

Designing and creating an educational app rubric for preschool teachers

Stamatios Papadakis¹  · Michail Kalogiannakis¹ · Nicholas Zaranis¹

Received: 30 November 2016 / Accepted: 12 January 2017 / Published online: 30 January 2017
© Springer Science+Business Media New York 2017

Abstract International studies indicate that the use of smart mobile devices and their accompanying educational applications (apps) can revolutionize young children’s learning experiences. Although there is a vast array of educational apps for preschoolers, they are not actually educational in their majority. In this context, it is important for preschool teachers to be able to assess each app for its effectiveness in educational practice. To evaluate educational mobile apps, this paper presents a rubric (abbreviated as REVEAC) in four areas: contents, design, functionality, and technical quality, each having multiple aspects. In this paper, we discuss the known problems in using educational apps and we present a review of relevant literature, as well as the process of formulating REVEAC. Finally, the paper concludes with a brief discussion of REVEAC limitations and future work.

Keywords Educational applications (apps) · Smart mobile devices · Rubric · Preschool children · Preschool teachers

✉ Stamatios Papadakis
stpapadakis@gmail.com

¹ Department of Preschool Education, Faculty of Education, University of Crete, Crete, Greece

1 Introduction

Since the launch of touch screen tablets, we have seen enormous growth in the industry of mobile applications (apps) (Kluser 2016a; Papadakis and Kalogiannakis 2017). A mobile app is a software application designed to run on mobile devices such as smartphones and tablet computers (Wikipedia 2016). Preschool¹ educators have all bought into the appeal of apps (Chiong 2012) and as a result, apps are rapidly emerging as a new medium for providing educational content to preschool age children (Chiong 2012; Lee and Cherner 2015; Papadakis et al. 2016a, 2016b; Shoukry et al. 2014). In Apple's and Google's app store educational apps for young children are among the top categories of apps most accessed or purchased by users (Avtar 2014; Apple 2016; Bouck et al. 2016; Cardenal and Lopez 2015; Hutchison et al. 2012).

Although there are now hundreds of media that review apps - websites, blogs, podcasts, and even print media (Big Ideas Machine 2015) - the majority do not use sufficiently scientific criteria. Preschool teachers have a limited number of reliable tools to evaluate the appropriateness of these apps on children's cognitive development (Emeeyou 2012; Kucirkova et al. 2014). There is a pressing need for an evaluation rubric that examines all aspects of educational apps (Lee and Cherner 2015) for preschool age children.

Using recent field research as a theoretical framework we present in this article the process of creating a rubric - as well as the rubric itself - which can be used by teachers to evaluate educational apps targeted at preschool age children.

1.1 The self-proclaimed educational apps

Although software companies currently produce a plethora of educationally relevant apps, many of which are meant to be suited for preschool learning (Richter 2015), for the majority this is misleading and there is no evidence of any learning value (Papadakis and Kalogiannakis 2017). Many educational apps for preschoolers are simply to entertain children (Higgins et al. 2005; Papadakis et al. 2016b). It is easy for developers to promote apps as educational that may not have any educational value for children at all (Guernsey et al. 2012; Kluser 2016a, 2016b; Watlington 2011). For example, most of the apps in Apple's store involve very basic literacy skills, such as letters, phonics and word recognition (Kluser 2016a; Kucirkova 2014a, 2016a). Far fewer apps involved more advanced early reading skills such as comprehension and grammar (Guernsey et al. 2012). Although many software developers know it is important to consult with educational experts, much software still is developed without consideration of key educational factors that may affect learning (Geisert and Futrell 1995).

The majority of apps in today's marketplace can be considered part of the "first wave" of the digital revolution (Hirsh-Pasek et al. 2015). Many of the educational apps (in the form of digital worksheets, games, and puzzles) have interactive yet repetitive game formats with "closed" content, that is, the content cannot not be changed or extended by the user (Flewitt et al. 2014). These games are primarily gamified literacy and numeracy

¹ In this paper the term preschool education is used to refer to the first formal pre-primary educational stage for children 4 to 6 years of age.

apps (Papadakis et al. 2016c), with content presented as a series of interactive tasks, the completion of which is recognized and rewarded with animated multimedia tokens of achievement (Lynch and Redpath 2012). Drill and practice may foster rote learning of facts, but it is not likely to promote deeper conceptual understanding (Hirsh-Pasek et al. 2015). The issue of apps of poor or doubtful educational value is encountered in both free and in most of full and paid-for versions of the apps (Falloon 2013).

Although educators have tried seeking information about apps of interest across app stores and expert review sites or at producers' websites, this kind of research is not sufficient or reliable (Vaala et al. 2015). The reason is that little information on the quality of apps is available beyond the star ratings published on retailers' web pages or digital stores (Stoyanov et al. 2015) and reviewer comments (Bouck et al. 2016). App reviews are subjective by nature and may come from unreliable sources (Bentrop 2014). Only half of the popular paid, free, and award-winning apps in Apple's store provide information about their development teams (Vaala et al. 2015).

Preschool educators need support in identifying quality apps or they risk wasting their time with inferior or inappropriate apps (Cooper 2012; Lee and Cherner 2015; Papadakis et al. 2016b). There is no concrete mechanism for them to determine whether an educational app is developmentally appropriate or not (Shuler 2009a), as studies have repeatedly shown that educators lack the knowledge to evaluate the appropriateness of educational apps for their students (Emeeyou 2012; Hirsh-Pasek et al. 2015; McManis and Parks 2011; Vincent 2012; Papadakis and Kalogiannakis 2017; PRWeb 2012; Watlington 2011; Zaranis et al. 2013). Despite the fact that educators have long been given guidance on which books can help children learn, no such help is on offer when it comes to apps (Kucirkova 2014b). Some educators are advanced and knowledgeable technology users themselves, but this does not mean that they necessarily understand the full implications of ICT products and services when used by young children, and this lack of understanding may, in turn, hinder their effectiveness when educating their students (Ebbeck et al. 2016).

1.2 A literature review of rubrics and frameworks for assessing – developing educational apps

Over the past few decades, there have been various attempts to evaluate the appropriateness of the educational software targeted at young children (Lee and Cherner 2015). Yet, rubrics that focus specifically on educational apps for preschoolers are extremely limited. One problem is that the evaluation criteria used in the majority of the rubrics is not clearly linked to previously conducted research, nor are the rubrics' evaluative dimensions clearly defined (Lee and Cherner 2015). For example, Reeves and Harmon's Systematic Evaluation of Computer-Based Education is widely recognized and adapted by many researchers as a foundation for many rubrics developed in the last two decades (Reeves and Harmon 1994). The fact is that Reeves and Harmon's model does not address some of the new functionalities that smart mobile technology utilizes, such as sharing capability, crossplatform integration, and ability to save progress. Another example is the use of the Developmental Software Scale (DSS) (Haugland 1999), which is another popular tool for evaluating developmentally appropriate computer software for preschool age children. Although Haugland's DSS includes several evaluating criteria such as the age appropriateness of an app, the adjustable

difficulty levels etc., these underline the importance of design without emphasizing personalized learning or the curriculum content compatibility (Chau 2014).

