International Journal of Mobile and Blended Learning
Volume 13 • Issue 1 • January-March 2021

A Socio-Cultural Approach to Evaluating and Designing Reading Comprehension Apps for Language Learning

Heydy Robles, Universidad del Norte, Colombia
Kevin Burden, University of Hull, UK
Karen Villalba, Universidad del Norte, Colombia

ABSTRACT

Mobile devices are increasingly promoted as tools to facilitate ubiquitous and individualized learning, allowing learners to work at their own pace in authentic and meaningful settings. However, in the case of second language learning, there is a paucity of apps and tools related to improving students' reading comprehension in both Spanish and English. Additionally, there are few studies that address the evaluation of applications for reading comprehension and innovation in this field and this is required in order to respond to the needs of transformation in language learning teaching. The authors present an original evaluation of 25 English language learning mobile apps using the iPAC app rubric, which identifies the pedagogical features of mobile learning: personalisation, authenticity, and collaboration. The results indicate that many of the existing apps fail to fully exploit the affordances of mobile learning and collaboration in particular. The findings suggest recommendations for app developers to design comprehension apps that address these shortcomings.

KEYWORDS

Authenticity, Collaboration, Higher Education, iPAC Framework, Mobile Learning, Pedagogical Affordances, Personalisation, Rubrics

INTRODUCTION

Mobile learning (referred to in this article as m-learning) is increasingly designed to exploit the different affordances and features of mobile devices, including their rich multimedia capabilities which make them highly attractive in standard educational practices (Reveiu, Smeureanu, & Dardala, 2009; Area & Adell, 2009; Olmedo, Grané & Crescenzi, 2012). Research suggests students are more engaged when learning is undertaken in a playful manner with digital devices, especially when this is associated with challenges and competitive actions (Trespalacios, Chamberlain, & Gallagher, 2011; Burden, Kearney & Schuck, 2019). Students currently inhabit a world in which different media contexts keep them permanently informed in real time, and in an interactive way. The use of mobile devices in educational contexts leads to increase motivation and has been found to be associated with faster, more individualized learning and more flexible and innovative collaborations (Area & Adell, 2009; Kearney, et al, 2012; Crompton & Burke, 2018; Burden, Kearney & Schuck, 2019). In Higher Education the majority of university students use mobile devices for tasks associated with their studies and almost half for specific learning tasks (López-Hernández & Silva-Pérez, 2016).

DOI: 10.4018/IJMBL.2021010102

However, teachers and university educators alike struggle to select appropriate resources, and particularly apps, to support their student's learning and there is limited support or guidance available for them to make informed choices about such resources based on sound pedagogical criteria (Papadakis & Kalogiannakis, 2017; Powell, 2014). It is estimated there are over one million apps labelled as 'educational' and this number continues to increase at an exponential rate making the task of selection ever more daunting (Cherner, Dix, & Lee, 2014; Stevenson, Hedberg, Highfield, & Diao, 2015). Many apps are poorly described or misappropriately labelled in the various app stores and a significant number of those that have been analyzed have been found to be largely 'drill and skill' in nature, based on an instructivist or transmissive pedagogy that is at odds with the more constructivists and collaborative affordances of mobile devices (Bano, Zowghi & Kearney, 2017; Goodwin, 2012; Murray & Olcese, 2011). Whilst there have been a number of attempts to develop typologies and recommendation lists to aid educators in their selection of appropriate mobile apps, most of these have had only partial success because they do not focus on the pedagogical orientation of the app or its utility for a particular pedagogical purpose (Cherner, Dix, & Lee, 2014; Lee & Cherner, 2015). In this paper, the researchers present the preliminary findings from a study exploring the utility and value of an internationally validated app evaluation rubric based on the iPAC framework developed by academics in Australia and the UK (Kearney, Schuck & Burden, 2020, in press). The rubric has been used by the authors to review a broad selection of reading comprehension apps for Spanish speakers and here we present the findings from this exercise and explore how these have been used to inform the mobile pedagogical characteristics and development of an innovative mobile learning app to support reading comprehension for Spanish learners.

BACKGROUND

M-Learning

M-learning is often described as more agile and spontaneous than other learning approaches (Ozdamli & Cavus, 2011;Traxler & Kukulska-Hulme, 2005) since the technologies that underpin it are ubiquitous and pervasive. M-learning frees students from the confines of traditional classrooms (Sha, Looi, Chen & Zhang, 2012) and with features such as GPS and location awareness, it is possible to tailor specific locations to particular types of learning experiences based on the needs of the individual learner, often referred to as location based learning.

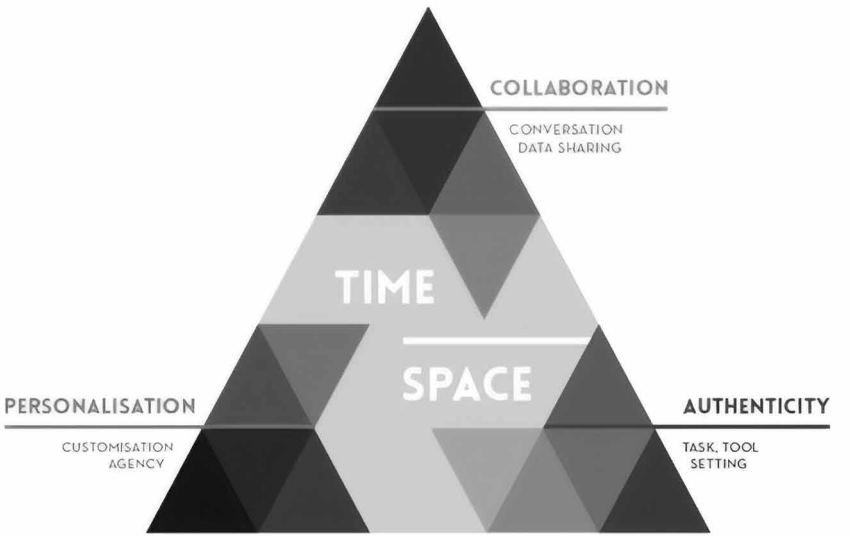
These features or affordances of mobile technologies indicate a shift in how we conceptualize and think about teaching and learning and they have therefore been accompanied by the development and promotion of different theoretical frameworks and models to help understand and analyze the growing phenomenon of m-learning. Some of these are briefly reviewed in the section that follows, leading to a justification for the use of the iPAC framework in this particular study.

M-Learning Frameworks and Models

A significant volume of theoretical frameworks and models have been developed to understand the phenomena of m-learning ranging from techno-centric models that focus primarily on the device, through to sociocultural frameworks that emphasize the interplay between the device, the user and the broader context within which this is situated (see Chapter 5, Kearney, Schuck and Burden, 2020 in press).

The researchers acknowledge the value of these different theoretical approaches but have chosen to underpin our work with the iPAC framework which is a well validated, socio-cultural model of mobile learning (Kearney, et al, 2012). This decision was based on the subject matter of the study which focuses on reading comprehension and is therefore well aligned to many of the constructs and sub-constructs that constitute the iPAC framework as explained below.

Figure 1. iPAC framework: constructs and sub-constructs (original version, Kearney et al, 2012)



The iPAC Framework

The iPAC framework was designed in 2012 and consists of three main constructs, or 'signature pedagogies' which are Personalisation, Authenticity and Collaboration (see Figure 1 below). ***Personalisation*** features include learner choice, agency and self-regulation as well as customisation which implies control over time, space, pace and autonomy over the learning content (Kearney, et al, 2012). ***Authenticity*** harnesses the potential of mobile devices to make learning more realistic and meaningful for students by the use of realistic, professional type tools and apps, the ability for learners to undertake real world tasks (either simulated or in real time) and the extent to which the context is realistic. ***Collaboration***, exploits the ability to involve learners in meaningful and extended conversations of many types and to enable them to generate and share their own multimodal content (Kearney, et al, 2012).

