on every skill. However, these students were a lot older than many of the other students reviewed here, which makes it important to state that children in middle school probably have a higher self-regulation than younger children and can thus more often be left to themselves when doing tasks on an iPad. However, the teacher should pay attention and provide help if necessary, since the danger of distraction is apparent even through adulthood (Hirsh-Pasek et al., 2015).

5.1.2 Literacy Skills

If building a knowledge base is important to make other learning more meaningful (Hirsh-Pasek et al., 2015), then drill and practice to improve literacy skills must be important to make any language learning meaningful for young children. Six studies examined the effectiveness of educational apps on emergent literacy skills: Arnold et al. (2021), Neumann (2018), Patel et al. (2021), Patchan and Puranik (2016), Christ et al. (2019), and O'Brien et al. (2019).

Arnold et al. (2021) focused on phonological awareness and print knowledge and used the *Khan Kids* app to study whether the intervention group would improve their emergent literacy skills as compared to the control group. The control group only intervened with two apps that did not specifically target emergent literacy skills. In short, Arnold et al. (2021) found that the Khan Kids group had a significantly larger increase in overall emergent literacy skills than the control group, thus proving to be efficient for language learning. The app was developed in cooperation with learning experts and based on research on how children learn to read. It was designed with interactive activities, instructional videos, as well as virtual books and other creative tools. Five different characters were used to guide the children through the learning path which rotates through a variety of topics. The app personalizes the learning experience individually, for every child, by dynamically adapting to the child's level of skill on each topic. These features are all found to be effective for learning through digital devices. Especially, Vogt et al. (2017) underlines the importance of the app being able to adapt to the individual child's proficiency and learning needs. The *Khan Kids* app bases its scaffolding on the correct and incorrect answers made by the child. The app also provides supportive feedback, tips, and other types of encouragement by the character named Kodi Bear. The design is active in that it involves activities where the child must interact with the app, for instance by tracing a letter correctly. This could potentially result in a more minds-on mind-set. Furthermore, the app was engaging in that it involved both extrinsic and intrinsic feedback, as well as other engaging aspects such as colorful images. Moreover, the interactivity could be experienced as meaningful in that they learned how they could combine different sounds to make words. Finally, the app could also be thought of as being socially interactive in that the child was interacting with five different characters on screen. Kodi Bear gave feedback and encouraging advice when needed, which could help the child develop some sort of a parasocial relationship to the characters (Hirsk-Pasek et al., 2015).

Neumann (2018) studied the *Endless Alphabet*, *Letter School*, and *Draw Buddy* app by comparing an intervention group with a control group who had classroom instruction as usual, without iPads. The *Endless Alphabet* is a letter-matching app that introduce all the letters in the English alphabet. With multimodal features, such as sound and animation, the content was engaging. Interactive features, such as dragging, swiping, and tapping, were easy to use to complete the letter-matching tasks. These tasks were open-ended activities with no scores, time limits, or different levels. Furthermore, the *Letter School* app is designed as a multimodal letter tracing game. It involved three different activities to form different letter shapes: tapping, writing, and tracing. When children completed an activity, they were rewarded with stars and clapping sounds. This form of extrinsic feedback could be engaging, however, Hirsh-Pasek et al. (2015, p. 13) note that one must be careful with only applauding intelligence because "praising children's intelligence leads them to avoid the inevitable risks of learning for fear of appearing stupid and losing face". Instead, the apps should have an additional focus on efforts. Praising children's efforts makes them understand that hard work pays off and that learning is not something that happens in an instant. Ultimately, children whose effort is praised will succeed more often.

The final app, *Draw Buddy* is a drawing app which allows for freely drawing of picture or letter forms on the screen interface by using their fingers. In this app, there were no interactive sounds or animations. Instead, children could use the app for experimenting, drawing, mark making, and writing their own letters, words, and name. While this app

allows for creativity, the lack of engaging elements such as feedback, sound, and colorful animations may have been the reason why children were observed to be less engaged with this app as compared to the *Letter School* and *Endless alphabet* which had interactive features. However, the features in *Letter School* and *Endless Alphabet* also had some features that were found to be more distracting than effective. Among other things, *Endless Alphabet* has continuous background music which made it difficult to hear the letter sound. Background music has been found by other researchers to be a distracting element (Barr, Shuck, Salerno, Atkinson & Linebarger, 2010 in Hirsh-Pasek et al. 2015). While such a feature can be engaging, the background music was meaningless for this task. Moreover, in the *Letter School* app, the home screen contained all the 26 letters of the English alphabet which continually and randomly flipped around the home screen. This was also a meaningless, distracting element, which could potentially disrupt the engagement and learning.