McManis and Parks (2011) created the Early Childhood Educational Technology Evaluation Toolkit, integrating technology framework principles in preschool education as expressed in the joint statement of educational institutions National Association for the Education of Young Children (NAEYC) and Fred Center Rogers (NAEYC and FRC 2012). Their rating form included 20 questions in Likert-scale questionnaire format, in an attempt to investigate whether an application is educational, age appropriate, child-friendly, and pleasant to use. However, as Chau (2014) states, McManis and Parks (2011), as well as Haugland's rating scale (Haugland 1999), mostly deal with the evaluation of software use in a classroom environment, lacking the specialization required for the evaluation of a new form of technology such as the tablet. McManis and Parks (2011) assessment tool aims to evaluate the appropriateness of the content rather than the design features, such as the use of visual and acoustic elements to enhance children's learning. Based on several studies, Chau (2014) created a rating scale, more commonly known as Developmentally Appropriate App Design Evaluation Form that provides a framework for developmentally appropriate design practices for children's educational apps. This framework comprises four basic design principles: interaction, visual, acoustic and instructional design.

Shoukry et al. (2012) created an evaluation framework in an attempt to investigate the effectiveness and suitability of educational games. The assessment framework they designed consisted of 15 categories, such as the screen design, the navigation and application control, the ease of use, the application design, the content presentation, as well as security, accessibility, usability, and cost issues. Hirsh-Pasek et al. (2015) offer a way to define the potential educational impact of current and future apps that belong to the first and the second "wave" of educational apps. Their aim is primarily to guide researchers, educators, and designers in evidence-based app development and secondly to create an evidence-based guide, in order to set a new standard for evaluating and selecting the most effective existing educational apps for children which are compatible with the second wave of app development.

Rodríguez-Arancón et al. (2013) created a rating scale which combines educational and technical criteria equally. In his assessment toolbox Vincent (2012) places great importance on the following five criteria: relevance, customization, feedback, thinking skills, participation and sharing. Walker (2010) is considered a pioneer in mobile application evaluation, as his rubric has been used as a template for subsequent scales by other researchers. Walker's rating scale is based on six criteria such as connection with the curriculum, authenticity, user friendliness etc. As Walker states, his rubric, which was mainly created for the evaluation of apps for iPod-type devices, can determine whether an application is associated with a targeted skill or a curriculum concept (Walker 2010). Buckler (2012) created a rubric with six dimensions to evaluate apps for people with special needs, including domains such as feedback, adjustability, ease of use, cost etc. (Lee and Cherner 2015). Another interesting app evaluation tool has been developed by the website YogiPlay, a mobile learning apps service designed specifically for children from ages 3 to 8 (PRWeb 2012). The YogiPlay evaluation system is based on the following key evaluation axes: design quality, ease of use, learning activities and the content structure. There are also some rating systems by Children's Technology Review, Common Sense Media, and a handful of parent-

oriented app services. Although these rating systems have not been scientifically evaluated, they are widely used in the field (Hirsh-Pasek et al. 2015). Stoyanov et al. (2015), after conducting a survey of the literature (taking into account explicit web or app quality rating criteria published between January 2000 and January 2013), formulated their assessment criteria for the development of the “Mobile App Rating Scale” (MARS). The MARS creators claim that it is a simple, objective, and reliable tool for classifying and assessing the quality of mobile apps. Additionally, it can also be used to provide a checklist for the design and development of high quality apps in terms of customisation, interactivity, functionality, aesthetics etc.

Lee and Cherner (2015) designed one of the most comprehensive rubrics with 24-evaluative dimensions tailored specifically to analyse the educational potential of instructional apps using previously published research. The 24 dimensions are categorized into three domains: (A) Instruction, (B) Design, and (C) Engagement. Although this rubric is designed to be as comprehensive and inclusive as possible, it has limitations. One of the main limitations in our opinion is that individuals must be able to label apps as skills-based, content-based, or function-based depending on their purpose, because instructional apps will score differently on the rubric based on their design.

Most recently, Kucirkova (2016b) published a framework (iRPD) in an attempt to bring together teachers, researchers and designers into one space specifically focused on iPad apps. That iRPD framework aims to provide the community with some thinking tools (five principles) in order to enrich traditional design-based research with novel affordances of twenty-first century technologies. The five principles are triple collaboration, shared epistemology, interconnected social factors, awareness of app affordances and a child-centred pedagogy (Kucirkova 2016b). In an attempt to make educators aware of the various tools available to support app evaluation, Bouck et al. (2016) present 5 different rubrics in their research paper (More and Travers 2013; Ok et al. 2015; Tammara and Jerome 2012; Walker 2010; Weng and Taber-Doughty 2015). For educators considering apps explicitly for students with learning disabilities the researchers propose the rubric by Ok et al. (2015).

2 Method

2.1 Draft rubric structure

The axes on which the initial construction of the rubric was based (criteria, levels of quality gradations, scoring strategy) are the following:

- Educational research. Early evidence indicates that children can learn from well-designed educational apps (as mentioned in the literature review). Children learn best when they are cognitively active and engaged, when learning experiences are meaningful and socially interactive, and when learning is guided by a specific goal (Hirsh-Pasek et al. 2015). Educational apps must provide unprecedented opportunities for children to create their own content and participate in rich and dynamic learning contexts (Kucirkova 2014c, 2015).
- App Rating Scales. Within the past five years, various app assessment tools have been developed (as mentioned in the literature review). The majority of rubrics lack

the specialization required for the evaluation of this new form of technology for preschoolers, or they do not emphasize all the features required for an educational app (see Table 1).

- Evaluation of the researchers. The apps which were available at the time of research on Play Google, had Greek content and targeted preschoolers.

More specifically, the procedure followed by the researchers for the evaluation of the apps was as follows: at first the researchers visited Google's online store (Play Google) and, based on the categorization of the apps and the use of keywords (such as preschool and / or game, education, children etc.), identified the free apps which targeted preschool age children. Randomly they chose 20 of these educational apps and installed them two different types of smart devices (smartphone & tablet). Subsequently, the researchers played each application separately until they reached the end. As in other studies, this phase focused on the technical and design specifications of the apps, as well as their pedagogic goals, in a more general sense. No in-depth methodological analysis of any particular app was therefore intended at this stage (Rodríguez-Aracón et al. 2013).

The researchers noted that the majority of applications were in edutainment format (matching cards, jigsaw puzzles). Learning goals could be attributed to drill and practice exercises. There were apps which were in the form of interactive electronic books, ("Read to Me" type). The questions were presented to children mainly as multiple choice or closed type and children could learn by the trial and error method. All apps were based on low level-thinking skills and did nothing more than promote rote learning instead of meaningful learning. Also, they tended to evaluate knowledge rather than introduce new concepts. In any case, they didn't encourage exploration, experimentation, problem solving, and creative thinking. The apps targeted literacy (70%) (alphabet learning: letter and sound recognition) and mathematics (30%) (numbers and basic operations).

