CrossMark

Teachers' Assessment of the Instructional Efficacy of Mobile Apps: a Formative Case Study

Robert F. Kenny [2] · Glenda A. Gunter [2] · Laurie O. Campbell [2]

Abstract Integrating console games into educational settings has been increasingly both applauded and criticized. The vast interest in game playing of people of all ages appears to be a motivating source for educators and trainers to find ways to effectively integrate serious games into their educational settings in order to inspire their own students towards the content they are providing in their classrooms. In the past few years, the surge in the interest in using mobile devices in the classroom has followed a similar track to that of console games, perhaps because their constructs appear to mirror those found in those console-based games. These similarities go beyond the coincidental fact that many of the most popular mobile apps also happen to follow game-like patterns. For this reason, the authors suggest that the need exists to examine the educational validity of the instructional apps that are being downloaded onto mobile devices using the same rationale that was previously extended in earlier studies. Using this premise, the researchers based their hypothesis that mobile apps can offer a unique and facile means to "gamify" a classroom. Based on a review of the current state of affairs in commercial off-the-shelf (OTS) apps, there may be too many on the market that profess to be "educational" when, in fact, they do little to actually support acquiring content knowledge. The authors suggest that this possible overabundance has contributed to the need to create a means for game designers to properly and accurately evaluate in a formative way what contributions their games would potentially make to education and training. The current study focuses both evaluating the value of using RETAIN for app designers to formatively assess their potential constructs and to inform educators as to which apps best meet the instructional and student learning needs once those apps enter the market.

Keywords Evaluating mobile apps · Mobile learning · Evaluation rubric for mobile apps

✉ Robert F. Kenny
rkenny@fgcu.edu

[1] Florida Gulf Coast University, Fort Myers, USA

[2] University of Central Florida, Orlando, USA

Introduction

Integrating serious games into educational settings has been both applauded and criticized. The vast interest in game playing by people of all ages appears to be a motivating source for educators and trainers to find ways to effectively integrate serious games into educational settings. They appear to have accepted the idea that the motivational aspects of these games can provide a compelling context to engage their students using a game's interactive constructs (Dominguez et al. 2013). Unfortunately, mixed results continue to be reported in the literature about the efficacy of games as a learning agent, especially as it relates to reaching consensus on how to effectively integrate games or how to consistently measure their instructional effectiveness (Kenny and McDaniel 2011). In addition, the move to integrate games has been slowed by a lack of consistent, positive disposition towards games on the part of teachers (Kenny and Gunter 2011).

In the past few years, a rapid surge in interest in using mobile devices in the classroom has been noted (Atkins et al. 2007). A recent Google search using the phrase "best apps to use in educational settings" resulted in well over 50,000 hits. It is not difficult to assume that educators are most likely overwhelmed with the amount of apparent contradictory information that exists

about the educational efficacy of mobile apps. The authors suggest that the constructs of mobile apps, in many ways, appear to mirror those found in console games due to the fact that many of the most popular mobile apps also happen to be scenario based or operate using game-like functions (Hao-feng et al. 2010). Given that these mobile apps, like their console game counterparts, have not yet been fully attributed to or correlated with empirically based learning gains, it is reasonable to suggest that the need exists to examine the educational validity of apps using a measuring tool that has been shown to be effective for games.

The effectiveness of the RETAIN rubric as a measurement tool in assisting with the evaluation of console games has been demonstrated in previous studies (Gunter et al. 2007; Kenny and Gunter 2011) and reaffirmed in several follow-up reviews (Prinsloo and Jordaan 2014). The current study was created with the expressed goal of providing direct evidence that RETAIN could also be utilized as a valid evaluation tool to assess the instructional efficacy of most mobile apps. Based on previous successes using RETAIN to formatively evaluate console games, it was hypothesized that the same process would also work to make the efforts more efficient in analyzing various apps intended for use in the classroom.