The iPAC framework has been used extensively to research, evaluate and design m-learning scenarios, both by the original authors and by many other academics and practitioners around the world. These studies include schools (Kearney, Burden & Rai, 2015), teacher education (e.g. Burden & Kearney, 2017; Schuck, Aubusson, Kearney & Burden, 2013), specific discipline studies (Burden & Kearney, 2016; Bano, Zowghi, Kearney, Schuck & Aubusson, 2018), Higher Education (Pegrum, 2019) and many studies which transcend the boundaries of each of these. The framework has been used to underpin a number of curriculum development projects and has been operationalised in the form of numerous research and evaluation instruments and resources (see for example www.mobilelearningtoolkit.com) including an app evaluation rubric which forms the focus of this article.

Mobile Apps in Language Learning

In order to understand the concept of apps and their influence on the education world, it is necessary to look at the definitions given by different authors. Castek & Beach (2013) state that an app is a

specialized program typically designed to run on mobile media platforms like smartphones, tablets and computers. Additionally, Islam and Mazumder (2010) add that a mobile app is an innovative and rapidly developing global information and communication technology piece that is simple, convenient, low-cost and downloadable.

The problem facing educators is how to select apps that are appropriate for a particular learning or teaching task. This is problematic partly because apps are often poorly described or labelled by their developers, but also because apps are extremely context sensitive. They are dependent on the specific context in which the learning is situated and as such they need to carefully designed to align with the various variables and features that make up the learning context. There should be specific instructional lessons and tasks as a way to support learning behind the use of the app (Beach & O'Brien, 2015).

In the case of reading and comprehension apps little is known about the pedagogical choices and decisions that teachers make when selecting such tools for use with their students. Selecting reading apps that can develop engagement and self-confidence, as well as reading strategies is a demanding process for teachers and there is a need to support them in this by providing more specific pedagogical criteria upon which they can search and identify potentially suitable apps.

Current Use of Mobile Apps in Language Learning

In the digital landscape that characterises university education there are many different tools that students can use to learn a second language (Pegrum, 2014). An obvious and well-known example is Duolingo, which is a free language learning platform that includes a website and a language learning app for mobile devices (Loewen, Crowther, Isbell, Kim, Maloney, Miller, & Rawal, 2019). This app boasts an interactive and competitive element given that users are awarded experience points when they perform a task correctly (Harous & Harahsheh, 2017). Multi-modality can be seen clearly in the recorded speech and playback feature on this app although it has been noted that background noise and interference when using this feature "on the go" makes this somewhat impractical, and the app has also been criticised for using outdated pedagogical approaches to language learning such as Behaviourism (Lotherington, 2018).

Mobile-Assisted Language Learning (MALL) deals with mobile devices characteristics as the ones mentioned in Duolingo and their use in language learning. Pachler, Bachmair & Cook defined MALL as "the processes of coming to know and being able to operate successfully in, and across, new and ever changing contexts and learning spaces with an emphasis on understanding and knowing how to utilize our everyday life-worlds as learning spaces," (2010, p. 6). Over recent years, enhanced mobile technology has led to ever more creative approaches to MALL, with researchers reporting on the potential benefits of augmented reality (AR) activities to provide authentic language learning experiences (Pegrum, 2019). In recent years there has been an increasing amount of research into the use of Mobile Instant Messaging (MIM) apps such as WhatsApp and how these can be used for language learning purposes. The effect of MIM on fostering student cognitive outcomes, development of social skills, building communication channels and enabling self-reflection activities have been some of the positive findings reported. Four categories of challenges emerged as well, namely insufficient facility support, inappropriate utilization and the lack of teachers' preparation towards the adoption of technology in the classroom (Tang & Hew, 2017).

Different authors have highlighted how apps can be used to enhance writing skills in a second or additional language (Awada, 2016; Bataineh et al, 2018). Other researchers stress the potential of apps to improve speaking skills and reduce anxiety related to oral production (Han & Keskin, 2016, Sun et al, 2017). Further work has highlighted the potential of MIM apps such as *WhatsApp* to extend the classroom by creating Communities of Practice (CoP) through meaningful dialogue (Keogh & Robles, 2018, Rambe & Bere, 2013) while vocabulary acquisition has also been identified as a significant benefits of mobile apps (Lai, 2016, Li & Deng, 2018).

Problems Associated With the Use of Educational Apps

Interest in the use of mobile applications to support reading and languages has increased significantly in recent years (Cheung & Slavin, 2013), however, the majority of applications are targeted at young adults or children, not university students, and there are few apps that provide opportunities for collaborative learning. Most English as a Second Language (ESL) apps are primarily form-focused which induces language learners to pay attention to linguistic form and most of the dominant methods are audio-lingual and task (test)-based (Kim & Kwon, 2012). In the same way, the majority of reading comprehension apps are aimed at children in school or second language learning and very few are targeted specifically at adults.¨Among the apps that do mention an age-range for users, the vast majority (90%) cite preschool-age children as at least part of their target audience¨ (Vaala, Ly & Levine, p.19, 2015).

In university undergraduate contexts, learners have to navigate through a variety of links, menus, web pages, texts and hyperlinks (García-Rodríguez, 2014). Consequently, the training of students in digital reading skills is required. Nonetheless, there are a considerable number of language related apps available but it is difficult for teachers and learners to evaluate all of them because they are not focused specifically on reading comprehension and without any clear pedagogical criteria or a rubric this process can be overwhelming for educators.

These are some of the many reasons why it was seen as important to design and build a bespoke reading comprehension app that would meet the requirements of post compulsory education students. Current systems for ratings apps are highly subjective and based on user reviews rather than any sound and validated pedagogical principles. Given the importance of designing such an app it was seen as imperative that we should review and evaluate existing reading comprehension apps using a pedagogically focused, valid rubric.

Reviewing Reading and Language Apps Based on Their Pedagogical Affordances

Since apps can be used for learning in many different contexts and for many different purposes, in designing a new app for reading comprehension, it is vital to assess existing apps both in terms of their technical specifications, but more importantly, in terms of their pedagogical orientation (Chen, 2016). Some of the earliest evaluation frameworks for apps were adapted from Reeves fourteen dimensions of computer-based education forms: epistemology, pedagogical philosophy, underlying psychology, goal orientation, experiential value, teacher role, program flexibility, value of errors, motivation, accommodation of individual differences, learner control, user activity, cooperative learning, and cultural sensitivity (1994). More recent evaluation frameworks related to language learning are based on the three approaches for evaluating Computer Assisted Language Learning (CALL) software: checklists, methodological frameworks, and Second Language Acquisition (SLA) (Levy & Stockwell, 2006). In terms of rubrics, REALL, a rubric for language learning applications was designed to focus on specific language learning and language learning dimensions and intended to fill the linguistic component gap by including the Common European Framework criteria (CEFR) (Martín-Monje, Arús, Rodríguez-Arancón & Calle-Martínez, 2014). In the last few years, a new rubric proposed by Rosell-Aguilar (2017), contains four primary categories: technology, user experience, pedagogy, and subject specific by using a list of questions to help educators decide whether an app meets their teaching needs.