Patel et al. (2021) assessed the effectiveness of a game-based phonics intervention for first and second grade English language learners in India. They examined the efficacy of *GraphoLearn* (GL) English Rime. The GL make use of adaptive technology which allows for individualized practice, which is suggested to be important for L2 learning through digital technologies (Vogt et al. 2017; Vulchanova et al., 2017). The GL English Rime uses "systematic rhyme family groups, where a small number of individual letter-sounds (grapheme-phoneme) correspondences are introduced, after which they are combined to form larger and more consistent orthographic rime units, and finally words" (Patel et al., 2017, p. 3-4). Throughout the game, the app provides extrinsic feedback and motivation in form of stars and coins. As with the *Letter School* app, only focusing on intelligence could potentially be problematic as it can lead to taking fewer learning risks in fear of losing face in front of others (Hirsh-Pasek et al., 2015). Nonetheless, Patel et al. (2021) found that the GL group made significant improvements on in-game assessments of word-recognition, letter-sound knowledge, as well as rime unit recognition. Moreover, the GL group showed that they were able to use their newly learned knowledge to recognize even larger units and words. However:

> From an educational perspective, the larger goal of using such games is to teach skills and not just to teach the game. In other words, it is essential that learners can transfer the skills learned in a game to out-of-game contexts (Patel, 2021, p. 10).

If learners cannot transfer their knowledge from an in-game context to an out-of-game context, the learning has not really succeeded in being efficient. Unfortunately, Patel et al. (2021) found no transfer from the skills learned in GL to the oral and paper-based measures, which may suggest that the children primarily learned the game, and not the skills. Phakiti & Plonsky (2018) points out that explicit instruction must be accompanied by opportunities to use the newly learned L2 in a communicative setting with a clear purpose and goal. By this, it is understood that explicit instruction alone will not be sufficient after a while. The potential problem with explicit instruction without the opportunity to use the newly attained L2, is that they may not be able to transfer the knowledge from the explicit instruction setting to another.

Patchan and Puranik (2016) assessed the effectiveness of the app *Writing Wizard* on the development of writing abilities, in particular: letter naming and letter writing skills. They had two intervention groups: an iPad and finger group, as well as an iPad and stylus group. These two intervention groups were then compared with a paper and pencil

control group. The *Writing Wizard* is a letter-writing app which first models the selected letter, then gives the child the opportunity to write it. If the child strays too far from the model, the app will stop and provide scaffolding by placing an arrow where the child should restart. Upon successful completion, the child would simply touch the screen, and the app would refresh so that the child could try again. The app also included the option to hide the model so that the child could practice writing the letter without guidance.

Patchan and Puranik (2016) found that using tablets and the *Writing Wizard* app could be effective for teaching children to write uppercase letters. However, children in the iPad finger condition wrote more letters correctly as compared to the children in the iPad stylus condition, suggesting that it is more beneficial to enhance a child's tactile experience during learning than it is to increase the similarity between the practice and transfer task. The stylus is more like a pen; however, it did not benefit or help the children to write more correct letters at the paper and pencil posttest. Patchan and Puranik (2016) also discussed the extrinsic feedback that was given by the app, suggesting that it was not helpful: "Perhaps, additional extrinsic feedback might not be necessary to support the development of writing skills […]. Additionally, by providing feedback as soon as an error was made, the tablet computer could have disrupted the encoding process" (Patchan & Puranik, 2016, p. 134). This is supported by Hirsh-Pasek et al. (2015). The immediate feedback and correction could have hindered the child's ability to stay on task. Patchan and Puranik (2016) suggest that it could possibly be better to receive the feedback after completely writing the letter rather than being interrupted, and thus distracted. A potential solution to the problem in *Writing Wizard* is to enable parents and children to turn of distracting elements like these.