Mobile Apps and the Gamification of Learning

Granic, Lobel, and Engels recently stated, "Console [sic] games are a ubiquitous part of almost all children's and adolescents' lives, with 97% playing for at least one hour per day in the United States" (2014, p. 66). Based on their previous research regarding the instructional efficacy of console games (Gunter et al. 2007; Kenny and Gunter 2011), we suggest that the benefit derived from using most mobile apps in the classroom is not as much about increasing the amount of academic content as it is about motivating students (especially reluctant learners) to engage in learning activities (Granic et al. 2014). It is well-known that motivation is necessary, but motivation is an insufficient instructional strategy. The concept of game-based learning has evolved into what is often referred to as "gamification," a term coined in 2002 by Nick Palling, a British-born computer programmer and inventor (Marczewski 2013). Palling described gamified instruction as a method that extracts the best elements of game design and gameplay that are inserted into classrooms, regardless of whether a game is actually present. In short, gamifying the classroom refers to a delivery method in which those aspects of games are included as a part of the strategy that makes those experiences interactive, meaningful, and motivating.

Using this premise, as well as their suggestion that mobile apps are similar in design to game constructs, the current researchers raised the question as to whether mobile apps can offer a unique and facile means to "gamify" a classroom, given that they are appropriately designed to do so. The

similarities between gameplay and the design of most mobile apps are direct and strong because:

- Both are personal.
- Both are problem based.
- Neither requires a manual to learn how to use them.
- Both mostly involve some type of visualization.
- Both offer "freedom to fail" and learn from it.
- They are fun, interactive, and provide direct feedback.

These observations helped the researchers further postulate that most mobile apps can be considered to be "gamified" even if their play mechanics are not strictly game based. Both enable students to become immersed in instructional content that help them act in meaningful ways allowing them to "play" so as to foster the internalization of content.

Using mobile apps in the classroom has drawn the attention of educators and researchers due to their potential to make "…not only a step but a leap further into the realm of learner-centered pedagogies" (Crompton 2013, p. 11). Many educators recognize that the future of education and training, regardless of how they are delivered, should include mobile technologies to remain relevant and effective. The math involved with enumerating the growth in the number of mobile devices used informally by today's learners provides a strong indication that mobile devices could potentially create effective personalized learning experiences, but only if their design includes effective, innate learning constructs. Kayaker (2015) suggests that students learn best by doing because it appears to promote deeper learning. The use of a play strategy to encourage students to learn/remember is an essential part of any effective gamification strategy. The authors suggest that the benefits provided by the engaging use of mobile devices as informal learning agents provides classroom educators with excellent opportunities as long as the apps are also properly designed to teach something, are targeted in their content, and are based on strong instructional theories and practices. In the current study, specific mobile apps were selected for evaluation based on an analysis of their advertised educational value.

The RETAIN Rubric

In 2007, and later supported by a follow-up study in 2011, Gunter, Kenny, and Vick proposed an extensive analysis to compare standard game design theory to the classical teaching and learning models. Their analysis uncovered several gaps between the two design models. The result was their creating a structured rubric that would evaluate the constructs and designs of console games. RETAIN was originally crafted based on this careful evaluation and aggregation of what was determined as best practices in instructional design and had been

shown effective over time in comparative studies (Prinsloo and Jordaan 2014). RETAIN is an acronym derived from a careful comparative review of both the game and instructional design genres (Relevance Embedding Transfer Adaptation Immersion and Naturalization). Care was taken to consider proper inferences, and representations were drawn of the terminologies, contexts, theories, and methods identified in both.

Several of the elements that make up the acronym correlate with one another. In a series of comparisons, which were supported with several follow-up studies, the authors found that the nuances and differences in goal oversights were the main causes for many games to fail in their attempt to teach academic content. Addressing these contradictions in meaning served as the basis for the weighting that was embedded into the RETAIN model and rubric. It was felt that the weighted model would help enhance the predictive powers of the games in their use in the classroom (Gunter et al. 2007; Hao-feng et al. 2010; Campbell et al. 2015a; Campbell et al. 2015b; Gunter et al. 2016).