An Alternative Pedagogical Rubric: iPAC

Although educators may start by searching for discipline specific apps which match their subject expertise, our own research suggests they are also drawn to use generic apps that are characterized by the absence of disciplinary content matter since these are more flexible and can be applied in more settings and contexts than discipline specific apps (Kearney et al., 2015). Therefore, there is a need to evaluate educational apps from both a discipline focus and from a more generic perspective where the emphasis is more likely to be focused on generative and creative activities (Goodwin & Highfield,

Table 1. Frameworks for evaluating language tools and apps

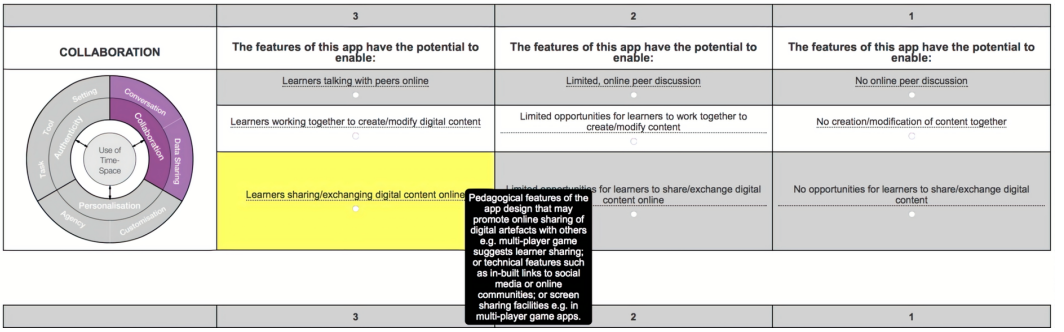
Frameworks and Rubrics for Apps Evaluation				
Reeves	**Chen**	**Monje et al.**	**Kearney & Burden**	**Rosell-Aguilar**
14 dimensions	LLM App rubric	REALL	iPAC framework/ rubric	LL App framework
1994	2014	2016	2016	2017
Wide range of dimensions	Technical and pedagogical criteria	Technical and pedagogical criteria	Quantitative and Qualitative	Technical and pedagogical criteria
No scores given	Scores given	Scores given	Scores given	No scores given
No international validity	No international validity	No international validity	International validity	No international validity

2013). The iPAC app rubric was developed to cater for both purposes and was specifically designed to help educators assess the pedagogical potential of apps with an emphasis on sociocultural models of learning since this is the pedagogical orientation of the iPAC framework itself. The rubric was developed iteratively between 20014-2017 as part of an international Erasmus+ project to develop a mobile learning toolkit (www.mobilelearningtoolkit.com), and was refined based on feedback from expert teachers, university experts, software developers, project members and an extensive set of trials with pre-service teachers in Australia and the UK.

Unlike many frameworks and typologies that are available to support the selection of educational apps, the iPAC app rubric purposely avoids context specific variables and questions. It aims to provide a generic tool for the selection of apps based on their pedagogical potential, not their disciplinary fit. This is related to the specific affordances or features of an app that may have been deliberately or accidentally designed into the app. So for example, apps created for star gazing (e.g. Star Walker) which are intended for use by individuals interested in astronomy and the stars, are designed to be used in a particular location (outside) and this is obvious to anybody who has ever used such an app. It takes advantage of the GPS affordances of mobile technologies to superimpose a representation of the stars against the night time sky when held up to the heavens. Similarly, some apps are deliberately designed to take advantage of the collaborative features of mobile devices that are always connected. Even though these features of an app may not have been deliberately designed by the developer, they nonetheless represent an inherent pedagogical affordance and opportunity for learning which the iPAC framework captures. Therefore, the iPAC rubric asks users of an app to evaluate its inherent pedagogical features, mapped against the three constructs and six sub-constructs of the iPAC framework. It does not ask users to evaluate the app based on how they actually used it since this is so context dependent it would be almost impossible to list every example. Hence, it is a unique rubric based on the potential of an app or tool to deliver a particular kind of pedagogical experience, based on the iPAC framework.

We have used the iPAC rubric because previous rubrics designed by others such as Rosell-Aguilar, Reeves and Martin Monje et al do not offer precise scores and the criteria used in the iPAC rubric can be quantified. In the case of Chen's rubric, the scores are included in a quantitative way but the rank is wider and there is a lack of linguistic components. By comparison the iPAC rubric offers both qualitative and quantitative evaluation data. The iPAC framework for example, takes into consideration other factors such as the teacher expertise and the tasks design. It clarifies the sub constructs and provides a useful lens to determine the features of pedagogical apps in a wide range of learning environments where the apps were actually used. All the constructs are informed by socio cultural theory. The iPAC rubric also allows reviewers to make comments in a freeway. There is a bar in which personal opinions and perceptions can be shared. Additionally, it offers the opportunity

Figure 2. Screenshot of the 'Collaboration' items in the iPAC app rubric



to see the potential of the app rather than judging it and does not contain items as assessment and aesthetic which can make it more accurate in terms of pedagogical features.

STUDY DESIGN

Context of the Study

The level of reading comprehension of students entering universities in Colombia is worryingly low. PISA claims that around 70% of teenagers over 15 in Colombia lack basic literacy (OECD, 2019) and students in general have shown a serious deficiency in text production and interpretation of what is read, seriously affecting their academic performance (Chinchilla, & Gómez, 2019). In fact, this is an international problem. Reading skills have not shown progress even in high income countries. In 2018, it was reported that almost ten million students were not able to complete even the most basic reading tasks (OECD, 2018).

Actually, some diagnostic tests that are applied at the beginning of each semester in a private university on the Colombian coast where this study is situated also revealed low levels of reading comprehension in students' mother tongue (Perez Zorrilla, 2005). This ability has been the subject of much research for years and is one of the weakest areas for students at this institution, resulting in problems such as retention and professional frustration (Gordillo and Florez, 2009).

These problems and issues associated with reading comprehension by students in one specific university are indicative of wider problems across Colombia and this was recognised recently by the award of funding to Universidad del Norte from Colciencias, a Colombian research agency to develop a mobile learning app to improve reading comprehension for students across the country. The app will be created using Design Based Research (Barab & Squire, 2004; Amiel & Reeves, 2008) and the first phase of this, which is the focus of this paper, is the identification of existing reading comprehension apps and their evaluation using the iPAC framework to ascertain which pedagogical features of m-learning are already evident in such apps and which features need to be developed further in future apps.

Selecting the Apps to Be Evaluated

In designing an app for reading comprehension we were cognisant of the need to design and develop it in line with sound pedagogical principles for m-learning. The researchers therefore made the decision to evaluate a selection of existing reading comprehension apps against the iPAC criteria, using the iPAC rubric tool in order to identify which, if any, of the three constructs were most apparent and what gaps exist for further development in this market sector. Therefore the following research questions underpinning this study are:

What are the mobile pedagogical characteristics of reading comprehension apps for Spanish language learners?
What pedagogical characteristics should feature in the design of a mobile reading comprehension app for Spanish learners?

The first step in the selection process was related to the revision of references related to reading comprehension and reading strategies. According to Guernsey & Levine, the first eight years of life represent the best stage in which a person can master different skills, including phonics, spelling, fluency and understanding of what is read (2015). In adults, the basic knowledge needed are the following: understanding moderately dense texts, determining cause and effect, making simple inferences and summaries and identifying the author's purpose (Kutner, Greenberg & Baer, 2006). Based on this, the first decision emerged: finding apps that would match the range of skills needed to become proficient readers through the creation of a checklist.