Christ et al. (2019) studied the relations amongst reading characteristics, app book characteristics, and comprehension outcomes by using the multimodal interactive app books *Frog*, *But not Hippo*, *Gustav the Goldfish*, *Troop*, *Fox in Socks*, and *Penguin*. These app books were chosen because they were highly interactive and allowed for user participation, control, and discovery. As with other interactive books, like those in Dore et al. (2021) and Teepe et al. (2017), it is important that the content is both entertaining and comprehensible (Phakiti & Plonsky, 2018). Furthermore, the hotspots in the books had to be primarily in line with the story, which agrees with the meaningful pillar of learning. Moreover, the apps had to be aesthetically appealing and include animation. It had to be easy for the children to use to enable them to interact and engage with them independently. Other features had to be developmentally appropriate, meaning that features such as font size and the amount of text per page had to align with the children's developmental literacy needs. Finally, the narrative had to be of good quality. All of these requirements to the interactive books are in line with research on second language learning and digital media.

In short, Christ et al. (2019) found that app book characteristics were important for reading comprehension. When children read app books that used more non-congruent hotspots, children were observed to have lower unprompted and prompted retelling scores, as well lower scores for inferences/critical thinking. Additionally, mode selection was critical to inference/critical reading scores. Children who used the Auto Read mode averaged an 18% lower inference/critical reading score and a less strategic use of hotspots. This can be seen in relation to Kolak et al. (2020) who suggested that the structure of the activity was important for intrinsic motivation and engagement: "Apps which give the opportunity for exploratory use alongside structured activities, might

increase children's intrinsic motivation and engagement" (Kolak et al., 2020, p. 6). However, Auto Read mode gave no such opportunities for exploratory use, which can explain the lower inference/critical reading scores. Moreover, reading in sequence was also related to 11% greater vocabulary depth, 18% higher prompted retelling scores, as well as 11% higher connection scores. It was also found that the greater amount of congruent hotspot used was positively related to vocabulary outcomes. However, in general it did not matter whether the hotspots were congruent, prompted, or directly supported a particular outcome, what mattered was that hotspots were used strategically to support positive comprehension and vocabulary outcomes.

Finally, O'Brien, Habib, & Onnis (2019) also studied English literacy skills, with special focus on reading skills. They divided the participants into three groups: A phoneme-level group, a rime-level group, and a word-level group. The study involved the *SeeWord Reading* app, as well as both the phoneme and rime version of the *GraphoLearn* app, and finally an in-house developed app at word level designed to build vocabulary and enhance word reading. The *SeeWord Reading* app was designed to teach students grapheme-phoneme correspondences. The GL Phoneme app was designed to teach letter-sound correspondences, while the GL Rime app focused on orthographic rime units and was designed to teach children to combine letter-sound correspondences into orthographic rime units. In short, O'Brien et al. (2019) found that spelling and reading outcomes increased across all groups. It was also found that those with poorer phonological awareness for reading fluency showed advantage for the word-level app, while children with poorer statistical learning abilities showed improvement at phoneme-level. In other words, the general pattern showed that different starting points would give different results, depending on the app. Everyone will not necessarily benefit from the same educational apps. However, scaffolding and personalized learning experiences (Vogt et al., 2017) for every child could potentially minimize the outcome differences.

5.1.3 Engagement

Al-Bogami and Elyas (2020) examined whether helpfulness, ease of use, and engagement could improve L2 learning in 12-13-year-old girls. Engagement is one of the four pillars of learning and is thus important for learning. However, is engagement alone sufficient for effective learning? Al-Bogami and Elyas found that by using different apps, like *Quizlet*, *iBook*, *Popplet Lite*, *Polleverywhere*, and *Pixton Comic Maker*, the children were more motivated to learn. The children of the study increased their participation, and the apps enabled them to stay focused and on task. Importantly, they had an increased sense of enjoyment in the EFL classes. The participants themselves stated that the iPad was beneficial for learning, and they found it helpful as the apps made both reading and vocabulary learning easier. They thought it improved their performances in the EFL.

The students became gradually less dependent on the teacher, which made them feel more in control over their own learning. Hirsh-Pasek et al. (2015) also agree that the level of 'control' is an important factor in why apps capture attention. There must be an *appropriate* level of control depending on factors such as age and experience. The children must be allowed to proceed at their own pace to sustain interest. This is one of the many advantages with tablets contra computers: touch-screen apps are more controllable by children of almost any age – depending on design – compared to a computer mouse or a keyboard. These results are directly relevant for L2 learning as

they did in fact observe L2 learning, suggesting that engagement can be efficient to ensure L2 learning of English.