The conceptual framework for motivation, for example, is universal to the design of both instructional strategies and games (Keller 1983). The fundamentals behind Gagne's "Nine Events of Instruction" (Gagné 1985) developed for the military and supply an appropriate checklist for instructional designers also played a significant role in the development of RETAIN. Further, practices similar to the ADDIE model have made significant contributions to the development of game design models and are common in software development. On the other hand, Bloom's (1956) hierarchical model for knowledge acquisition and retention of information was found to differ, with transfer being one of the most obvious. In game design, this roughly correlates to the idea of leveling up. The authors found in their original studies that learning strategies in gameplay and cognitive knowledge acquisition differed. The focus in a game has traditionally been to maximize gameplay mechanics, whereas in academia the concept of knowledge building relates directly to content knowledge (Kenny and Gunter 2011). Due to this significant discrepancy, transfer element was weighted more than the others. Similarly weighted was adaption. Console games teach gameplay mechanics through the use of challenges. Few of them, however, were found to follow through using Piaget's (1969) notions on schema and dis-equilibration. Lastly, it is common knowledge that, for a game to be a commercial success, a crucial element is for it to integrate a convincing fantasy or story line that fosters immersion. Similarly in learning situations, story/narrative constructs have been shown to be an effective contextualizer (Havens 2007). Gunter et al. (2007) determined in their discussions with game designers that, even though fantasy was a key element of game design, there were only small, connotational differences. Thus, the usefulness of the embedding and immersion characteristics found in successful console game designs more closely correlated to one another

than the other elements. Because of this, reality, story, immersion, and engagement are weighted slightly less in the calibrated rubric found in RETAIN.

Based on our current review of the current state of affairs in commercial, off-the-shelf mobile apps, we believe there exist significant parallels between our previous analyses of commercial console games and mobile apps, causing us to also believe that there is a need for app developers to take advantage of RETAIN. Like with console games, the basic premise is that, to follow best instructional practices, mobile applications intended for the classroom should also follow the same evaluation for content acquisition, knowledge transfer, and automasticity/naturalization to ensure that those apps selected for classroom use actually teach what the teacher or instructor intends and at the appropriate level. Based on previous successful experiences with RETAIN in evaluating console games, the authors adapted slightly the rubric to determine its efficacy in the current situation. For these reasons, we hypothesized that RETAIN would be a useful, formative tool.

Methodology

As mentioned previously, there was a dual purpose for the current study. The first was to determine if teachers could successfully utilize the RETAIN rubric to examine mobile applications in the same way that was done previously for console games (Pellegrino and Hilton 2012). In this study, a second goal was also established. Based on the knowledge gained in several corollary studies (Campbell et al. 2015a; Campbell et al. 2015b), we wanted to measure whether a teacher's perceptions about integrating mobile apps would positively correlate based on their having a useful tool to more easily assess the value of the app. Based on previous successful administrations of RETAIN studies (Gunter et al. 2007) and corollary studies that delved into teachers' dispositions and attributions about games (Hao-feng et al. 2010), we were fairly confident that using RETAIN would fill in gaps in the literature about how to successfully implement mobile apps in the classroom, regardless if they were actual games.

The study employed a mixed method phenomenological research (MMPR) methodology (Mayoh and Onwuegbuzie 2015). The MMPR methodology provides a robust, formative explanation of teachers' quantitative responses to surveys. Each administration of the survey employed qualitative, quantitative, and phenomenological explanations that were treated with equal importance. Through triangulation, the authors felt they could better explain both the use of RETAIN as a tool to evaluate the mobile applications as well as identify teachers' attitudes towards the integration of mobile apps in general. Both quantitative and qualitative data were collected concurrently (Creswell 2015). The study aimed to both quantify and qualitatively describe/explain the use of RETAIN among

teachers and to determine the level of appropriateness of using it to analyze mobile apps regardless if they were actual games in the same way the RETAIN model was used for console games. Our approach included assessing participants' perceptions of using the RETAIN model to determine whether the rubric would help to overcome any misgivings they might have about the apps or their use in their classrooms. Data were collected using surveys (pre- and post-), completed RETAIN evaluations, observations, and participant self-reflections.