Initially, nine skills were catalogued to check if the titles of the apps contain these keywords. Early childhood skills were: Phonemic awareness, spelling, fluency, reading comprehension (Main Ideas). Adults essential reading skills were: Reading comprehension (Main ideas and details), summarizing, cause and effect, making inferences and identifying the author's purpose. Children's skills were basic and addressed to pre-school and elementary audiences. However, the search of apps in Apple and Android stores revealed more advanced literacy skills available for young learners and even adults who are in second language certification process. Other words such as 'reading', 'comprehension', 'reading skills', 'reading strategies' and 'text' were also used to search both of the Android and iOS (Apple) stores. Additionally, we referred to a number of web ranking sites such as 'appbot.co', 'appbrain' and 'apptrace.co' to identify suitable apps based on their popularity rankings and descriptions from actual users.

From these two stages, around fifty apps were identified. A more refined search of these apps followed leading to a number being discarded based on criteria that included: no target language; no updated version; no longer available in the stores; insufficient supporting information about the app or no reading skills included. Using these specific criteria and approaches twenty-five reading comprehension apps were eventually selected for the study.

The Evaluation Team

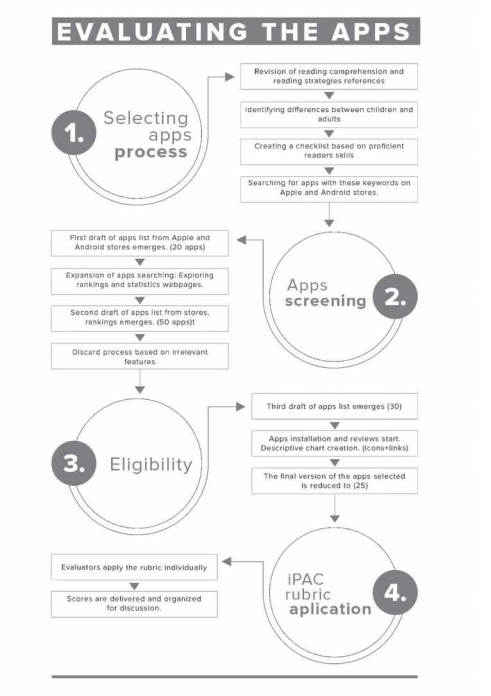
The evaluation of the relevant reading comprehension app was undertaken by three academics in a local university in the Caribbean coast of Colombia who have experience in language education and digital technologies. Each of these evaluators has a high level of expertise in mobile learning and in the use of a variety of apps for education in second language learning.

The Review Process

In order to undertake the evaluation using the iPAC app rubric we organised a two-week training and moderation exercise to ensure each evaluator understood and applied the iPAC rubric consistently. At the end of this exercise we met as a team to share our common understanding of the iPAC rubric and to evaluate a small sample of apps (separate from the twenty-five) as a pilot study. This exercise revealed some inconsistencies and misunderstandings about the different sub-constructs of the iPAC framework which were subsequently rectified by further discussion and exploration.

Following this training and moderation exercise all three evaluators were allocated the twenty-five apps to review and score against the iPAC rubric using the scale of 1-3 for each of the separate sub-constructs. A score of 1 indicated low or no evidence of the particular iPAC criteria and 3 indicates a high level of evidence for this criteria. This process was undertaken independently over a period of two weeks (December 2019) using the online rubric tool (see http://richprocter.co.uk/survey/mttep/rubric/) to submit individual scores which were then automatically aggregated by the tool (see Table 4).

Figure 3. The app evaluation process



Inter-Rater Reliability

At the end of this period the scores were analysed to identify how this batch of reading comprehension apps matched the three constructs underpinning the iPAC framework. Before this could be undertaken the data collected online needed to be 'cleaned' before it was ready for analysis. In the case of three apps only one evaluator had scored them and these were therefore removed from the final sample. To check for inter-rater reliability between each of the evaluators we aggregated the sub-construct scores for each construct and then aggregated the scores for each construct to find a mean score as shown in Table 4.

There are a number of methods that can be used to measure inter rater reliability including the traditional approach which produces a percentage number based on the total number of agreement scores divided by the total number of scores. However, this approach is less effective when a large number of variables are being evaluated or when there are more than two evaluators. We therefore used a statistical package (SPPS) and applied Fleiss Kappa, a variation of Choen's Kappa which is designed for three or more evaluators. It also takes account of the fact that some evaluators might actually guess on some of the sub-constructs that they are uncertain about. Fleiss Kappa is looking for the level of correlation between different raters and can produce a score from -1/+1.

In the case of the Collaboration construct the three evaluators were almost unanimous in the ratings they awarded for each app (their three scores agreed in 92% of the cases) although for the other two constructs they were less so as shown in Table 4 above. Nonetheless the combined score for all three constructs (0.74) was considered to be moderately consistent (see Table 3).

Low Scores for Most Apps

The iPAC framework consists of three principal pedagogical constructs (Personalisation; Authenticity and Collaboration) which are further sub-divided into six sub-constructs (see Figure 1). The online

Table 2. Kappa correlation raters

Value of Kappa	Level of Agreement	% of Data that are Reliable
0–.20	None	0–4%
.21–.39	Minimal	4–15%
.40–.59	Weak	15–35%
.60–.79	Moderate	35–63%
.80–.90	Strong	64–81%
Above.90	Almost Perfect	82–100%

iPAC app rubric automatically aggregates the sub-construct scores into a single score for each principal construct based on a score between 1-3. A score of 1 indicates a very low score for a particular construct and 3 indicates a very high score. For the purposes of this article we report the aggregated mean score for each construct along with an overall mean score for each app based on all three constructs (see Table 4). In Table 4 the apps are organised in rank order based on their overall mean score. The lowest mean score was 1.1 whilst the highest was 2.1. The mean score based on all twenty-five of the apps that were evaluated was 1.5.

Based on this relatively small but concentrated selection of apps it is evident that reading comprehension apps for language learners score relatively poorly when measured against the 'signature pedagogies' of m-learning as measured using the iPAC framework, with an average score between low and medium (1.5), where low (a score of 1) can indicate both low or no evidence of a particular pedagogical criteria. Given the socio-cultural orientation of the iPAC framework itself which emphasises the social and collaborative aspects of learning set within the malleable and flexible opportunities that mobile devices enable in terms of time and space, the overall results of this evaluation are somewhat surprising because it could reasonably be expected that the development of language skills would be grounded in many of the pedagogical affordances that the iPAC framework foregrounds such as conversation, data sharing, personal agency and the importance of situating learning in authentic and meaningful contexts using real world tools.

A more detailed examination of the results from this evaluation are also illuminating since they highlight an even more pronounced shortcoming around the use of the collaboration construct which might be expected to be an important element for language learning.

Examples of Apps and Their Evaluation Scores

Whilst the mean score for all of the apps that were evaluated was relatively low (1.5) as measured against the iPAC rubric, some apps performed better and were judged to be more pedagogically oriented in terms of socio-cultural learning and the affordances of m-learning. For the purposes of illustration three of these apps are described in the section below.