As regards the app designs, the researchers found that the majority were not created based on best design practices (Sesame Workshop 2012). In brief, no app included an on-screen character, there was a low quality use of multimedia elements (audio, image), a lack of "palm rest", incorrect design of app interface, ineffective use of interactive hotspots, inability to configure, no levelling (e.g. at least three levels: easy, medium and difficult). In several apps, the researchers found in-app purchases, subscriptions, and advertising. No app monitored preschoolers' progress, offered a portfolio system, or content sharing or synchronous play.

2.2 Rubric creation process

2.2.1 First stage

Researchers worldwide have set various standards for the construction of a scale-classified criteria (Allen and Tanner 2006; Moskal 2000; Roblyer and Wiencke 2003). More specifically, a rubric must:

- Include an appropriate and optimal number of criteria. When the number of criteria is very large, the rubric becomes dysfunctional, whereas when the number of

Table 1 Relevant literature and existing app rating scales related to rubric dimensions

Subsector	Relevant literature and existing app rating scales
1.1 Knowledge package appropriateness	Bentrop 2014; Chau 2014; Harrold 2012; Chiong and Shuler 2010; Falloon 2013; Falloon and Khoo 2014; Flewitt et al. 2014; Haugland 1999; Kucirkova et al. 2014; Lee and Cherner 2015; McManis and Parks 2011; Ok et al. 2016; Rodríguez-Arancón et al. 2013; Shoukry et al. 2012; Vincent 2012; Walker 2010; Wartella 2015; Weng and Taber-Doughty 2015
1.2 Learning provision	Chau 2014; Chiong and Shuler 2010; Falloon 2013; Flewitt et al. 2014; Hirsh-Pasek et al. 2015; Kucirkova et al. 2014; Lynch and Redpath 2012; McManis and Parks 2011; Rodríguez-Arancón et al. 2013; Shoukry et al. 2012; Wartella 2015
1.3 Levelling	Buckler 2012; Chau 2014; Chiong and Shuler 2010; Falloon 2013; Fu et al. 2009; Hirsh-Pasek et al. 2015; Haugland 1999; Hutchison et al. 2012; Lynch and Redpath 2012; Rodríguez-Arancón et al. 2013; Shoukry et al. 2012; Shuler 2009b; Tammaro and Jerome 2012; Wartella 2015; Weng and Taber-Doughty 2015
1.4 Motivation - Engagement	Bentrop 2014; Chau 2014; Chiong and Shuler 2010; Falloon 2013; Falloon and Khoo 2014; Hirsh-Pasek et al. 2015; Hutchison et al. 2012; Kucirkova et al. 2014; McManis and Parks 2011; Ok et al. 2016; Sesame Workshop 2012; Sharples and Beale 2002; Stoyanov et al. 2015; Vincent 2012; Walker 2010; Wartella 2015
1.5 Error correction and feedback provision	Bentrop 2014; Buckler 2012; Falloon 2013; Fu et al. 2009; Hirsh-Pasek et al. 2015; Lee and Cherner 2015; McManis and Parks 2011; Ok et al. 2016; Sesame Workshop 2012; Shoukry et al. 2012; Tammaro and Jerome 2012; Vincent 2012; Walker 2010; Weng and Taber-Doughty 2015
1.6 Progress Monitoring/Sharing	Bentrop 2014; Chau 2014; Lee and Cherner 2015; McManis and Parks 2011; Ok et al. 2016
1.7 Bias free	Chau 2014; Chiong and Shuler 2010; Haugland 1999; Shoukry et al. 2012
2.1 Graphics	Chau 2014; Chiong and Shuler 2010; Falloon 2013; Ok et al. 2016; Sesame Workshop 2012; Sharples and Beale 2002; Shoukry et al. 2012; Tammaro and Jerome 2012
2.2 Sound	Chau 2014; Chiong and Shuler 2010; Falloon 2013; Ok et al. 2016; Sesame Workshop 2012; Sharples and Beale 2002; Shoukry et al. 2012; Tammaro and Jerome 2012
2.3 Layout / scenery	Chau 2014; Chiong and Shuler 2010; Hirsh-Pasek et al. 2015; Rodríguez-Arancón et al. 2013; Sesame Workshop 2012; Sharples and Beale 2002; Shoukry et al. 2012; Tammaro and Jerome 2012; Weng and Taber-Doughty 2015
2.4 App/menu design	Chau 2014; Chiong and Shuler 2010; Falloon 2013; Falloon and Khoo 2014; Hirsh-Pasek et al. 2015; Ok et al. 2016; Sesame Workshop 2012; Sharples and Beale 2002; Shoukry et al. 2012; Shuler 2009a, 2009b; Tammaro and Jerome 2012; Weng and Taber-Doughty 2015
3.1 Child friendly	Bentrop 2014; Buckler 2012; Chau 2014; Chiong and Shuler 2010; Hirsh-Pasek et al. 2015; McManis and Parks 2011; Ok et al. 2016; Sesame Workshop 2012; Shoukry et al. 2012; Shuler 2009a, 2009b; Walker 2010
3.2 Child autonomy	Chau 2014; Chiong and Shuler 2010; Haugland 1999; Kucirkova et al. 2014; McManis and Parks 2011; Sesame Workshop 2012; Weng and Taber-Doughty 2015
3.3 Instructions	Falloon 2013; Haugland 1999; Sesame Workshop 2012; Shoukry et al. 2012
3.4 Configuration ability / personalization	Bentrop 2014; Chiong and Shuler 2010; Ok et al. 2016; Sesame Workshop 2012; Shoukry et al. 2012; Shuler 2009b; Vincent 2012
4.1 Performance and Reliability	Chiong and Shuler 2010; McManis and Parks 2011; Rodríguez-Arancón et al. 2013; Sesame Workshop 2012
4.2 Electronic Transactions / Advertisements	Chiong and Shuler 2010; Hirsh-Pasek et al. 2015; Sesame Workshop 2012; Shoukry et al. 2012;
4.3 Social interactions	Chiong and Shuler 2010; Fu et al. 2009; Hirsh-Pasek et al. 2015; Hutchison et al. 2012; Kucirkova et al. 2014; Lynch and Redpath 2012; Vincent 2012; Wartella 2015

criteria is too small, the rubric does not provide sufficient information regarding the evaluated object.

- Have operational criteria and performance-level descriptions. The rating scale should range from one as the worst performance to four as the best performance.
- Contain performance-level descriptions that are as informative as possible.

Considering the relevant literature and based on the findings of the app evaluations, the researchers decided to create a rubric entitled “Rubric for the Evaluation of Educational Apps for preschool Children” (REVEAC) in order to evaluate the quality of educational apps for preschoolers in the following four key areas: educational content, design, functionality and technical characteristics.