Participants

Participants included a convenience sample of an intact group of in-service teachers in a graduate degree program at a large university located in the southeastern USA. A smaller intact group of educators located in Brazil volunteered for the study to see if any cultural biases or anomalies existed that might prevent them from also adopting RETAIN. The collaboration between the Brazilian and US educators was convenient because a visiting scholar from Brazil was spending time at the university. For the most part, the Brazilian participants appeared to hold similar suspicions and misgivings about the use of games and apps. All participants indicated that they had taught less than 15 years, with the majority of them teaching 8 years or less. The participants represented a cross section of teaching levels, including those who taught kindergarten as well as those who taught high school. The participants' ages ranged from 25 to 55 years, with the majority of them between ages 25 and 40. Over three quarters of the participants were female.

Pre-study Survey

Participants took a pre-study survey that the researchers developed specifically for this study to collect demographic information as well as questions to ascertain their familiarity with the term "serious" or "console" games, their experience with mobile educational applications, and any previous knowledge of the RETAIN model. Further, the participants were asked to indicate whether they previously incorporated mobile apps in their classrooms. Participants were also asked to choose verbs that correlated to Bloom's taxonomies as an indicator of their knowledge about the use of games for learning. Finally, participants rated their specific familiarity with three of the mobile game applications (those that the researchers had previously identified as educational apps) that they were going to be asked to evaluate. Upon completion of the survey, the students were taught the RETAIN model/rubric and its use.

Implementation

The RETAIN rubric was modified slightly to target mobile apps rather than console games. Similarly to the previous

studies, each element was weighted according to its perceived importance (Kenny and Gunter 2011). As with previous administrations, Transfer and Naturalization were assigned the greatest weights. Again, the highest rating an app could receive was 63 points. The higher the score, the stronger the app supported or supplemented knowledge transfer and naturalization. Participants recorded and calculated the appropriate points for each app they evaluated. Participants followed up with an open-ended, written reflection that included their rationale in support of their score (see Attachment A).

After the pre-study survey, participants were introduced in detail to the RETAIN model and were taught the nuances of using it to evaluate mobile apps. Next, participants were asked to use the rubric to evaluate a total of five mobile applications. Three of the apps were assigned by the researchers. These were selected because they had been marketed as educational by their developers or the researchers had found anecdotally in informal reviews that they appeared to be useful as classroom tools. Two additional apps were chosen by each participant, who knew of them either through firsthand knowledge or were ones that their districts had recommended. While it was not specifically identified beforehand, one of the pre-selected apps was a mobilized version of the classic game, *Oregon Trail*. The other two were scenario-based but not necessarily pre-categorized as games. All apps worked on both Android- and iOS-based platforms. Students rated all five apps over a 3–4-week period. At the end of 4 weeks, participants turned in their evaluations and shared their completed results with other participants through a virtual/online discussion board. Finally, all participants completed a post-administration survey. The study took 5 weeks to complete.

The post-study survey, like the pre-study survey, was administered online through Qualtrics. Participants indicated their perceptions as to the usefulness of using RETAIN to evaluate the mobile applications, their current use of these apps in their classrooms, the time it took to evaluate them, and the helpfulness of their being able to read other participants' reviews, plus open-ended reflections as to whether having access to the rubric changed their views on using apps in the classroom in general. The post-study survey also included questions related to the likelihood of their using the rubric in the future and whether they would encourage others to do so also.

Discussion

Based on the results of the pre-survey, no participant indicated that they were familiar with the RETAIN rubric. Many of them used mobile applications in their classrooms. None of them had previously vetted mobile applications using a rubric similar to RETAIN. Of those who indicated that they had

previously used mobile applications in their classroom, they indicated that they were motivated to do so because (a) of recommendations they received from other teachers, (b) the applications were free or had been purchased by the district and were readily available, and/or (c) because they saw the app in use at another class, school, or conference (Table 1).