Table 3. Combined Inter-rater correlation scores

	Combined Correlation Scores
Personalisation	0.61
Authenticity	0.70
Collaboration	0.92
Overall correlation	0.74

Table 4. Mean scores for each app and iPAC construct

	Name of App	Code	Mean Score for Each Construct (1-3)			Total PAC Mean Score
			P	A	C	
1	Reading Raven	RAV	1.1	1.2	1	1.1
2	Reading Skills	RSK	1.1	1.2	1	1.1
3	Little Robot	LRT	1	1.7	1	1.2
4	Reading Coach	REC	1	1.7	1	1.2
5	RC Test Series	RCT	1.8	1.0	1.2	1.3
6	Reading Comprehension	RCO	1.4	1.6	1.0	1.3
7	IELTS	IEL	1.9	1.0	1.1	1.3
8	Inference Ace	ACE	1.8	1.0	1.2	1.3
9	Online reading tutor	ORT	1.9	1.0	1.1	1.3
10	Galexia	GAL	1.7	1.0	1.3	1.3
11	SSAT	SSA	1.5	1.8	1.0	1.4
12	Kids reading	KRE	1.6	1.8	1	1.4
13	Reading Eggs	REE	1.3	2.0	1	1.4
14	Reading comprehension GRE	GRE	1.6	1.8	1	1.4
15	Reading Prep Comprehension	RPC	1.6	1.8	1	1.4
16	QuotEd	QUO	1.3	2.0	1.0	1.4
17	Reading Comprehension Fun	RCF	1.5	1.8	1	1.4
18	Short Articles	SHA	1.6	1.8	1	1.4
19	Skimming & Scanning	S&S	2	1.7	1.0	1.6
20	Comprehension Builder	COB	1.8	1.9	1	1.6
21	Reading comprehension passages	RCP	1.8	1.8	1.0	1.6
22	TOEFL	TOE	1.9	1.8	1	1.6
23	Reading Cat	CAT	1.9	1.8	1.0	1.6
24	Speed reading	SRE	2	2	1	1.7
25	Elevate	ELE	2.3	2.8	1.2	2.1
Mean scores for each constructs			1.6	1.7	1.1	1.5

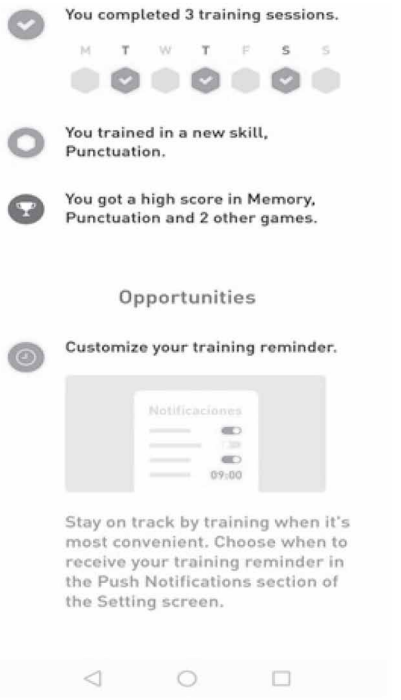
Elevate

[overall score of 2.1 (Personalization = 2.3; Authenticity = 2.8; Collaboration = 1.2]

Elevate is an app designed to train your brain in different aspects of learning which include oral skills, math, memory and reading among others. Users are provided with their own personalized training program that adjusts and adapts to their performance over time. As users train the app adapts to their skills and the daily training focus can be customized. Unsurprisingly this app scored highly against the customisation sub-construct of Personalisation (2.3).

The reading section of the app contains many different comprehension elements including agility, association, comprehension, connotation, context, extraction, processing, visualization and word parts.

Figure 4. Screenshot of the Elevate's achievement report



In the case of processing, for example, a passage is shown progressively as the number of words per minute increases. The idea is to reduce sub vocalization and prevent learners from mentally saying words and increase the reading speed. This feature of the app illustrates how the task undertaken by users is highly authentic because this is something that the students usually face every day when they are exposed to tweets, articles and comments on social media or when they simply deal with texts in the school or university. They can choose if they want to do the activity or switch it to another one depending on their performance. They can also customize their training reminder and repeat the activity as many times are needed to master it. Hence the authenticity score for this app was 2.8, the highest score recorded for any app in the evaluation.

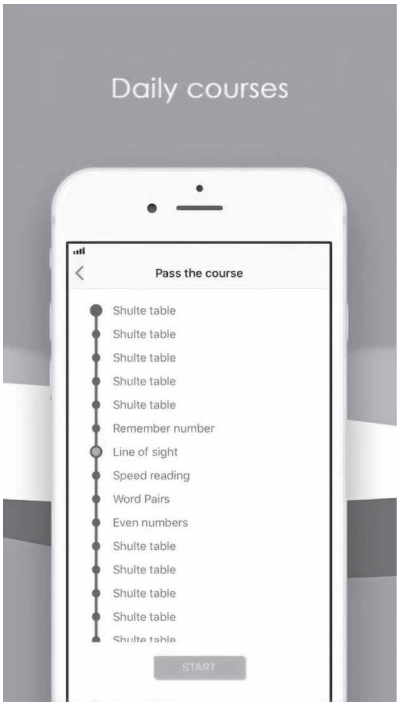
Nonetheless despite its high ranking and mean score, the Elevate app scored poorly in terms of collaboration (1.2) since this app only provided limited opportunity for digital conversations, allows people to send mails to share their doubts and questions but no space to participate in chat rooms or forums in which interaction levels can increase. Learners can also see their rankings and test results but they cannot follow other users or share their achievements on social networks. Therefore, this app scores poorly against both of the sub-constructs of Collaboration - conversation and data sharing.

Speed Reading

(Overall score = 1.7: Personalisation = 2; Authenticity = 2; Collaboration = 1)

Speed reading was the second highest scoring app in our evaluation with a mean score of 1.7. It is designed to encourage students to improve their peripheral vision by using the Shulte board and to improve their memory with word games and different reading sections in the premium version. There are also sections about neural speeder, numbers reminder, view field, concentration games, color confusion and attention focus. The app scored highly on customisation (a sub construct of

Figure 5. Screenshot of the Speed reading app layout



Personalisation) since learners have access to unique information tailored to themselves through the aggregation of statistics of their performance. They can also access a personalised library and some of the app settings can be modified to their individual preferences, such as the selection of activities. In general, this app reveals a balance between customization and agency (both sub constructs of Personalisation) because the users can move through the app freely without level restriction or intervention. These are the reasons that support a 2.0 score in terms of personalization.

Regarding authenticity, the app features a combination of real and simulated activities but there are few opportunities to use it in an actual realistic setting or z learning spaces. The teaching mechanics behind the app are engaging and the exercises about photographic memory engage the learners in real life activities such as following sequences, organizing words and making associations. Hence the app scored highly for authenticity (2.0).

However, like most of the apps in this evaluation, even those with the highest mean score, the collaboration construct obtained a low score, mainly because users have no opportunity to participate in online peer discussion and there is no facility in the app that enables them to share the content they have created with other learners online.

Skimming and Scanning

Overall score = 1.6 (Personalisation = 2; Authenticity = 1.7; Collaboration = 1)

The final high scoring app illustrated in this review is 'Skimming and Scanning' with a mean score of 1.6. Whilst this overall score is not high given the mean for all apps was only 1.5, it does include a relatively high score for the Personalisation construct which we therefore felt warranted illustration. This app is notable since it includes a range of resources and activities that enable users to select their own pace and level thereby making it a very customised learning experience. The learning materials

Figure 6. Screenshot of the Skimming and Scanning app

are categorized into three levels (easy, medium and difficult) in order to foster 2 reading strategies. The high personalisation score (2.0) was further bolstered by the ability of users to select their own avatars, play mode and articles to read, giving them considerable agency and choice in what and how they learned through the app. Authenticity, on the other hand, obtained a lower score of 1.7, less because most of the activities were artificial or simulated and were not judged to be realistic representations of real world situations.