5.2 What Are the Costs of Learning?

The focus of this comprehensive qualitative review is efficient learning. However, it is important to discuss the consequences of such as focus. Learning should be effective, but at what cost? The findings in this study suggest that educational apps can indeed lead to effective learning, however, there are a few consequences of the use of educational apps that should be considered.

5.2.1 Apps Replacing Teachers?

Apps could never replace the teacher in a classroom setting, but several apps in these studies were said to be possible to use independently, thus limiting the need for a teacher. For instance, the apps in Al-Bogami and Elyas' (2020) study were found to provide the 12-13-year-olds with some sense of control over their own learning. They became less dependent on their teacher. Nonetheless, Al-Bogami and Elyas (2020, p. 13) conclude that "integrating mobile technology in the sphere of classroom does not eliminate the crucial role of the teacher".

Other apps in these studies have in-built scaffolding features which were also hypothesized to make the teacher less needed. However, the learning outcomes when using these apps are quite limited, and a teacher could be needed to provide additional scaffolding which could help improve other language skills than those intended by the app. For instance, the *Letter School* and *Endless Alphabet* apps mentioned in Neumann (2018) had in-built scaffolding features, nonetheless, Neumann concluded that these apps would be strengthened if they were designed with young children's individuals learning needs in mind. The apps in Neumann's (2018) study contributed to the significant differences that were found between the intervention group and the control group for letter knowledge, print concepts, and name writing. The intervention group performed better than the control group on all skills. However, Neumann (2018) also found that letter writing did not improve by using the apps, suggesting that a more active scaffolding approach by a teacher may be required to support emergent writing during experiences with such apps. A more knowledgeable person such as a teacher or parent is an important factor for influencing the effectiveness of such apps in supporting emergent literacy.

5.2.2 Apps Replacing Children?

There is a risk that children become passive observers when interacting with educational apps. If you read a book, the book does not provide you with pictures of the main character. You must use your imagination to fill in the blanks, to give the story life and colors. This sort of mentally involved reading is what Hirsh-Pasek et al. (2015) calls minds-on. If the book app provides you with too many pictures, it could be argued that creativity is lost. Even when the learning is active, the app often provides the children with everything but the answer. All the child is left to do is to answer that specific task. Everything else is lost. For instance, the open-ended *Beck and Bo* app examined in Vatalaro et al. (2018), provided children with backgrounds to choose from and other items to fill the background with. While the app provides some freedom of creativity, the backgrounds to choose from were still set. The items to use were also set. The child is given a choice, but some of the creativity is lost when all the child has to do is choose.

The choosing can be done as easily as to simply tap the first one. Thinking and creating is no longer a necessity. When apps become too specific, other aspects of learning becomes irrelevant.

When apps replace the children, there is also a risk that they do not really learn anything. For instance, when the child is tasked with transferring knowledge learned in an in-game context to an out-of-game context, it sometimes becomes evident that the learning is flawed. To illustrate, Patel et al. (2021) found that the skills children learned while using the *GraphoLearn* (GL) – Rime app did not transfer to the oral and paper-based measures. This might suggest that the interactivity involved in the GL app did not mentally engage the children properly.

A further examination of interactivity with touchscreens was conducted by Russo-Johnson et al. (2017) who examined the effect of interactive features in different apps, like dragging, swiping, and tapping. They found that such interactions with the screen were easy to use, however, the study found no main effects for which of these behaviors led to best word learning. They did, however, find that tapping seemed to be more a reflexive action than a reflective one. In the study, tapping on named objects was expected to promote more learning than watching without interacting. Interestingly, this was not the case. While the children were engaging with the app, the children in the watch condition tapped a total of only 10 times while the children in the other condition tapped on average four times as often during instruction. They were instructed to listen and not tap on the screen during instruction, still, tapping did occur and this could have been disruptive for learning. Interacting with the screen in such a way may have primed the children's prepotent tendency of reflexive tapping, possibly distracting them from focusing on the words. Therefore, tapping may have been more a reflexive act than a reflective one, which may not have effectively directed the children's attention. Reflexive tapping on screens does not instigate a minds-on mind-set, and the child is somehow lost in the activity. Russo-Johnson et al. (2017) thus states that:

> Interactivity from touchscreens is a double-edged sword: on the one hand, haptic engagement (including touches on a response screen) can direct attention and focus and contribute to learning in adults [...] and children [...]. On the other hand, research indicates that interactivity in the form of hotspots and games can actually distract from learning [...] due to the need for a child to «task switch» or disengage from the interactive feature and selectively re-focus on educational content. (Russo-Johnson et al., 2017, p. 13)