After defining the concepts to be measured, the aim in this first stage was to create a list of items and determine the format of measurement. A few key considerations about creating this rubric need to be clarified. First, each of the evaluative dimensions together comprise the entire rubric, and the dimensions discussed in the following sections refer back to it. Second, each evaluative dimension was designed to follow a consistent format. The format includes a prompt that focuses the dimension on a central question, and four indicator descriptors that describe the ways in which an app’s content, functionality or design may behave in response to a prompt. Table 2, illustrates the first row in the rubric. The way in which we proceeded was to first fill in the cell corresponding to the maximum rating, i.e. 4 (exemplary) in which all the sub-criteria were fulfilled, and gradually decrease the sub-criteria that were fulfilled as we moved down the scale, until the minimum rating, i.e. 1 (unsatisfactory/poor), was reached, where none of the sub-criteria was fulfilled.

After the first draft of the app evaluation rubric was created, the scale was distributed to a sample that consisted of preschool teachers and undergraduate students in the department of preschool education. Prior to this, the authors received Institutional Review Board (IRB) approval. The researchers wanted to check whether the proposals of the scale created interpretation problems for people who were not familiar with the

Table 2 Learning provision sub criterion in the educational app evaluation rubric

Evaluation Sector	Score			
	1 Unsatisfactory / Poor	2 Needs improvement	3 Good quality	4 Exemplary
Educational content: Learning provision	Learning is provided with an emphasis on information presentation, (typically of eBook or memory flashcards format). The app focuses on knowledge evaluation through drill and practice activities based on mechanical memory	Learning is mainly provided through drill and practice activities but includes some activities in simulation game format. Children are rarely called upon to apply what they learn	Learning is provided through an authentic learning environment. Children partially act on and discover knowledge on their own, based on prior knowledge.	The app emphasizes discovery learning, creative thinking and problem solving. Children deal with problems in a familiar context, experiencing situations and practices through which they learn.

use of educational apps for preschoolers. After semi-structured interviews and first stage data analysis, both the structure of some sentences, as well as some terms or expressions that were originally used on the scale were modified, merged, or removed in order to keep pace with the sample observations. For example, the preschool teachers proposed that when selecting educational software, the content and features should be appropriate not only for the targeted student's chronological age, but also for a student's developmental age, and suggested merging two subsectors into one (Age and Learning package to Knowledge package appropriateness). In summary, nine subsectors were merged or removed (Educational content sector: Age and learning level, Critical Thinking, Multiple Learning Styles Support, Evaluation Existence, Design sector: Scenery, Multimedia Elements Usage, Functionality sector: Interactivity, Technical Characteristics: Update mode, Compatibility). Two subsectors were moved and merged from the Functionality to Technical Characteristics sector (Electronic Transactions, Advertisements) (See Fig. 1).

Based on this procedure, the draft rubric was redesigned so as to facilitate the app evaluation process. The rubric's sub-axes were as follows:

- The educational content section consists of seven subsectors: knowledge package appropriateness, learning provision, levelling, motivation/engagement, error correction/feedback provision, progress monitoring/sharing and bias free
- The design section consists of four subsectors: graphics, sound, layout/scenery and app/menu design.
- The functionality section consists of four subsectors: child friendliness, autonomy, instructions and customization.
- The technical characteristics section consists of three subsectors: performance and reliability, advertising / electronic transactions and social interactions.

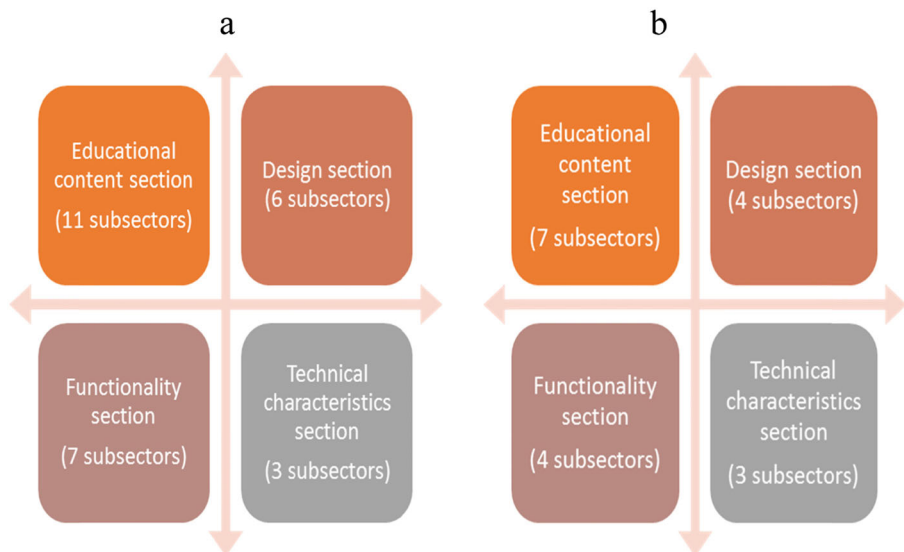


Fig. 1 The axes and sub axes of the rubric before (a) and after the completion of the first phase (b)

Calculating the scores of the subscales and an overall app total score is how the REVEAC is scored.

2.2.2 Second stage

A rubric should be shown to have sufficient validity and reliability to establish its usefulness in clarifying expected performance (Roblyer and Wiencke 2003). According to Taggart et al. (1999) rubric content validity can be improved by involving experts in its development. After the rubric was drafted, two experts (experienced instructional technology professionals) were asked to evaluate five different apps using the evaluation tool created by the researchers in the first stage for assessment of inter-rater reliability. The experts were provided with a form to complete ratings on each app, and then answering a few short-answer questions about the rubric. Their feedback was used to improve the clarity and comprehensiveness of rubric elements (Roblyer and Wiencke 2003).

Modifications were then made to the rubric based upon the experts' suggestions (wording and relevance). The apps that were evaluated were selected randomly by the researchers. They were listed in Google's digital store at the time of the research (April, 2016), where available in the Greek language, targeted at preschool-age children, were free and included the following: Play & Learn – Kindergarten, Kids ABC Numbers & Colors free, Greek Kindergarten Lite, Learn the numbers, Infant Tasks.

After the revised rubric was finalized, it was presented to the experts again to ensure that they understood each of the rubric's dimensions and indicators. After reviewing the revised rubric, all the experts confirmed that its clarity had increased and that they understood each of the rubric's dimensions and indicators. The inter-rater reliability between the two raters was analysed using Spearman's correlations (see Table 3). Data were analysed with IBM SPSS Statistics version 23.

The scale had a high level of internal consistency (Cronbach alpha = .79) and the average inter-rater reliability was $rs(5) = .72, p < .01$. Progress monitoring/sharing, $rs(5) = .91, p < .01$, Error correction /feedback provision, $rs(22) = .92, p < .01$, and Levelling, $rs(5) = .82, p < .01$, had the strongest correlations. All other correlations between individual items ranged from .58 to .78 and were large (Shrout and Fleiss 1979).