All participants universally agreed that apps should be evaluated and analyzed for instructional effectiveness prior to their use in the classroom. After the pre- and post-surveys were completed, a convenience group was selected to participate in a focus group and was asked to discuss if they agreed that the apps used in their classrooms needed to be systematically evaluated. Responses included, among others, (1) that all digital programs should be reviewed just like textbooks and other media that are similarly reviewed for accuracy, (2) to prevent assigning an app that did not meet the intended learning objective, and (3) to make sure that the app was age and content appropriate.

Participants acknowledged that they did not know whether any of the mobile apps that they were currently using in their classrooms aligned with the educational practices identified by the RETAIN rubric. Several participants in the focus group indicated that, in the past, it was unclear as to which elements or characteristics should be included in a rubric to assess the efficacy of the apps. Finally, using Bloom's Taxonomy (1956) as a basis, the focus group participants were asked how/why they would use the apps in their classroom (see Table 2). They were not provided Bloom's taxonomy in an intact form; instead, they were provided representative verbs. Participants were allowed to choose more than one verb for each app. The majority indicated they would use the apps to apply knowledge and to support deeper understanding, analysis, and synthesis. Some commented that they tended to assign math apps for skill and practice rather than using the app to supply a context to the problem-solving process.

Post-study Survey Qualitative Responses

Research Question 1: What are teachers' impressions of using the RETAIN model to evaluate mobile and web-based game applications?

All participants indicated that either they were somewhat likely or extremely likely to use the RETAIN rubric in the future. Further, more than half of the participants indicated that, as a result of evaluating the apps in this manner, they would continue to use one or more of those that rated the highest in their classroom. On the contrary, 47% of the participants stated that they would not use any of the reviewed applications if they scored low. (See Fig. 1.)

Scanning the responses for commonly used words, several themes evolved. These outlined a gamut of impressions about participants' use of the RETAIN rubric to gain support for the benefits of its usage to evaluate applications. Some participants mentioned that, after reviewing the results of an evaluation of some of the apps that they chose to review, they decided that they would no longer use those that earned a low score, specifically because they earned a low score on knowledge transfer. This appears to confirm the validity of the researchers placing a higher valuation on this element in the rubric. Others noted that, while some apps received low scores and caused them not to use a particular app in the future, they were encouraged to look for others that would earn a higher RETAIN score. After one participant found that one district-approved mobile app had earned a low score, she stated "I am surprised that my school district does not preevaluate [sic] the apps that that they recommend."

One comment about the benefits of using RETAIN was echoed by many. "I think the model is a great way to effectively decide what the purpose of the game is and if students will really gain what you want them to get out of the games. I am now going to review other games I use in my classroom with it [RETAIN] [sic]." After reviewing RETAIN evaluations submitted by their peers, one participant stated, "Some of the other students' evaluations gave me great information. I looked up some of the games to use with my students." Seventy-three percent of these participants confirmed that they were going to consider other apps based on the other's reviews. The benefits of reviewing an app went beyond determining a score from the rubric. Some participants discovered that the content was not being integrated properly into the mobile app but that some of them were not age appropriate. During the focus group, participants discussed potential plans of action for apps that were flawed or contained inaccurate content and how the shortcomings could become teachable

Questions	Yes	No
Do you use mobile applications in your classroom?	100%	–
Have you formally evaluated the applications used in your classroom?	–	100%
Do you choose mobile applications based on the recommendation or indicated use of others?	80%	20%
Is there a need for a formal evaluation of games and mobile learning apps for classroom usage?	100%	–

Table 1 Percentage of participants answers pre-study survey

Table 2 Bloom's taxonomy reference chart (old and new)

Comprehend Understanding (2)	Synthesis Creating (6)	Application Applying (5)	Evaluate Evaluating (4)	Knowledge Remembering (1)	Analysis Analyzing (4)
Had my students use the game to classify, compare, estimate, show.	Had my students use the game to analyze, arrange, examine, discover.	Had my students use the game to experiment, practice, select, solve.	Had my students use the game to estimate, propose, simplify, speculate.	Had my students use the game to identify, label, recognize, repeat.	Had my students use the game to deduct, defend, persuade, give reason.

Participants only saw the second row of this chart. Numbers indicate the level of cognition from the lowest to the highest

moments. Including reteaching the content, some would have their students determine what content was inaccurate in the game. Others would open up discussions about how the information could be corrected.