Young children have limited self-regulation and may struggle to regulate attention and action. Developers of educational apps for children must therefore bear this in mind when adapting interactivity to their apps. They must think strategically about how the interactivity could benefit learning, and not distract from it (Russo-Johnson et al., 2017). Russo-Johnson et al. (2017) concludes their study by stating that the content of an app is important, however, the way in which children interacts with the app may be just as important in determining how much they learn, and thus how effective the apps are for learning. Children's self-regulation, age, gender, screen experience, as well as the physical requirements of engagement must be considered when designing educational apps for children so that the apps do not replace the children in the activity.

5.2.3 The Displacement Hypothesis

Are apps displacing other, more enriching activities? According to Hassinger-Das et al. (2020), the displacement hypothesis was first introduced when the television started occupying children. However, with the tablet and apps becoming more available and mobile, the hypothesis has resurfaced once more. In Great Britain, "a rising number of toddlers are now put to bed with a tablet instead of a bedtime story" (Vulchanova et al., 2017, p. 1), which might suggest that tablets and apps are in fact replacing other activities. People are critical of the time spent on these screens, and they are worried that they displace other, more important activities, such as parent-child interactions (Hassinger-Das et al., 2020). Parent-child interactions are of outmost importance when it comes to language acquisition. Vulchanova et al. (2017) explained a view on language acquisition as presented by Tomasello (2003), where language learning for a child deeply depends on the linguistic environment around them. Specifically, the linguistic environment provided to them through parent-child communication and interaction. A few studies in this review were in fact studying parent-child interactions and how they were affected using certain educational apps, as well as the learning outcome of these apps. These studies were the ones conducted by Teepe et al. (2017) and Rowe et al. (2021).

Teepe et al. (2017) examined the efficacy of TES, *Jeffy's Journey*, on stimulating parent-child interaction and consequently developing vocabulary knowledge. In short, the intervention groups improved their vocabulary as compared to the no-treatment control group, and TES was indeed found to stimulate active child involvement and generate parent-child interaction. More time on story phases was associated with more and higher quality parent-child interactions. Usage of prompts, especially prompts that were more related to the story, was also associated with improved parent-child interaction quality. TES was furthermore evidenced to improve children's productive vocabulary knowledge. In other words, Teepe et al. (2017) found that a good-quality TES app can in fact benefit parents in interacting with their children, and improved vocabulary occur as a result of that high-quality interaction.

Moreover, Rowe et al. (2021) also studied apps and their potential to encourage quality interactions between parents and child. By measuring children's language skills, as well as the quantity and quality of parent-child interactions and speech, it was found that interactive apps can help transform smartphones into opportunities for parent-child conversations and learning. More specifically, results show that "children who engaged in more app use had greater MLU at visit 2 ($r=0.45$, $p< .001$) controlling for MLU at visit 1" (Rowe et al., 2021, p. 7). MLU is the number and length of morphemes in an utterance. In other words, children who used the apps more often were the ones who produced longer utterances during app use at the last visit. The apps were thus positively associated with producing longer and more complex utterances.

As indicated by Rowe et al. (2021) and Teepe et al. (2017), apps do not necessarily replace other important activities, such as parent-child interactions. In fact, some apps may help support and improve it. According to Hirsh-Pasek et al. (2015), such interactive book apps that encourage parent-child interactions or dialog in order to stimulate story comprehension in a child, demand active involvement. "Parents and children can be actively involved with an app when they use it as a platform to discover new information about a content area" (Hirsh-Pasek et al., 2015, p. 10).

5.3 Limitations

Some limitations should be considered when interpreting the findings in this study. First, the keywords used in this study, and in the chosen databases, resulted in a limited number of studies on learning and learning apps. By limiting the search to studies in recent years, from 2014 onwards, the result was even smaller. This weakens the reliability of the study. While the study was supposed to involve studies solely focusing on second language learning through apps, there were simply too few studies. The search focus would have to be extended and involve all types of language learning.