We analysed convergent validity using Pearson correlations. Each of the scores assigned on the rubric as well as an overall average score was compared to the scores given by researchers to the apps. Researchers used the rubric created by Lee and Cherner (2015) to rate the apps that were unrelated to this rubric. Researchers chose to use this rubric because they found it to be the most comprehensive rubric tailored specifically to analyse the educational potential of instructional apps. Results show that overall the rubric correlated with actual scores with a mean of $r(5) = .52, p < .01$. Individual items varied, but the Content appropriateness, $r(5) = .74, p < .01$, Motivation, $r(5) = .63, p < .01$, Error correction /feedback provision, $r(5) = .71, p < .01$, Progress monitoring/sharing, $r(5) = .65, p < .01$, Bias free, $r(5) = .67, p < .01$, Layout / scenery, $r(5) = .61, p < .01$, Child-friendliness $r(5) = .65, p < .01$, and Social interactions $r(5) = .67, p < .01$, yielded significant correlations. All other individual item correlations were not significant.

Table 3 Mean, Standard Deviation, Inter-Rater Reliability and Convergent Validity

		M	SD	IRR <i>rs</i>	CV <i>rp</i>
1.1	Knowledge package appropriateness	2.1	1.2	.58*	.68*
1.2	Learning provision	1.7	1.3	.78*	.35
1.3	Levelling	1.6	1.1	.82*	.48
1.4	Motivation / Engagement	2.4	1.1	.61*	.63*
1.5	Error correction /feedback provision	2.9	1.2	.89*	.71*
1.6	Progress monitoring/sharing	1.1	1.1	.91*	.65*
1.7	Bias free	3.2	1.2	.59**	.67*
2.1	Graphics	2.7	1.3	.60*	.48
2.2	Sound	2.6	1.1	.61*	.62
2.3	Layout / scenery	2.4	1.4	.60*	.61*
2.4	App/menu design	2.1	1.5	.58*	.45
3.1	Child friendliness	2.2	1.1	.59**	.65*
3.2	Autonomy	3.1	1.1	.60*	.48
3.3	Instructions existence	2.1	1.1	.61*	.62
3.4	Configuration ability	1.1	1.1	.60*	.51
4.1	Performance and Reliability	3.3	1.1	.59**	.45
4.2	Advertisements / Electronic Transactions	1.1	1.1	.60*	.44
4.3	Social interactions	1.1	1.1	.74*	.67*

* $p < .01$ ** $p < .05$

2.2.3 Third stage

For this stage, the researchers chose a new sample of five apps (Listen, find, learn; Place in order of size; Find the different thing; Write the word; Days, months, seasons), with the same criteria as the second stage, and installed them in a same type of mobile devices (a 10-in. tablet combined with the latest Android OS, Lollipop). Then, researchers recruited ten raters, five of whom were preschool educators and five students of the department of preschool education. Participation in the rubric evaluation was voluntary. The completed evaluations were anonymous with no personal information or identifying factors.

Independent ratings on the overall REVEAC total score of the 5 apps demonstrated an excellent level of inter-rater reliability (2-way mixed $ICC = .81$, 95% CI 0.79–0.87) (Landers 2015). The REVEAC total score had excellent internal consistency (Cronbach alpha = .85). Internal consistencies of the REVEAC subscales were also very high (Cronbach alpha = .80–.89, median .85), and their inter-rater reliabilities were fair to excellent ($ICC = .50$ –.80, median .65). Detailed item and subscale statistics are presented in Table 4.

Correlations also were done between scores on rubric ratings and the sample's experience in using mobile and educational apps to determine whether the sample's evaluation results were related to their experience as users of mobiles and/or educational apps. Similarly, correlations were done between the evaluation results and the sample's preference for mobile devices, gender, etc. No significant correlations were found between any of these variables. Although the number of apps so far evaluated is still too small to statistically measure the raters' agreement, the results shown in Table 4 do

Table 4 Results of the analysis of the REVEAC items

Subscale/item	Corrected item-total correlation	Mean	SD
1. Educational content $\alpha = 0.88$, ICC = 0.81 (95% CI 0.72–0.84)			
1.1 Knowledge package appropriateness	.65	1.71	0.69
1.2 Learning provision	.71	2.15	1.10
1.3 levelling	.64	1.44	0.98
1.4 Motivation / Engagement	.66	3.11	1.50
1.5 Error correction /feedback provision	.62	1.32	0.52
1.6 Progress monitoring/sharing	.71	1.45	0.85
1.7 Bias free	.69	3.34	0.96
2. Design $\alpha = 0.89$, ICC = 0.84 (95% CI 0.73–0.86)			
2.1 Graphics	.58	2.91	1.12
2.2 Sound	.69	2.15	0.84
2.3 Layout / scenery	.74	1.99	0.66
2.4 App/menu design	.81	2.14	1.11
3. Functionality $\alpha = 0.91$, ICC = 0.87 (95% CI 0.76–0.92)			
3.1 Child-friendliness	0.84	3.16	0.99
3.2 Autonomy	0.54	3.22	0.87
3.3 Instructions	0.49	1.25	0.66
3.4 Configuration ability	0.51	1.37	0.49
4. Technical characteristics $\alpha = 0.89$, ICC = 0.83 (95% CI 0.75–0.89)			
4.1 Performance and Reliability	0.49	2.54	0.65
4.2 Advertisements / Electronic Transactions	0.62	2.48	0.75
4.3 Social interactions	0.68	1.19	0.54

seem to show consistency between the evaluators and, therefore, allow us to be optimistic as to the usability of the rubric. The REVEAC is presented in Appendix Table 5.

3 Discussion and conclusions

We cannot insulate children from technology, but we need to ensure that they are not harmed in any way by it (Ebbeck et al. 2016). As Parette et al. (2010 p.2) state “*with regard to specific technology applications available to young children, the issue is not, however, whether technology should be considered and used in education settings but how and whether it makes a difference in children’s learning and development*”.

3.1 Implications of this study for preschool teachers

While electronic toys and video games were introduced to children as early as 1972, the release of the iPad in 2010 precipitated a dramatic new shift toward digital activities by young children (Huber et al. 2016). Applications on touchscreen devices have the potential of revolutionizing early childhood education and helping to build a stronger base for lifelong learning in the twenty-first century (Papadakis and Kalogiannakis 2010). However, this momentum needs to be systematically exploited since it is not certain that just putting digital devices and apps in the hands of young children-students will assist them in various different aspects of their development. The literature review shows that only fragmented efforts have been made to create principles that will be applicable to the construction of mobile educational apps for preschool age children.

There is an immediate need for apps that go beyond the examples of “skill & drill” apps, which currently dominate a great part of the preschool mobile educational market. Only if educational researchers closely collaborate with the app producers and educators, can this become a source of inspiration for improved app designs and a novel way of implementing research insights into practice (Kucirkova 2016b).

The aim of this study was to develop an educational preschool app evaluation instrument (REVEAC) that incorporated research-based evidence and practitioner expertise. We constructed this assessment rubric under the guidelines of previous papers on rubric development and standardization, but also reflecting the fact that apps are very much dependent on technology, and should therefore not be evaluated from an exclusively pedagogical perspective. We have the notion that the rubric we propose in this paper could serve as a comprehensive tool, helping preschool teachers to select high quality mobile educational apps. It can also be a useful tool in the hands of software developers and designers, providing a flexible framework and offering guidelines for designing developmentally appropriate educational applications for preschool-age children.