As with all of the apps reviewed here, the collaborative component scored low. Although learners are able to 'play' with another partner in the app in a competitive mode, there is no opportunity in the app to create, modify or share content. Like most of the apps reviewed in this evaluation, there is a surprising absence of chat rooms or forums in the app 'Skimming and Scanning' and little or no opportunity to have discussions with peers.

Other Apps

In the case of apps that scored below the mean (1.5) they were also characterised by a significant lack of opportunity to undertake any form of collaboration with other learners which in terms of the iPAC framework means digital conversations (e.g. Twitter; online chats and forums) and the creation and sharing of data (e.g. sharing an assessment artefact with other learners in order to receive feedback). In the most extreme cases none of these features were available for learners and the apps were extremely instrumentalist in nature with learners reduced to basic skill and drill type activities. Many of the apps featured in this evaluation gave learners very limited opportunities to exercise choice or agency as it is referred to in the iPAC framework. Although the average score for the authenticity construct was above the mean for apps as a whole (1.7) in many cases the authenticity score was also very low and it was noticeable how few apps made use of realistic learners tasks and tools such as real life conversations and texts rather than simulated ones or the use of the app in real life settings beyond the classroom, for example.

CONCLUSION

Motivated by a national phenomenon and concern regard reading comprehension amongst post compulsory students in Colombia, which is likely to resonate with university professionals in many parts of the world, this study has employed an innovative and original mobile learning app rubric to

identify the characteristics of existing reading comprehension apps for language learners in order to inform the design and development of a new bespoke app which is itself the focus of a forthcoming but separate article.

In response to the first of our two research questions, the findings from this study reveal that current reading comprehension apps for use by language learners in higher education (research question 1) are generally weak in many of the salient features and affordances of mobile learning and particularly so in the case of collaboration that is recognised as both a key affordances of mobile learning and an essential skill that underpins effective language learning and reading comprehension. Some of the more innovative and creative apps such as 'Elevate' and 'Skimming and Scanning' showcased above, do exploit the personalisation potential of mobile learning by giving students more control and choice over their learning and by adapting to the learners' progress and needs in a form of customisation. However, these features are not universal in the apps we reviewed and in many of the examples it was a 'one size fits all' learning experience in which the learner was treated with little or no freedom to demonstrate agency or independent learning, both key elements of personalisation.

With a few exceptions this was also the case for Authenticity since few of the apps exploited the opportunities to set the learning in a real world context (e.g. by inviting real time language speaking via a native speaker) or to situate it outside of formal classroom spaces. Real world tools and tasks for language learning were seldom evident in these apps which tended to rely instead on highly simulated and often very artificial activities and exercises. Simulated forms of authenticity are very common in many software applications but mobile technologies offer opportunities to make these less artificial than traditional learning exercises or to replace them with entirely realistic real world learning opportunities (see Burden & Kearney, 2016). In the apps reviewed here there appear to be missed opportunities to make reading comprehension more meaningful, engaging and realistic for university students by grasping the affordances of mobile devices which are increasingly ubiquitous and available to students all of the time, not just in formal classes.

However, the low scores for collaboration is arguably the most significant and concerning finding from this study since conversation and data sharing - the two sub-constructs that constitute the Collaboration construct in the iPAC framework - are so important in language learning and indeed in the literature around mobile learning. In schools and where younger students are involved, the failure to exploit the collaborative benefits of mobile learning might be attributed to the concern teachers and parents often display around privacy and safe-guarding associated with the exchange of data and online conversions (Kearney, Burden & Schuck, 2019) but this is more difficult to understand or ascribe to apps that are designed for an older age group over the age of consent. Except of course, many of the apps we have reviewed here are not exclusively designed for university age students or adults, mainly because this was identified as a significant gap in the market in our initial review of literature. There is, therefore, an opportunity for app designers and teachers to consider the opportunity to design for an older, adult demographic which is the focus of our second research question.

Our second research question focuses on what we can learn from this evaluation of existing reading comprehension apps in terms of the pedagogical design and use of mobile apps in post compulsory learning settings such as university. The lack of opportunities for collaboration suggests that designers and creators of pedagogical apps are more focused on traditional learning approaches where information is 'delivered' to the student, rather than a socio-cultural model in which the learning is deemed to be more participative, social and mediated through technologies such as the mobile device itself. The findings from this study indicate that apps need to be designed to exploit more opportunities for collaboration between learners, both in and beyond formal contexts, and that learning from peers is, and always has been, a key element of language learning that should not be neglected in the rush to develop education apps for mobile devices (Tudge, 1992).

Social Interaction and active dialogue with peers is part of the sociocultural theory that supports authentic learning (Vygotsky 1978). In fact, Kearney et al (2012) claimed that the social interactivity corresponds to the Collaboration sub construct that is divided into Conversation and Data Sharing.

If the apps omit these two aspects, rich peer interactions will be limited as well as access to other mediations. We would recommend that for the design of an app for reading comprehension enhancing, short authentic texts such as aphorisms are an alternative to be included as part of the input that learners should be exposed to, since they broaden intellectual and cultural contexts in which aspects of daily life topics are revealed (Davis, 1999). In terms of collaboration, more opportunities for students to share new aphorisms, encouraging them to submit their own examples and share them through the sub constructs mentioned above should be identified.

On the other hand, the evaluation of current existing apps for reading comprehension also revealed a surprising lack of apps that focus on native speakers, in this case Spanish. Consequently, it is essential to develop an app that allows Spanish speakers to improve their native language not just English and respond to the necessity of improving Colombian students reading skills by introducing more demanding reading strategies such as inferences. Colombia is among the three countries with the lowest reading comprehension levels (OECD, 2019) and the intervention of this problem is imperative. Besides, most of the reading comprehension apps evaluated do not offer authentic contexts in which learners can interact with real texts. A great number of texts exposed in the results were artificial or created for educational purposes and only a few apps, such as those designed for exam preparation, were designed with this in mind. Additionally, agency (granting more control and choice to users) is seldom granted due to the need to structure the learning process in terms of ability levels. Based on this, it might be said that aphorisms are a suitable option to introduce authentic texts in reading comprehension apps.

All these aspects mentioned are part of the importance of raising awareness for apps developers and teachers in order to ensure they design and select apps using a sound and validated set of pedagogical criteria. The iPAC app rubric was designed specifically for this purpose in order to encourage both designers and users of apps to rise above the purely contextual factors that often inform their design and choice of an app but in ways that restrict and limit its potential use. The iPAC app rubric deliberately asks users to evaluate apps based on their inherent characteristics of affordances, not on how they have actually used them in their own context. This can be challenging and almost counter-intuitive for both teachers and designers and further research needs to be undertaken in this respect to fully comprehend the processes experienced by both designers and by educators when they use the app for this purpose. We acknowledge in this part that our sample of apps is relatively small and limited and also the fact that the level of agreement between the app evaluators was only moderate, not strong. However, the almost perfect score (92%) recorded for the Collaboration construct suggests that for this feature of mobile learning, at least, language educators hold a common understanding and shared vocabulary about what constitutes high quality reading comprehension exercises and also how inadequate or lacking many of the current apps designed for this purpose are. This has enabled the researchers involved in this project to identify a clear focus and purpose for their proposed app around authentic collaboration in what will be an iterative design based research process that we hope to report on shortly.