Furthermore, the study relied on a relatively limited number of databases for the identification of potentially eligible studies. One of the databases resulted in over 14 000 results after conducting a search, however, this result was too overwhelming when the database did not include any specific ways of limiting the number without changing or adding to the selected keywords. Moreover, there is always the possibility that relevant articles and studies were excluded due to misleading titles or abstracts due to the screening method. Adding to this, the study is also difficult to duplicate. Some data extraction processes, like the screening, could have resulted in other findings for other researchers. While the method section is detailed, there is no guarantee that another researcher would evaluate the studies the same way as in this study. There are several processes going on in a researcher's mind when screening. Hopefully, the research result would have resulted in similar findings if someone else were to conduct the study. Another limitation of the study is that some studies included had incomplete reports of method. Among other things, two studies failed to inform where exactly the app intervention was completed.

Furthermore, the sample age needs to be further limited. Preschool children differ a lot from middle school children. The age gap between 3-year-olds and 13-year-olds is a relatively large one, as this study shows that children differ greatly in their learning even when 3- and 5-year-old are compared. However, due to the already limited number of relevant articles, the studies involving middle school children were included.

Luckily, the study has some strengths as well. By conducting a systematic search review, it is easy to summarize and present relative recent studies on the topic. The oldest study in this review is from 2016, while there were three studies from 2021. The study may be a potentially influential study for teachers and researchers as it illuminates areas on language learning with apps, but more importantly, it illustrates the importance of more research on the topic. This study clearly shows what topics within language learning and educational apps have been touched upon, and which topics have not.

6 Conclusion and implications

Today, second language teachers as well as other teachers make use of apps in their classroom teachings. While the result is limited in the number of studies taking on the exact issue with second language learning and apps, the findings suggest that it can be beneficial to use apps in second language learning as well as in other subjects. However, the teacher should be careful when implementing apps in the second language classroom. It is not as much a question of whether you should make use of apps or not – they seem to be a permanent change in the digital classroom – but *how* to use them. The teacher should not simply engage the children in the new app, and then leave them alone. Children need scaffolding from the teacher as well as from the app; actual face-to-face interaction will never be out of fashion in the second language classroom as well as any other classroom. By guidance from a teacher, the apps can be used in a way that is in line with the four pillars of learning. Apps themselves are often limited in their learning outcomes, but good teachers will know how to take the learning a step further.

In this comprehensive qualitative review, existing literature on educational apps and its efficacy was examined. The literature reviewed agree that there can be effective educational apps, however, not all apps are as successful. Some are simply engaging, while others are socially interactive and promote active learning, but there is nothing specific to be learned. More cooperation between educational app developers and child development researchers would result in educational apps with a common goal: efficient learning.

6.1 Professional relevance

The findings of this comprehensive qualitative review have direct relevance for me as a teacher. Teachers need to be informed about their technological reality concerning educational apps. Tablets and apps are being used in many schools on a daily basis. This technology was set in motion even before there was any evidence suggesting that it would be beneficial for learning. Many teachers and parents may be tempted to use all available apps in the market labelled 'educational', however, not all educational apps are efficient for learning. Such apps should be used with caution. As a teacher myself, working at an elementary school, I have realized that I need to be more critical of the educational apps that we make the children use. We need to have the four pillars of learning in mind when we evaluate the use of the different apps and should only encourage educational apps that activate and engage children, as well as make the learning meaningful and socially interactive for the child. Simply touching or swiping on a tablet will not improve learning. However, touching and swiping may help children engage with the learning material in a different way, but the design and content of the different apps must be evaluated to decide whether there are any real learning outcomes relevant for the child. For teachers, the job becomes to pay attention to the use and learning outcomes of these apps and provide active scaffolding when necessary.

Nevertheless, apps can never replace a face-to-face learning interaction, and a teacher can never be replaced by an app. However, it is proven that apps can be used in an educational way and thus benefit the children and their learning outcomes. A combination of using apps and teacher-student interaction is to be preferred, and teachers should work on how to make such a combination work the classroom.