Research Question 2: Can RETAIN be applied to mobile applications to determine their support for content transfer?

Seeing as Transfer and Naturalization were the highest rated elements in the rubric, a separate discussion took place in the focus groups to validate the weighted values for these two categories. It is important to note that the main focus of RETAIN was originally created to evaluate the efficacy of console games as stand-alone tools. The rubric was not originally intended to be used on games that played a supplementary role for activities that were also guided and supported by teacher intervention. Participants indicated unanimously that the model was appropriate for evaluating mobile apps regardless of whether they included gameplay, as long as they were scenario based. Participants indicated that the ability to score for knowledge transfer and naturalization provided very important information and weighting these two elements higher than the others was valid (Fig. 2).

One participant stated, "I think most apps are always going to score low on RETAIN because they are missing some of the simulation and gameplay qualities that are found in console games." All participants concurred that they were concerned that some apps that they expected to score high did not because they did not scaffold content or increase in difficulty level.

Another concern arose because a number of what many previously thought were useful apps scored low in many of the content areas. One teacher stated, "I found it hard to find apps that relate to my content. I teach middle school language arts and there are very few that are appropriate for that content and level." Another said, "Most apps designed for my classroom are based on 'drill and kill' arcade games. I would love to read about some apps that are directly applicable to what I do." Another indicated, "I almost could not find a serious game or game-like application. Almost all of them are arcade style or are repetitive games that really don't require much thinking." Lastly, one participant actually went on to ask rhetorically: "Do game designers consult with teachers about the academic content to include within a game?" In short, almost all participants expressed satisfaction that RETAIN allowed them to consider the academic value of mobile apps that they would be selecting in the future.
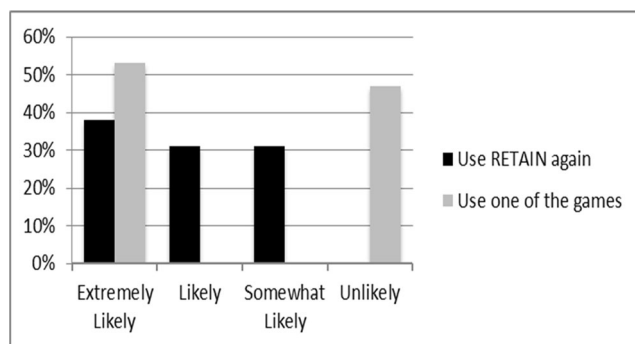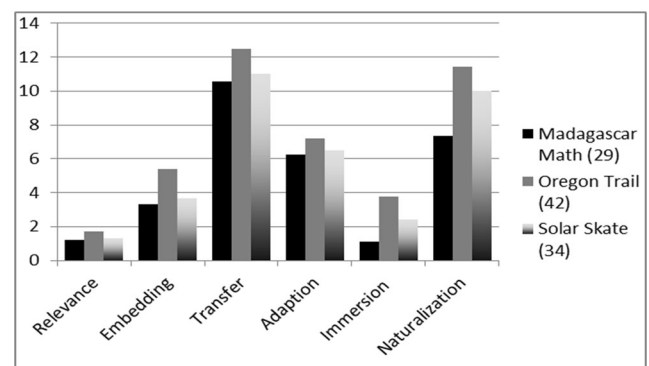

Fig. 1 Likelihood of use

Fig. 2 Assigned mobile applications weighted scores by category. (*Numbers in parentheses* indicate average score)

Conclusion

The use of mobile devices in K-12 educational settings continues to grow, especially with the implementation of one device to one person (1:1) access and Bring Your Own Device (BYOD) initiatives (Madden et al. 2013; Jonas-Dwyer et al. 2012). As a result of these programs and the growing numbers of school districts committing to digital curriculums, educators are incorporating more games and apps into their classrooms. The rush to find apps that are age and content appropriate increases the need to have a rubric that supports them in ensuring that the apps directly correlate to learning goals, standards, and intended outcomes (Powell 2014). The present study provided information that the use of the RETAIN rubric demonstrated at least an indirect correlation on how an evaluation influences attributions of effectiveness on the part of teachers towards using games and apps in general.