3.2 Limitations of the study and future research directions

Although this rubric was designed to be as comprehensive as possible, following the latest developments in the pedagogical and technological area, there are several limitations to its use. One limitation of the study is a lack of inter-rater reliability, as the raters only coded each manuscript once. Another limitation is the relatively small sample and the type of apps (only free – only in the Greek language) analysed for convergent validity due to the restrictions we had set for the selection of the sample (apps). Both experts asked that an N/A (Not applicable) option be included on some rubric subscales as some apps by design are not assessable on all dimensions. On the other hand, the existence of an N/A option should be problematic as the users would find it difficult to determine when lack of information within an app should be rated as N/A or as a flaw. In a new revised version of the rubric we have to consider including an N/A option on some rubric subscales. The REVEAC was piloted on Android, rather than iOS apps. Although the scale has been partially applied to multiple iOS apps and no compatibility issues were encountered, future research should explore the reliability of the REVEAC with iOS apps both free and paid. A further analysis would be interesting for a correlation between the iTunes or Google star rating system and the total rubric score.

As Lee and Cherner (2015) state, creating a single rubric to evaluate all varieties of educational apps is not possible. Considering this, it is necessary to create comprehensive rubrics for the evaluation of different types of educational apps. On the other hand, we have the notion that the rubric presented here, is already too comprehensive for educators and probably this limitation will affect how the rubric can be used, especially outside the formal school environment. With this in mind, it is necessary to create two rubrics adapted for different audiences: parents and preschool teachers. In the future, we intend to create a rubric which will be primarily aimed at parents of preschool age children.

Compliance with ethical standards

Conflict of interest None declared.

Appendix

Table 5 Rubric for the Evaluation of Educational Apps for preschool Children[™] - REVEAC

Evaluation Sector	Score			
	1	2	3	4
	Unsatisfactory / Poor	Needs improvement	Good quality	Exemplary
Knowledge package appropriateness	Existence of multiple errors. Failure to comply with learning object. The content is not appropriate for the targeted student's chronological age.	Incomplete compliance with learning object. The app is occasionally too easy or difficult for child age.	Accurate, mostly suitable for use. It is associated with the learning object. Potentially some of its parts are very easy or difficult.	Accurate, error-free, suitable for use. It responds excellently to the learning object. The content is appropriate for the targeted student's chronological and developmental age.
Learning provision	Learning is provided with an emphasis on information presentation, (typically of eBook or memory flashcards format). The app focuses in knowledge evaluation through drill and practice activities based on mechanical memory	Learning is mainly provided through drill and practice activities but includes some activities in simulation games format. The child rarely called upon to apply what it learns	Learning is provided through an authentic learning environment. The child partially acts and discovers the knowledge on its own based on its prior knowledge.	The app emphasizes on discovery learning, creative thinking and problem solving. The child deals with problems in a familiar context, experiencing situations and practices through which it conquers learning.
Levelling	Lack of multiple difficulty levels-settings. The app does not adjust functionalities to better suit child needs.	One difficulty level. Minimum child experience adaptation.	At least three difficulty levels (e.g. easy, medium, hard). Customizable to some degree.	The app offers multiple levels of difficulty. It provides personalized learning experiences.
Motivation / Engagement	Does not encourage the child. Activities are not enjoyable and/or interesting, motivating child to learn.	Activities rarely keeping learner motivated. The child may be engaged in a learning activity for about 2–3 min.	Motivates learning. The child is involved in substantial learning activity (more than 5 min).	Provides increased incentives for learning. The child is engaged in a meaningful learning activity. Emotionally excite the child.
	Without feedback.	Limited feedback.		

Table 5 (continued)

Score		Score	
Evaluation Sector	1 Unsatisfactory / Poor	2 Needs improvement	3 Good quality
			4 Exemplary
Error correction / Feedback provision		Feedback provision that enhances learning.	Feedback provision that enhances and supports learning. Frequent rewards, punishments absence, immediate support, and assistance.
Progress Monitoring / Sharing	Child progress monitoring does not exist.	Child progress monitoring is not sufficient. Does not provide adequate evidence of its progress. Does not retain progress history.	Complete child progress monitoring and analysis. Keeps child progress history.
Bias free	App is clearly gender, race, ethnicity, culture biased.	The app contains direct references to gender, cultural or ethnic bias.	App content is free from gender, racial, ethnic and cultural stereotypes.
Graphics	Poor quality images, not attractive graphics	Medium quality images, not so attractive graphics	High-quality images, very attractive graphics.
Sound	Poor sound quality. Unnecessary sounds, such as background music which it distracts child attention.	Moderate sound quality. Slightly distracting.	Excellent sound quality. Enhancing the learning experience.
Layout / scenery	Layout elements are confusing. Is often unreasonable and unpredictable. Incorrect use of color and contrast for text and/or graphics.	The layout is not intuitive or often is unreasonable. Text or graphics are hard to read or distinguish as they are often incompatible with the background	Clearly visible elements, easy to use, efficient and responsive layout.
App/menu design	Incorrect menu design, which complicates app handling. There is a lack of design consistency	App/menu design is partially correct. There are inconsistencies or errors in the design.	Proper menu design, which facilitates app handling. There is a design consistency between

Table 5 (continued)

Score		2		3		4	
Evaluation Sector	Unsatisfactory / Poor	Needs improvement	Good quality	Exemplary			
Functionality	Child-friendliness	Partially difficult to understand, learn and use. Lack of “palm rest”.	Partially difficult to understand, learn and use. Easy to understand, learn and use. Lack of “palm rest”.	the different levels / screens of the app.			
	Autonomy	The child needs support or supervision by an adult for app use.	The child periodically needs support or supervision by an adult.	The child may need support in some parts of the app.	The child can deal with the app without having the need of an adult, at least after the first time.		
	Instructions existence	Complicated, incomplete, inaccurate or incomprehensible instructions.	Complicated or not clear enough instructions.	The instructions are relatively easy and complete.	The instructions are simple and complete.		
	Configuration ability	None. App settings cannot be altered from their default values.	Notification sounds, audio, and some messages may be disabled.	New messages, sounds or other features of the app can be added.	All app elements are fully configurable.		
Technical Characteristics	Performance and Reliability	Freezes, running slow, crashes or hangs frequently. A lot of technical issues.	Loading and running slow, crashes frequently. Some technical issues.	Loading and running quickly. Minor technical issues.	Loading and running quickly. It is reliable.		
	Advertisements / Electronic Transactions	Ads existence that completely disrupt child’s attention. The child is forced to conduct electronic transactions	Ads existence that partially disrupt child’s attention. The app indirectly encourages the child in conducting electronic transactions.	Ads existence which doesn’t disrupt child’s attention. The app uses unattractive graphics for the payments section but it has not a separate parental control section.	No Advertisements. The app contains a separate (parent region) section with an additional “Baby gate” window for performing transactions.		
	Social interactions	The app doesn’t provide a potentially social context.	The app partially allows collaborative visual communication through typing, drawing, or interacting with virtual objects	Enable some forms of social interaction and/or synchronous play and/or screen share.	Enable social interaction: face to face interactions, mediated interactions through technologies, on-screen characters (2-way interactions with children).		

