Computers & Education 75 (2014) 65–71

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Mobile app selection for 5th through 12th grade science: The development of the MASS rubric

Lucy Santos Green[a,*], Richard P. Hechter[b], P. Dawn Tysinger[a], Karen D. Chassereau[a]

[a] Georgia Southern University, P.O. Box 8131, Statesboro, GA 30460, United States
[b] University of Manitoba, 234 Education Building, Winnipeg, Manitoba R3T 2N2 Canada

ARTICLE INFO

Article history:
Received 9 October 2013
Received in revised form
10 February 2014
Accepted 11 February 2014

Keywords:
Mobile apps
Tablets
STEM
K-12
m-Learning

ABSTRACT

Tablets devices for student use present several advantages over laptops and desktops including portability, touch-screen features and numerous applications. However, the magnitude of apps available also presents a challenge for secondary science educators who struggle to select content-appropriate applications that support the development of science literacy and science content acquisition. This paper details the process of creating, developing and testing a mobile science application rubric so as to aid secondary science classroom teachers in selecting and rating science applications for a K-12 student target population and its curricular needs. Quantitative and qualitative data collected during four design cycles resulted in the Mobile App Selection for Science (MASS) Rubric, comprising six items on a four-point response scale. Further comparison of the science content-specific MASS rubric with a general mobile app selection rubric (Evaluation Rubric for Mobile Applications; ERMA) revealed expected results with three item pairs (Pair A, Pair C, and Pair D) demonstrating concurrent validity through significant correlations and one pair (Pair B) displaying the expected divergent validity. Additionally, paired *t*-tests among each pair indicated a significant difference in participants' ratings of the apps using the two rubrics. The differences in ratings were also in the expected direction given the content-specific nature of MASS versus the more general nature of ERMA.

1. Introduction

Helping students to develop a lifelong love of learning and exploration is considered a vital component of modern education. In support of this goal, ubiquitous mobile technology holds possibilities for the design and implementation of learning experiences that take the student beyond the traditional classroom space (Ahmed & Parsons, 2013). Ubiquitous technology considers learners who are constantly on the move. Thus, ubiquity "refers not to the idea of 'anytime, anywhere' but to the 'wide-spread', 'just in time' and 'when-needed' computing power for learners" (Peng, Su, Chou, & Tsai, 2009, p. 175). This characteristic of ubiquity is perceived, by educators and parents, to differentiate learning with mobile devices from other types of technology integrated learning activities (Sha, Looi, Chen, & Zhang, 2012). Recent educational trends such as Bring Your Own Device (BYOD) or Bring Your Own Technology (BYOT) reflect this public demand for ubiquitous technology, and have led to a corresponding popularity in mobile and handheld devices in K-12 educational settings (Banister, 2010).

The iPad, in particular, presents several advantages for student use over laptops and desktops. In addition to sharing many of the same capabilities as laptops and desktops, it is relatively affordable, lightweight, portable, equipped with wireless network capacity and allows for the quick installation of a vast array of topic-based applications (Brand & Kinash, 2010). Many of these applications incorporate multi-sensory input features which enable students to experience content in auditory, visual and tactile forms, a promising feature for the support of student learning of science concepts (Carr, 2012; Castelluccio, 2010; Hill, 2011). There is considerable precedence for the use of handheld computers and devices in K-12 classrooms to support scientific inquiry (Roschelle, Penuel, Yarnall, Shechtman, & Tater, 2005).

---

* Corresponding author. Tel.: +1 912 478 0275.
E-mail addresses: lgreen@georgiasouthern.edu (L.S. Green), richard.hechter@umanitoba.ca (R.P. Hechter), dtysinger@georgiasouthern.edu (P.D. Tysinger), kchassereau@georgiasouthern.edu (K.D. Chassereau).

http://dx.doi.org/10.1016/j.compedu.2014.02.007
0360-1315/

Research details the use of mobile devices in science to improve science learning through collaborative projects, participatory simulations, observation and concept mapping (Chen, Kao, & Sheu, 2003; Kaput & Hegedus, 2002; Stroup, 2002; Tinker & Krajcik, 2001; Yarnall et al., 2003).