Works cited

Al-Bogami, B., & Elyas, T. (2020). Promoting Middle School Students' Engagement Through Incorporating iPad Apps in EFL/ESL Classes. *SAGE Open*, *10*(2), 1-18. https://doi.org/10.1177/2158244020926570

Arnold, D. H., Chary, M., Gair, S. L., Helm, A. F., Herman, R., Kang, S., & Lokhandwala, S. (2021). A randomized controlled trial of an educational app to improve preschoolers' emergent literacy skills. *Journal of Children and Media*. https://doi.org/10.1080/17482798.2020.1863239

Cavus, N., & Ibrahim, D. (2017). Learning English using children's stories in mobile devices. *British Journal of Educational Technology*, *48*(2), 625-641. https://doi.org/10.1111/bjet.12427

Christ, T., Wang, X. C., Chiu, M. M., & Cho, H. (2019). Kindergartener's meaning making with multimodal app books: The relations amongst reader characteristics, app book characteristics, and comprehension outcomes. *Early Childhood Research Quearterly, 47*(2), 357-372. https://doi.org/10.1016/j.ecresq.2019.01.003

Dore, R. A., Shirilla, M., Hopkins, E., Collins, M., Scott, M., Schatz, J., Lawson-Adams, J., Valladares, T., Foster, L., Puttre, H., Toub, T. S., Hadler, E., Golinkoff, R. M., Dickinson, D., & Hirsh-Pasek, K. (2019). Education in the app store: using a mobile game to support U.S. preschoolers' vocabulary learning. *Journal of Children and Media*, *13*(4), 452-471. https://doi.org/10.1080/17482798.2019.1650788

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). https://doi.org/10.1542/peds.2019-1579

Hassinger-Das, B., Brennan, S., Dore, R. A., Golinkoff, R. M., & Hirsh-Pasek, K. (2020). Children and screens. *Annual Review of Developmental Psychology, 2*(1), 69-92. https://doi.org/10.1146/annurev-devpsych-060320-095612

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in 'educational' apps: lessons from the science of learning. *Psychological Science in the Public Interest, 16*(1), 3-34. https://doi.org/10.1177/1529100615569721

Kolak, J., Norgate, S. H., Monaghan, P, & Taylor, G. (2020). Developing evaluation tools for assessing the educational potential of apps for preschool children in the UK. *Journal of Children and Media*. https://doi.org/10.1080/17482798.2020.1844776

Neumann, M. N. (2018). Using tablets and apps to enhance emergent literacy skills in young children. *Early Childhood Research Quarterly*, *42*, 239-246. https://doi.org/10.1016/jecresq.2017.10.006

O'Brien, B. A., Habib, M., & Onnis, L. (2019). Technology-Based Tools for English Literacy Intervention: Examining Intervention Grain Size and Individual Differences. *Frontiers in Psychology, 10*, 1-21. https://doi.org/10.3389/fpsyg.2019.02625

One World Rental (2019). History of iPad. Retrieved from: https://oneworldrental.com/blog/all-ipad-list-and-history

Patchan, M. M., & Puranik, C. S. (2016). Using tablet computers to teach preschool children to write letters: Exploring the impact of extrinsic and intrinsic feedback. *Computers & Education*, *102*, 128-137. https://doi.org/10.1016/j.compedu.2016.07.007

Patel, P., Torppa, M., Aro, M., Richardson, U., & Lyytinen, H. (2021). Assessing the effectiveness of a game-based phonics intervention for first and second grade English language learners in India: A randomized controlled trial. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12592

Phakiti, A. & Plonsky, L. (2018). Reconciling Beliefs about L2 Learning with SLA Theory and Research. *RELC Journal, 49*(2), 217-237. https://doi.org/10.1177/0033688218781970

Rowe, M. L., Turco, R. G., & Blatt, J. H. (2021). Can interactive apps promote parent-child conversations? *Journal of Applied Developmental Psychology*, *76*. https://doi.org/10.1016/j.appdev.2021.101326