References

- Allen, D., & Tanner, K. (2006). Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE-Life Sciences Education*, 5(3), 197–203.
- Apple (2016). iPad in Education. <http://www.apple.com/education/apps-books-and-more/>. Accessed 25 Oct 2016.
- Avtar, T. (2014). How app markets are adding a lot of value in education, think tab blogs. <http://goo.gl/D1gCMp>. Accessed 17 Oct 2016.
- Bentrop, S. M. (2014). Creating an educational app rubric for teachers of students who are deaf and hard of hearing. *Independent Studies and Capstones*. Paper 680. Program in Audiology and Communication Sciences, Washington University School of Medicine. http://digitalcommons.wustl.edu/pacs_capstones/680. Accessed 28 Sept 2016.
- Big Ideas Machine. (2015). 2015 App reviewer survey. <http://bigideasmachine.com/app-reviewer-survey/>. Accessed 19 Oct 2016.
- Bouck, E. C., Satsangi, R., & Flanagan, S. (2016). Focus on inclusive education: evaluating apps for students with disabilities: supporting academic access and success. *Childhood Education*, 92(4), 324–328.
- Buckler, T. (2012). Is there an app for that? Developing an evaluation rubric for apps for use with adults with special needs? *The Journal of BSN Honors Research*, 5(1), 19–32.
- Cardenal, F., & Lopez, V. (2015). Education apps. One step beyond: it is time for something more in the education apps word. In Á. Herrero, B. Baruque, J. Sedano, H. Quintián, & E. Corchado (Eds.), *International Joint Conference: CISIS'15 and ICEUTE'15* (pp. 571–581). Spain: Burgos.
- Chau, C. L. (2014). Positive technological development for young children in the context of Children's mobile apps. In *Doctoral dissertation*. USA: Tufts University.
- Chiong, C. (2012). Living up to the hype: can mobile technology really make kids smarter? *Early Childhood Matters*, 117, 90–95.
- Chiong, C., & Shuler, C. (2010). *Learning: is there an app for that? Investigations of young children's usage and learning with mobile devices and apps*. New York: The Joan Ganz Cooney Center at Sesame Workshop.
- Cooper, R. J. (2012). Hi ConnSENSE fans! RJ Cooper here. ConnSENSE Bulletin. Retrieved from <http://www.connsensebulletin.com/2012/09/hi-connsense-fans-rj-cooper-here/>. Accessed 25 Oct 2016.
- Ebbeck, M., Yim, H. Y. B., Chan, Y., & Goh, M. (2016). Singaporean parents' views of their young children's access and use of technological devices. *Early Childhood Education Journal*, 44(2), 127–134.
- Emeeyou. (2012). Should parents expect more? Making the case for evidence-based early childhood education mobile applications. <http://goo.gl/WwBR25>. Accessed 11 Oct 2016.
- Falloon, G. (2013). Young students using iPads: app design and content influences on their learning pathways. *Computers & Education*, 68, 505–521.
- Falloon, G., & Khoo, E. (2014). Exploring young students' talk in iPad-supported collaborative learning environments. *Computers & Education*, 77, 13–28.
- Flewitt, R., Messer, D., & Kucirkova, N. (2014). New directions for early literacy in a digital age: the iPad. *Journal of Early Childhood Literacy*, 15(3), 289–310.
- Fu, F., Su, R., & Yu, S. (2009). EGameFlow: a scale to measure learners' enjoyment of e-learning games. *Computers & Education*, 52(1), 101–112.
- Geisert, P. G., & Futrell, M. K. (1995). *Teachers, computers, and curriculum*. Boston: Allyn & Bacon.
- Guernsey, L., Levine, M., Chiong, C., & Severns, M. (2012). *Pioneering literacy in the digital wild west: empowering parents and educators*. Washington, DC: Campaign for Grade-Level Reading.
- Harrold, R. (2012). The iPad effect: leveraging engagement, collaboration, and perseverance. <http://www.tieonline.com/article/100/the-ipad-effect-leveraging-engagement-collaboration-and-perseverance>. Accessed 14 Sept 2016.
- Haugland, S. W. (1999). *Developmental software scale*. Cape Girardeau: K.I.D.S. & Computers, Inc..
- Higgins, K., Boone, R., & Pierce, T. B. (2005). Evaluating software for use by students with disabilities to foster inclusion in general education. Paper presented at the inclusive and supportive education congress international special education conference. <http://goo.gl/TKPGiF>. Accessed 14 Sept 2016.
- 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.
- Huber, B., Tarasuik, J., Antoniou, M. N., Garrett, C., Bowe, S. J., Kaufman, J., & Team, T. S. B. (2016). Young children's transfer of learning from a touchscreen device. *Computers in Human Behavior*, 56, 56–64.