RETAIN has previously been shown in several studies to be an effective evaluation tool for stand-alone console games (Kenny and Gunter 2011; Prinsloo and Jordaan 2014; Zhang et al. 2010). In this study, the RETAIN rubric was adapted slightly and then reaffirmed as an effective tool to evaluate/assess mobile applications, regardless if they are game based as long as they were at least scenario based. Mobile applications that earned the lowest scores often did so because they lacked the ability to promote transfer and naturalization, even if they scored high on immersion and engagement. While these latter two elements are significantly important to a commercially developed game or app, they are related to being necessary but insufficient in educational settings. Scoring immersion and engagement for apps only comes into play if their user interface is not friendly and attractive. Based on the evaluation of the qualitative, open-ended comments made by participants in the study, if the shortcomings in the apps could be found during the design stage of development, much of the time wasted making bad selections and negative attributions towards apps and games could be minimized. Teachers and instructional designers working with game and app builders during design and development work is referred to as educational design research. This method is more of a formative role, rather than them acting as summative reviewers once the games have been developed (McKenney and Reeves 2012), resulting in an increased overall educational efficacy of the apps at a far less cost (Wood 2013).

Even though some mobile apps scored low on the RETAIN scale, they may still have a place in the classroom. It is commonly argued that there is hardly ever one single digital tool that has been developed that can fit all learning situations. Most scenario-based mobile applications, when combined with appropriate supportive instructional strategies and other technologies, also have the potential to gamify a classroom. The results of this study seem to indicate that RETAIN can be a powerful predictive tool in helping to identify weaker apps and indirectly suggest the kinds of supplemental strategies that could be utilized to help appropriately gamify the classroom experience.

The pre-study survey indicates that participants typically used the recommendation of others to help them choose the learning apps that they might use in their classroom, often citing that they (a) lacked the time to review them, (b) did not know what to look for in the app, or (c) were required by school authorities to utilize those provided for them or provide arduous justifications for why those chose others that were not pre-approved. In the post-study survey, participants were asked how long it took them to evaluate an app. Participants took between 45 min to over 3 h. The Brazilian participants provided more comprehensive reviews of the apps, which may explain why it took them longer to perform the reviews.

One unintended outcome was that most participants voluntarily disseminated the results of their RETAIN evaluations to their colleagues, administration, and interested co-workers. One reviewed her list of the apps provided to her students' parents and revised it based upon how they scored on the RETAIN rubric. In one case, this individual decided to refer her students' parents to Web-based games rather than mobile apps that covered equivalent content because she discovered that the Web-based games tended to score higher on RETAIN, indicating that continued research would be needed to determine the causes of this unexpected outcome.

In the end, many of the mobile apps evaluated in this study did not score well. While most participants did find an abundance of apps that were labeled "educational," they seemed to have difficulty finding those that even advertised that they would engage their learners in support of their acquiring content knowledge.

References

Atkins, D.E, Brown, J.S., & Hammonds, A.L. (2007). A review of the Open Educational Resources (OER) movement: achievements, challenges, and new opportunities. Creative Commons Attribution. Retrieved: March 1, 2017 from: https://pdfs.semanticscholar.org/8d16/858268c5c15496aac6c880f9f50afd9640b2.pdf.

Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook I: the cognitive domain*. New York: David McKay Co, Inc..

Campbell, L.O., Gunter, G., & Braga, J. (2015a). Utilizing the RETAIN model to evaluate mobile learning applications. In proceedings of *society for information technology & teacher education international conference 2015* (pp. 670–674). Chesapeake: Association for the Advancement of Computing in Education (AACE).

Campbell, L.O., Gunter, G., & Kenny, R. (2015b). The gamification of mobile learning evaluated by the RETAIN model. A*ssociation for educational communication and technology*, 2015. Accelerate Learning: Racing into the future, conference, Indianapolis, Indiana, November 3–7, 2015.