Conflicts of Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Funding Statement

No funding was received for this work.

Process Dates:

Received: March 13, 2020, Accepted: July 7, 2020

Corresponding Author:

Correspondence should be addressed to Heydy Robles, hrobles@uninorte.edu.co

REFERENCES

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Journal of Educational Technology & Society*, *11*(4), 29–40.

Area, M., & Adell, J. (2009). E-Learning: Enseñar y aprender en espacios virtuales. In J. De Pablos (Coord.), *Tecnología Educativa. La formación del profesorado en la era de Internet* (pp. 391-424). Málaga: Algibe. Recuperado de: http://andremalraux.edu.pe/wpblog/wp-content/uploads/2016/10/eLearning.pdf

Awada, G. (2016). Effect of WhatsApp on critique writing proficiency and perceptions toward learning. *Cogent Education*, *3*(1), 1264173. doi:10.1080/2331186X.2016.1264173

Bano, M., Zowghi, D., & Kearney, M. (2017, July). Feature based sentiment analysis for evaluating the mobile pedagogical affordances of apps. In *IFIP World Conference on Computers in Education* (pp. 281-291). Springer. doi:10.1007/978-3-319-74310-3_30

Bano, M., Zowghi, D., Kearney, M., Schuck, S., & Aubusson, P. (2018). Mobile learning for science and mathematics school education: A systematic review of empirical evidence. *Computers & Education*, *121*, 30–58. doi:10.1016/j.compedu.2018.02.006

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*(1), 1–14. doi:10.1207/s15327809jls1301_1

Bataineh, A. (2018). The Effect of Using Computerized Accounting Information Systems on Reducing Production Costs in Jordanian Pharmaceutical Companies. *International Journal of Business and Management Invention*, *7*(7), 1–10.

Beach, R., & O'Brien, D. (2015). Fostering students' science inquiry through app affordances of multimodality, collaboration, interactivity, and connectivity. *Reading & Writing Quarterly*, *31*(2), 119–134. doi:10.1080/10573569.2014.962200

Burden, K., & Kearney, M. (2016). Conceptualising authentic mobile learning. In *Mobile learning design* (pp. 27–42). Springer. doi:10.1007/978-981-10-0027-0_2

Burden, K., & Kearney, M. (2016). Future Scenarios for Mobile Science Learning. *Research in Science Education*, *46*(2), 287–308. doi:10.1007/s11165-016-9514-1

Burden, K. J., & Kearney, M. (2017). 'Investigating and critiquing teacher educators' mobile learning practices'. *Interactive Technology and Smart Education*, *14*(2), 110–125. doi:10.1108/ITSE-05-2017-0027

Castek, J., & Beach, R. (2013). Using apps to support disciplinary literacy and science learning. *Journal of Adolescent & Adult Literacy*, *56*(7), 554–564. doi:10.1002/JAAL.180

Chen, X. (2016). Evaluating language-learning mobile apps for second-language learners. *Journal of Educational Technology Development and Exchange*, *9*(2), 3. doi:10.18785/jetde.0902.03

Cherner, T., Dix, J., & Lee, C. (2014). Cleaning up that mess: A framework for classifying educational apps. *Contemporary Issues in Technology & Teacher Education*, *14*(2), 158–193.

Cheung, A. C., & Slavin, R. E. (2013). Effects of educational technology applications on reading outcomes for struggling readers: A best-evidence synthesis. *Reading Research Quarterly*, *48*(3), 277–299. doi:10.1002/rrq.50

Chinchilla, C. M. D., & Gómez, A. A. R. (2019). La comprensión lectora y el rendimiento académico en estudiantes de ingeniería. *Revista Colombiana de tecnologías de avanzada (RCTA), 1*(33).

Crompton, H., & Burke, D. (2018). The use of mobile learning in higher education: A systematic review. *Computers & Education*, *123*, 53–64. doi:10.1016/j.compedu.2018.04.007

Davis, M. S. (1999). Aphorisms and clichés: The generation and dissipation of conceptual charisma. *Annual Review of Sociology*, *25*(1), 245–269. doi:10.1146/annurev.soc.25.1.245

García-Rodríguez, A. (2014). Aplicaciones de lectura infantil y competencias digitales: evaluar antes de enseñar. *Vega journal. org, 10*(1), 21-39.

Goodwin, K. (2012). *Use of tablet technology in the classroom*. NSW Department of Education and Communities. Retrieved from http://rde.nsw.edu.au/files/iPad_Evaluation_Sydney_Region.pdf

Goodwin, K., & Highfield, K. (2013). A framework for examining technologies and early mathematics learning. In *Reconceptualizing early mathematics learning* (pp. 205–226). Springer. doi:10.1007/978-94-007-6440-8_11

Gordillo Alfonso, A., & Flórez, M. D. P. (2009). Los niveles de comprensión lectora: hacia una enunciación investigativa y reflexiva para mejorar la comprensión lectora en estudiantes universitarios. *Actualidades pedagógicas, 1*(53), 95-107.

Guernsey, L., & Levine, M. (2015). *Tap, click, read: Growing readers in a world of screens*. Jossey-Bass.

Han, T., & Keskin, F. (2016). Using a mobile application (WhatsApp) to reduce EFL speaking anxiety. *Gist: Education and Learning Research Journal,* (12), 29-50.

Harous, S., & Al Harahsheh, A. R. (2017, December). An app for language learners. In *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)* (pp. 436-440). IEEE. doi:10.1109/ICTUS.2017.8286047

Islam, R., Islam, R., & Mazumder, T. (2010). Mobile application and its global impact. *IACSIT International Journal of Engineering and Technology*, *10*(6), 72–78.

Kearney, M., Burden, K., & Rai, T. (2015). Investigating teachers' adoption of signature mobile pedagogies. *Computers & Education*, *80*, 48–57. doi:10.1016/j.compedu.2014.08.009

Kearney, M., Burden, K., & Schuck, S. (2019). Disrupting education using smart mobile pedagogies. In *Didactics of smart pedagogy* (pp. 139–157). Springer. doi:10.1007/978-3-030-01551-0_7

Kearney, M., Schuck, S., Burden, K (in press). *Theorising and implementing mobile learning: Using the iPAC framework to inform research and practice*. Springer.

Kearney, M., Schuck, S., Burden, K., & Aubusson, P. (2012). Viewing mobile learning from a pedagogical perspective. *Research in Learning Technology*, *20*(1), n1. doi:10.3402/rlt.v20i0.14406

Keogh, C., & Robles, H. (2018). *WhatsApp as a Site for Meaningful Dialogue*. International Association for Development of the Information Society.

Kim, H., & Kwon, Y. (2012). Exploring smartphone applications for effective mobile-assisted language learning. *Multimedia-Assisted Language Learning*, *15*(1), 31–57. doi:10.15702/mall.2012.15.1.31

Kutner, M., Greenberg, E., & Baer, J. (2006). *A First Look at the Literacy of America's Adults in the 21st Century. NCES 2006-470*. National Center for Education Statistics.

Lai, A. (2016). Mobile immersion: An experiment using mobile instant messenger to support second-language learning. *Interactive Learning Environments*, *24*(2), 277–290. doi:10.1080/10494820.2015.1113706

Lee, C. Y., & Sloan, T. (2015). A comprehensive evaluation rubric for assessing instructional apps. *Journal of Information Technology Education*, *14*, 14. doi:10.28945/2097

Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer-assisted language learning*. Lawrence Erlbaum.