However, there is very little known about the *process* of learning with mobile devices, what impact the integration of these technologies might have on student learning, and what aspects of the applications themselves nurture lifelong inquiry in students (Sha et al., 2012). Studies that have been conducted oftentimes report conflicting results. While Murray and Olcese (2011) struggled to find iPad apps that supported “truly innovative teaching and learning [that]…[extended] what teachers and students could otherwise do” (p. 46); the use of tablets as a ‘cognitive companion’ was found to positively impact student learning in a Hong Kong primary school (Li & Pow, 2011). In the field of K-12 science, several research studies concluded that students engaged in science learning activities supported with mobile applications demonstrated enhanced scientific content acquisition as compared to students engaged in more traditional methods of science learning (Huang, Lin, & Cheng, 2010; Hwang & Chang, 2011). In contrast, Park, Parsons, and Ryu (2010) found no significant difference in student knowledge regardless of the use of mobile apps.

Sung and Mayer (2013) suggest that the reason for these conflicting results may lie within the focus of this body of research, much of which is based on a “seemingly reasonable assertion…that learning on an iPad in a comfortable place is more fun and therefore students will try harder to learn than when they learn in a lab [or on a computer]” (p. 639). They propose that iPad applications ought to be evaluated through the lens of Clark’s (2005) method-not-media hypothesis which states that “the most promising approach to learning is [the assumption] that it is caused by instructional methods that can be embedded in instruction and presented by a variety of media” (Clark, 2005, p. 99). In other words, while app characteristics such as usability, interface and absence of advertisements are considered, the pedagogical applications of the app ought to carry the most weight in the selection process (Ting, 2012; Walker, 2011). Further, a shift in evaluative focus from generalized app characteristics to science pedagogy more accurately reflects the existing close alignment between teacher technology integration practices and their pedagogical beliefs (Ertmer, Ottenbreit-Leftwich, Sadik, Sendurur, & Sendurur, 2012).

Currently, there are websites dedicated to curating and reviewing educational mobile apps. A quick Google search of the phrase “educational apps for iPad” yields over forty million page results, most of these a quick list of selected favorites. Other web resources curate, tag and encourage review postings of educational apps (e.g.TechChef4U.com, IEar.org and Learninginhand.com). Even while these websites “provide overviews and useful insights into the pros and cons of specific apps, [they] do not provide a common language or structure for evaluation” (Walker, 2011, p. 60). The same mobile app might be highly rated by one educator, and completely derided by another. In an attempt to standardize app evaluation, and develop a common language, several mobile app evaluation rubrics have been developed. The two most disseminated rubrics are Schrock’s (2013) rubrics for content-based and creation-based apps, and Walker’s (2011) Evaluation Rubric for Mobile Applications. Schrock’s (2013) rubrics function as a 10 point checklist of app characteristics and observable student behaviors such as student motivation to use the app and student perception of the helpfulness of app instructions. Walker’s (2011) ERMA rubric, comprised of 6 categories and 4 levels, also evaluates app characteristics and observable student behaviors such as student motivation and engagement with the app, as well as the degree to which teacher supervision is needed for app usage.

Despite all of these available resources, a framework for the evaluation of science education applications, using a common language structure associated with lab-based technologies and scientific tools is still absent in the literature. Therefore, our design team endeavored to develop an evaluative tool that would allow science teachers to critically explore and evaluate the nature of individual digital applications towards their pedagogical suitability for curricular integration and support.

2. The pedagogical perspective

A pedagogical perspective in mobile app selection translates into the use of a mobile app for more than its technological functions, stressing the context of its integration instead. Through this lens, the classroom teacher first considers how a mobile app supports his or her pedagogical approach (Ting, 2012). When designing an evaluative rubric to support this pedagogical perspective, it is important to consider the entirety of the mobile learning environment, including learner experiences and potential interactions (Traxler, 2007). Consequently, the mobile learning pedagogical framework pictured in Fig. 1 was used to inform the design process of MASS (Kearney, Schuck, Burden, & Aubusson, 2012). The m-learning environment framework was developed to support “an examination of m-learning which foregrounds pedagogy rather than technology, a perspective in which pedagogy is central and technology is under investigation only for what may be distinctive about the learning afforded by that technology” (Kearney et al., 2012, p. 2).