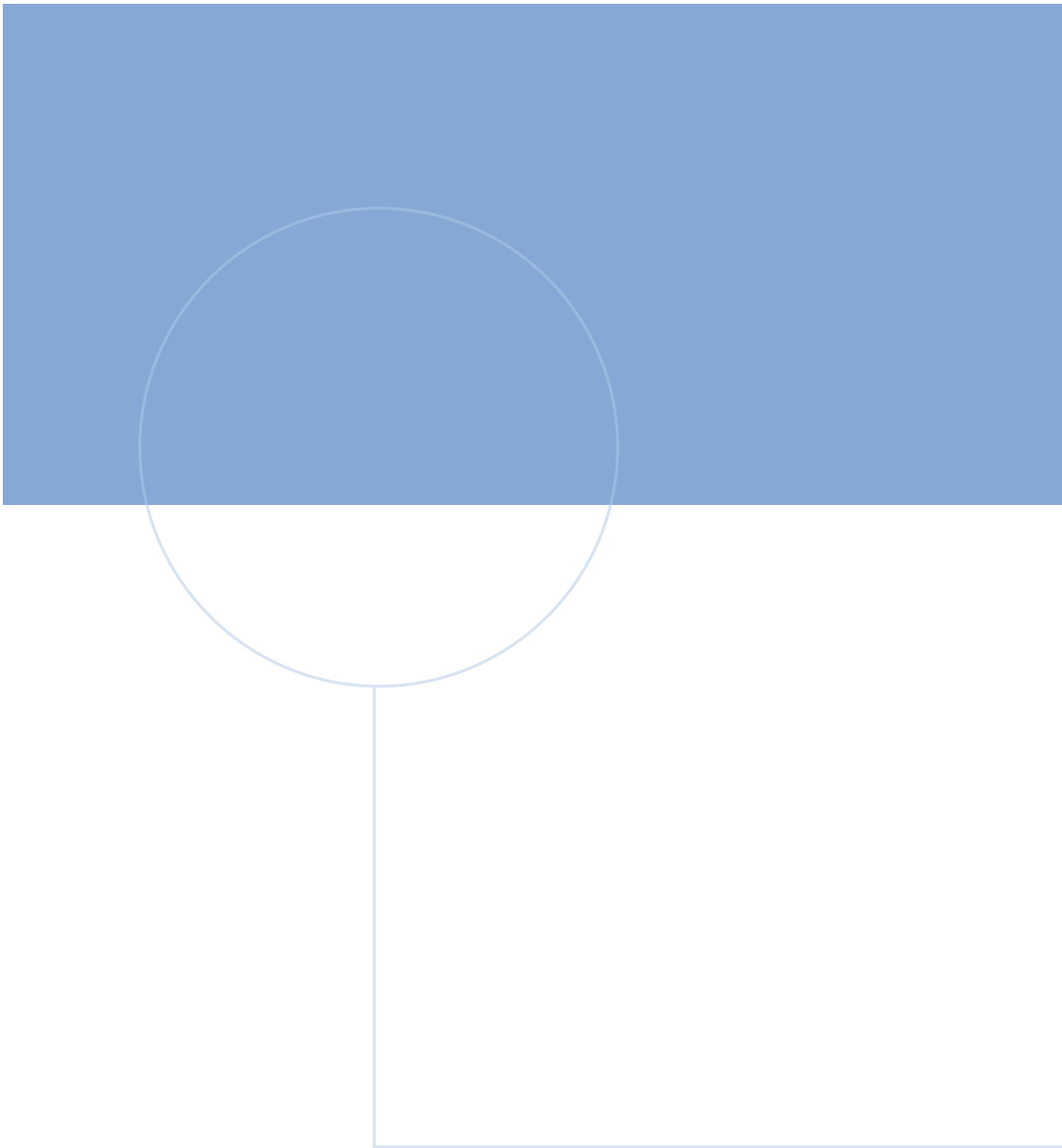
Russo-Johnson, C., Troseth, G., Cuncan, C., & Mesghina, A. (2017). All tapped out: Touchscreen interactivity and young children's word learning. *Frontiers in Psychology*, *8*(578). https://doi.org/10.3389/fpsyg.2017.00578

Teepe, R. C., Molenaar, I., & Verhoeven, L. (2017). Technology-enhanced storytelling stimulating parent-child interaction and preschool children's vocabulary knowledge. *Journal of Computer Assisted Learning*, *33*(2), 123-136. https://doi.org/10.1111/jcal.12169

Vatalaro, A., Culp, A. M., Hahs-Vaughn, D. L., & Barnes, A. C. (2018). A quasi-experiment examining expressive and receptive. Vocabulary knowledge of preschool head start children using mobile media apps. *Early Childhood Educ J*, *46*, 451-466. https://doi.org/10.1007/s10643-017-0877-3

Vogt, P., de Haas, M., de Jong, C., Baxter, P., & Krahmer, E. (2017). Child-Robot Interactions for Second Language Tutoring to Preschool Children. *Frontiers in Human Neuroscience, 11*(73). https://doi.org/10.3389/fnhum.2017.00073

Vulchanova, M., Baggio, G., Cangelosi, A., & Smith, L. (2017). Editorial: Language Development in the Digital Age. *Frontiers in Human Neuroscience, 11*(447). https://doi.org/10.3389/fnhum.2017.00447



NTNU

Norwegian University of
Science and Technology