Li, J., & Deng, Q. (2018). What influences the effect of texting-based instruction on vocabulary acquisition? Learners' behavior and perception. *Computers & Education*, *125*, 284–307. doi:10.1016/j.compedu.2018.06.017

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, *31*(3), 293–311. doi:10.1017/S0958344019000065

López-Hernández, F. A., & Silva-Pérez, M. M. (2016). *Factores que inciden en la aceptación de los dispositivos móviles para el aprendizaje en educación superior*. Academic Press.

Lotherington, H. (2018). Mobile language learning: The medium is^ not the message. *Journal of Linguistics and Language Teaching*, *10*(2). Advance online publication. doi:10.5070/L210235576

Martín-Monje, E., Arús, J., Rodríguez-Arancón, P., & Calle-Martínez, C. (2014). *REALL: Rubric for the evaluation of apps in language learning*. Academic Press.

Murray, O. T., & Olcese, N. R. (2011). Teaching and learning with iPads, ready or not? *TechTrends*, *55*(6), 42–48. doi:10.1007/s11528-011-0540-6

OECD. (2018). *PISA 2018. Drop in literacy rates*. OECD Publishing.

OECD. (2019). *PISA 2019. Expanding and upgrading upper secondary education*. OECD Publishing.

Olmedo, K., Grané, M., & Crescenzi, L. (2012). Uso y percepciones de uso de dispositivos móviles. Una visión desde la triangulación metodológica. In Investigación presentada en Comunicación y riesgo. In *III Congreso Internacional de la Asociación Española de Investigación de la Comunicación*. AE-IC.

Ozdamli, F., & Cavus, N. (2011). Basic elements and characteristics of mobile learning. *Procedia: Social and Behavioral Sciences*, *28*, 937–942. doi:10.1016/j.sbspro.2011.11.173

Pachler, N., Bachmair, B., Cook, J., & Kress, G. (2010). Mobile learning. New York, NY: Springer. doi:10.1007/978-1-4419-0585-7

Papadakis, S., & Kalogiannakis, M. (2017). Mobile educational applications for children: What educators and parents need to know. *International Journal of Mobile Learning and Organisation*, *11*(3), 256–277. doi:10.1504/IJMLO.2017.085338

Papavlasopoulou, S., Giannakos, M. N., & Jaccheri, L. (2019). Exploring children's learning experience in constructionism-based coding activities through design-based research. *Computers in Human Behavior*, *99*, 415–427. doi:10.1016/j.chb.2019.01.008

Pegrum, M. (2014). *Mobile learning: Languages, literacies and cultures*. Springer. doi:10.1057/9781137309815

Pegrum, M. (2019). Mobile Learning Experiences. In *Mobile Lenses on Learning* (pp. 207–248). Springer. doi:10.1007/978-981-15-1240-7_6

Powell, S. (2014). Choosing iPad apps with a purpose: Aligning skills and standards. *Teaching Exceptional Children*, *47*(1), 20–26. doi:10.1177/0040059914542765

Radinsky, J., Bouillion, L., Lento, E. M., & Gomez, L. M. (2001). Mutual benefit partnership: A curricular design for authenticity. *Journal of Curriculum Studies*, *33*(4), 405430. doi:10.1080/00220270118862

Rambe, P., & Bere, A. (2013). Using mobile instant messaging to leverage learner participation and transform pedagogy at a South African University of Technology. *British Journal of Educational Technology*, *44*(4), 544–561. doi:10.1111/bjet.12057

Reeves, T. (1994). Evaluating what really matters in computer-based education. In M. Wild & D. Kirkpatrick (Eds.), *Computer education: New perspectives* (pp. 219–246). MASTEC.

Reeves, T. (2006). Design research from a technology perspective. In *Educational design research* (pp. 64–78). Routledge.

Reveiu, A., Smeureanu, I., & Dardala, M. (2009). Generating multimedia components for m-learning. *Informações Econômicas*, *13*(3), 88–95.

Rosell-Aguilar, F. (2017). State of the app: A taxonomy and framework for evaluating language learning mobile applications. *CALICO Journal*, *34*(2).

Schuck, S. R., Aubusson, P. J., Kearney, M. D., & Burden, K. (2013). Mobilising teacher education: A study of a professional learning community. *Teacher Development*, *17*(1), 1–18. doi:10.1080/13664530.2012.752671

Serdyukov, P. (2017). Innovation in education: what works, what doesn't, and what to do about it? *Journal of Research in Innovative Teaching & Learning*.

Sha, L., Looi, C. K., Chen, W., & Zhang, B. H. (2012). Understanding mobile learning from the perspective of self-regulated learning. *Journal of Computer Assisted Learning*, *28*(4), 366–378. doi:10.1111/j.1365-2729.2011.00461.x

Stevenson, M., Hedberg, J., Highfield, K., & Diao, M. (2015). Visualizing solutions: Apps as cognitive stepping-stones in the learning process. *The Electronic Journal of e-Learning, 13*(5), 366-379. https://eric.ed.gov/?id=EJ1084237

Sun, Z., Lin, C. H., You, J., Shen, H. J., Qi, S., & Luo, L. (2017). Improving the English-speaking skills of young learners through mobile social networking. *Computer Assisted Language Learning*, *30*(3-4), 304–324. doi:10.1080/09588221.2017.1308384

Tang, Y., & Hew, K. F. (2017). Is mobile instant messaging (MIM) useful in education? Examining its technological, pedagogical, and social affordances. *Educational Research Review*, *21*, 85–104. doi:10.1016/j.edurev.2017.05.001

Traxler, J., & Kukulska-Hulme, A. (2005). *Evaluating mobile learning: Reflections on current practice*. Academic Press.

Trespalacios, J., Chamberlin, B., & Gallagher, R. R. (2011). Collaboration, engagement & fun: How youth preferences in video gaming can inform 21st century education. *TechTrends*, *55*(6), 49–54. doi:10.1007/s11528-011-0541-5

Tudge, J. (1992). *Vygotsky, the zone of proximal development, and peer collaboration: Implications for classroom practice*. Academic Press.

Vaala, S., Ly, A., & Levine, M. H. (2015). *Getting a Read on the App Stores: A Market Scan and Analysis of Children's Literacy Apps*. Full Report. Joan Ganz Cooney Center at Sesame Workshop.

Vygotsky, L. S. (1978). *Mind in society*. MIT Press.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, *53*(4), 5–23. doi:10.1007/BF02504682

Zorrilla, M. J. P. (2005). Evaluación de la comprensión lectora: Dificultades y limitaciones. *Review of Education*, 126.

Heydy Robles is a research professor in the Department of Foreign Languages at Universidad del Norte, Barranquilla, Colombia. She holds a PhD in Educational Innovation from Tecnológico de Monterrey, Mexico. She has a degree in Languages from Universidad del Atlántico, Colombia. She also holds a M.A in Education with emphasis in English from Universidad del Norte.

Kevin Burden is an international expert on the application and impact of digital technologies in learning. His work focuses on the transformational uses of mobile technologies in both formal, but increasingly, informal public spaces that include schools, colleges, universities, museums, heritage sites and the wider environment. Professor Burden has led international collaborative projects in mobile learning with partners including the EU, the British Council and UNESCO.

Karen Villalba is a foreign language professor. She holds a Master degree in English teaching from Universidad del Norte. Experience as a virtual tutor and visually impaired student tutorial sessions. Current English teacher for undergraduate programs at Universidad del Norte and Ph.D in Electronic research and technology-enhanced learning candidate at Lancaster University.