- Hutchison, A., Beschoner, B., & Schmidt-Crawford, D. (2012). Exploring the use of the iPad for literacy learning. *The Reading Teacher*, 66(1), 15–23.
- Kluser, C. (2016a). Have children's apps lost their luster? <http://goo.gl/QuMRGJ>. Accessed 24 Oct 2016.
- Kluser, C. (2016b). What's the state of the app economy? *An Interview with Sarah Towle*. <http://goo.gl/9zcVFI>. Accessed 19 Oct 2016.
- Kucirkova, N. (2014a). At what age should we put babies on a digital media diet? <https://goo.gl/DqhO7Q>. Accessed 22 Oct 2016.
- Kucirkova, N. (2014b). How to choose the best educational app for your child. <https://goo.gl/cf67ey>. Accessed 11 Oct 2016.
- Kucirkova, N. (2014c). iPads in early education: separating assumptions and evidence. *Frontiers in Psychology*, 5, 715.
- Kucirkova, N. (2015). Story-making with iPad apps: baking stories in the twenty-first century. *Exchange*, 222, 47–50.
- Kucirkova, N. (2016a). Parents express concerns as more toddlers switch on tablet computers. <https://goo.gl/h9GMmJ>. Accessed 24 Oct 2016.
- Kucirkova, N. (2016b). iRPD—A framework for guiding design-based research for iPad apps. *British Journal of Educational Technology*. doi:10.1111/bjet.12389.
- Kucirkova, N., Messer, D., Sheehy, K., & Panadero, C. F. (2014). Children's engagement with educational iPad apps: insights from a Spanish classroom. *Computers & Education*, 71, 175–184.
- Landers, R. N. (2015). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *NeoAcademic: Technology, education and training research from an industrial/organizational (I/O) psychologist in the ivory tower*: <http://neoacademic.com/2011/11/16/computing-intraclass-correlations-icc-asestimates-of-interrater-reliability-in-spss>. Accessed 2 Oct 2016.
- Lee, C.-Y., & Cherner, T. S. (2015). A comprehensive evaluation rubric for assessing instructional apps. *Journal of Information Education: Research*, 14, 21–53.
- Lynch, J., & Redpath, T. (2012). 'smart' technologies in early years literacy education: a meta-narrative of paradigmatic tensions in iPad use in an Australian preparatory classroom. *Journal of Early Childhood Literacy*, 14(2), 1–28.
- McManis, L. D., & Parks, J. (2011). *Evaluating Technology for Early Learners. Ebook and toolkit*. Winston-Salem: Hatch Early Learning.
- More, C. M., & Travers, J. C. (2013). What's app with that? Selecting educational apps for young children with disabilities. *Young Exceptional Children*, 16(2), 15–32.
- Moskal, B. M. (2000). Scoring rubrics: what, when and how? *Practical Assessment, Research & Evaluation*, 7(3), 1–5.
- NAEYC & Fred Rogers Center for Early Learning and Children's Media. (2012). Technology and interactive media as tools in early childhood programs serving children from birth through age 8. In *Joint position statement*. Washington, DC: NAEYC.
- Ok, M. W., Kim, M. K., Kang, E. Y., & Bryant, B. R. (2015). How to find good apps: an evaluation rubric for instructional apps for teaching students with learning disabilities. *Intervention in School and Clinic*, 51(4), 244–252.
- Papadakis, S., & Kalogiannakis, M. (2010). eTwinning in the early childhood as starting line of innovative practices for the didactic of natural sciences. In M. Kalogiannakis, D. Stavrou, & P. G. Michaelides (Eds.), *HSci 2010: 7th international conference hands-on science "Bridging the Science and Society Gap"*. Greece: University of Crete. July 25–31, 2010
- Papadakis, S., & Kalogiannakis, M. (2017). Mobile educational applications for children. What educators and parents need to know. *International Journal of Mobile Learning and Organisation* (Special Issue on Mobile Learning Applications and Strategies), (in press).
- Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2016a). Comparing tablets and PCs in teaching mathematics: an attempt to improve mathematics competence in early childhood education. *Preschool and Primary Education*, 4(2), 241–253.
- Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2016b). Developing fundamental programming concepts and computational thinking with ScratchJr in preschool education: a case study. *International Journal of Mobile Learning and Organisation*, 10(3), 187–202.
- Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2016c). Improving mathematics teaching in kindergarten with realistic mathematical education. *Early Childhood Education Journal*. doi:10.1007/s10643-015-0768-4.
- Paratte, H. P., Quesenberry, A. C., & Blum, C. (2010). Missing the boat with technology usage in early childhood settings: a twenty-first century view of developmentally appropriate practice. *Early Childhood Education Journal*, 37(5), 335–343.

- PRWeb. (2012). *The YogiMeter*. <http://goo.gl/CuYP6f>. Accessed 12 Oct 2016
- Reeves, T. C., & Harmon, S. W. (1994). Systematic evaluation procedures for interactive multimedia for education and training. In S. Reisman (Ed.), *Multimedia computing: preparing for the twenty-first century* (pp. 472–505). Harrisburg: Idea Group Publishing.
- Richter, F. (2015). Apple's App Store Is Growing by 1,000+ Apps a Day. <https://goo.gl/pMQImo>. Accessed 22 Oct 2016.
- Roblyer, M. D., & Wiencke, W. R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. *The American Journal of Distance Education*, 17(2), 77–98.
- Rodríguez-Arancón P., Arús, J., & Calle, C. (2013). The use of current mobile learning applications in EFL. In *Proceedings of the 13th International Educational Technology Conference -IETC 2013*, (pp. 1219–1232). Kuala Lumpur, Malaysia
- Sesame Workshop (2012). Best Practices: Designing Touch Tablet Experiences for Preschoolers. <http://www.sesameworkshop.org/our-blog/2012/12/17/sesames-best-practicesguide-for-childrens-app-development/>. Accessed 14 Oct 2016.
- Sharples, M., & Beale, R. (2002). Design guide for developers of educational software. *Report produced for BECTA*. http://www.becta.org.uk/technology/software/curriculum/reports_pdf/designguide.pdf. Accessed 2 Oct 2016.
- Shoukry, L., Sturm, C., & Galal-Edeen, G. H. (2012). Pre-MEGa: a proposed framework for the design and evaluation of preschoolers' mobile educational games. In *In the proceedings of The International Conference on Engineering Education, Instructional Technology, Assessment, and E-learning (EIAE 12)*. Bridgeport: USA.
- Shoukry, L., Sturm, C., Galal-Edeen, G. H., & Göbel, S. (2014). Conducting evaluation studies of mobile games with preschoolers. *CEUR Workshop Proceedings*, 1227, 262–269.
- Shrout, P., & Fleiss, J. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428.
- Shuler, C. (2009a). *iLearn: a content analysis of the iTunes app Store's education section*. New York: The Joan Ganz Cooney Center at Sesame Workshop.
- Shuler, C. (2009b). *Pockets of potential: using mobile technologies to promote children's learning*. New York: The Joan Ganz Cooney Center at Sesame Workshop.
- Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth*, 3(1), e27.
- Taggart, G. L., Phifer, S. J., Nixon, J. A., & Wood, M. (Eds.). (1999). *Rubrics: a handbook for construction and use*. Lanham, MD: Scarecrow Press.
- Tammaro, M. T., & Jerome, M. K. (2012). Does the app fit? Using the apps consideration checklist. In M. J. Ault & M. E. Bausch (Eds.), *Apps for all students: a teacher's desktop guide* (pp. 23–31). Reston: Exceptional Innovations.
- 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*. New York: The Joan Ganz Cooney Center at Sesame Workshop.
- Vincent, T. (2012). Ways to evaluate educational apps. <http://learninginhand.com/blog/ways-to-evaluate-educational-apps.html>. Accessed 14 Oct 2016.
- Walker, H. (2010). Evaluating the effectiveness of apps for mobile devices. *Journal of Special Education Technology*, 26(4), 59–63.
- Wartella, E. (2015). Educational apps what We do and do not know. *Psychological Science in the Public Interest*, 16(1), 1–2.
- Watlington, D. (2011). Using iPod touch and iPad educational apps in the classroom. In M. Koehler & P. Mishra (Eds.), *Proceedings of Society for Information Technology & teacher education international conference 2011* (pp. 3112–3114). Chesapeake: AACE.
- Weng, P.-L., & Taber-Doughty, T. (2015). Developing an app evaluation rubric for practitioners in special education. *Journal of Special Education Technology*, 30(1), 43–58.
- Wikipedia (2016). Mobile app. https://en.wikipedia.org/wiki/Mobile_app. Accessed 7 Oct 2016.
- Zaranis, N., Kalogiannakis, M., & Papadakis, S. (2013). Using Mobile Devices for Teaching Realistic Mathematics in Kindergarten Education. *Creative Education*, 4(7A1), 1–10.