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Los Angeles: Sage.

Crompton, H. (2013). A historical overview of mobile learning: towards learner-center education. In Berge & Muilenburg (Eds.), *Handbook of mobile learning*. New York: Routledge.

Dominguez, A., Saenz-De-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pages, C., Martinez-Herráiz, J.J. (2013). Gamifying learning experiences: practical implications and outcomes. *Computers &Education, 63*, 380-392.

Gagné, R. (1985). *The conditions of learning* (4th ed.). New York: Holt, Rinehart & Winston.

Granic, I., Lobel, A., & Engels, R. (2014). The benefits of playing video games. *American Psychologist, 69*(1), 66-78.

Gunter, G., Kenny, R., & Vick, E. (2007). Taking educational games seriously: using the RETAIN model to design endogenous fantasy into standalone educational games. *Educational Technology Research & Development, 56*(5/6), 511–537.

Gunter, G. Campbell, L.O., Braga, J., Racilan, M., Souza, V. (2016). Using the RETAIN Model to evaluate mobile educational games for language learning. *Revista Brasileira de Linguistica Aplicada.*

Hao-feng, Z., Xi-yan, F., & Hai-feng, X. (2010). Research on the design and evaluation of educational games based on the RETAIN model. *2010 3rd International symposium on Knowledge Acquisition & Modeling (KAM)*, 375. doi:10.1109/KAM.2010.5646186.

Havens, K. (2007). *Story proof: the science behind the startling power of story*. Greenwich: Libraries Unlimited.

Jonas-Dwyer, D., Clark, C., Celenza, A., & Siddiqui, Z. S. (2012). Evaluating apps for learning and teaching. *International Journal of Emerging Technologies in Learning, 7*(4), 54–56.

Kayaker, J. (2015). Deeper learning in practice. Edutopia. Retrieved: March 1, 2016, from http://www.edutopia.org/blog/deeper-learning-in-practice-jennifer-kabaker, http://www.online-journals.org/index.php/i-jet/issue/view/135.

Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), *Instructional design theories and models: an overview of their current status* (pp. 383–434). New York: Lawrence Erlbaum.

Kenny, R., & Gunter, G. (2011). Factors affecting adoption of video games in the classroom. *Journal of Interactive Learning Research, 22*(2), 259–276 **Chesapeake, VA: AACE**.

Kenny, R., & McDaniel, R. (2011). The role teachers' expectations and value assessments play in their adopting and integrating video games into the curriculum. *British Journal of Educational Technology, 42*(2), 197–213.

Madden A., Lenhart, M., Duggan, M, Cortesi, S., Gasser, U. (2013). Teens and technology 2013: main findings. Pew Research. http://www.pewinternet.org/2013/03/13/main-findings-5.

Marczewski, A. (2013). *Gamification: a simple introduction* (p. 46). Amazon Digital Services, Inc.

Mayoh, J., & Onwuegbuzie, A. (2015). Toward a conceptualization of mixed methods phenomenological research. *Journal of Mixed Method Research, 9*(1), 91–107.

McKenney, S., & Reeves, T. C. (2012). *Conducting educational design research*. New York: Routledge.

Pellegrino, J. W., & Hilton, M. L. (Eds.). (2012). *Developing transferable knowledge and skills in the 21st century*. National Academies Press http://www.nap.edu/catalog/13398/education-for-life-and-work-developing-transferable-knowledge-and-skills.

Piaget, J. (1969). *The mechanisms of perception*. London: Rutledge & Kegan Paul.

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

Prinsloo, J. W., & Jordaan, D. B. (2014). Selecting serious games for the computer science class. *Mediterranean Journal of Social Sciences, 5*(21). doi:10.5901/mjss.2014.v5n21p39.

Wood, D. C. (2013). *Principles of quality costs: financial measures for strategic implementation of quality management* (4th ed.). Milwaukee: Quality Press.

Zhang, H.F., Fan, X.Y., & Xing, H.F. (2010). Research on the design and evaluation of educational games based on the RETAIN Model. Third international symposium on knowledge acquisition and modeling. p. 375–378.