Based on the principles of socio-cultural theory (Vygotsky, 1978) which describes learning as a situated, social endeavor, it considers three specific characteristics of m-learning: authenticity, personalization and collaboration (see Fig. 1). Authenticity refers to tasks that most closely approximate tools and procedures used in the ‘real world’ (Radinsky, Bouillion, Lento, & Gomez, 2001). The framework further breaks down authenticity into *situatedness*, authentic tasks that place students in actual practice (e.g. recycling programs) and *contextualization*, authentic tasks that simulate real-world settings within a learning space (Kearney et al., 2012). Personalization refers to mobile application characteristics such as learner choice, self-pacing, self-regulation, and customization (McLoughlin & Lee, 2008). Within this framework, personalization is further extended to the m-learning environment when activities “create a personalized, tailored learning journey” (Kearney et al., 2012, p. 9). The third characteristic, collaboration, heavily emphasizes social interaction as a component of learning where the learning experience is modified by the mobile applications used. This construct considers not only collaborative conversations present in the m-learning environment, but the ability to share representations of learning with others, incorporating the responses elicited by those representations (Ackermann, 2009).

3. The design process

The m-learning environment framework guided and informed the creation, development and testing of a mobile science application rubric to aid secondary science classroom teachers in selecting and rating science applications for a K-12 student target population, curricular needs and pedagogical appropriateness. The Mobile App Selection for Science (MASS) rubric, a holistic instrument that measured different levels of mobile app performance or quality, was developed over the course of four design cycles which will be described in

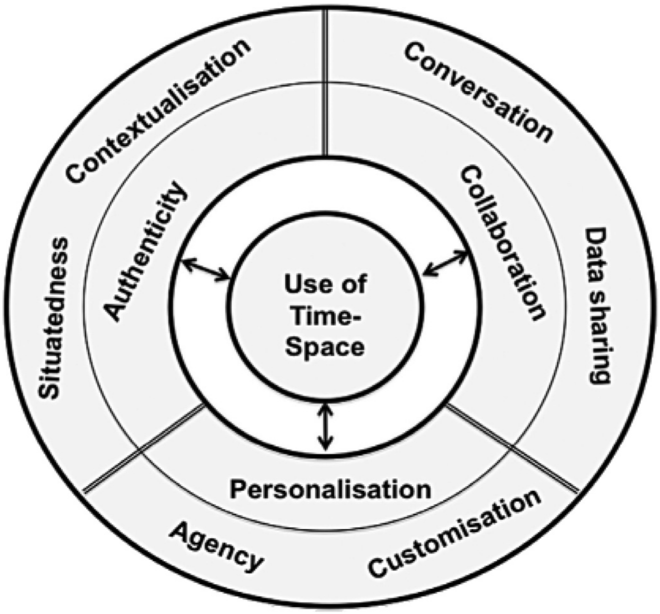

Fig. 1. Kearney et al. (2012) m-learning framework.

subsequent sections (see Appendix A). For the development and testing of the MASS rubric, procedures sought to (1) adequately identify patterns in participant use of the MASS rubric, (2) redesign MASS based on reviewer and participant feedback and; (3) test its strength as a content-specific mobile app evaluation tool as compared to a general mobile app rubric through the determination of concurrent and discriminant validity (Table 1).

Throughout the development of MASS, a mixed methods investigative approach was selected in order to more fully inform its design (de Waal, 2001). Mixed methods "combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study" (Johnson & Onwuegbuzie, 2004, p. 17). The benefit of this investigative approach is two-fold. First, collection of both qualitative and quantitative data can highlight complex and interdisciplinary phenomena present in educational settings so that results further inform each type of data set. Second, and most critical to the design process described in this article, a mixed methods approach aids in the description and development of procedures and tools that more closely resemble those used by educational practitioners (Onwuegbuzie & Leech, 2004). Finally, the mixed-method data collection enabled context-dependent inquiry and inductive data analysis which informed and established procedures for each subsequent design cycle (Guba & Lincoln, 1988). Therefore, in order to preserve contextual accuracy and transparency of the design process, analysis of data collected during each cycle will be discussed following its description.

3.1. Design cycle I

In design cycle I, the design team reflected on the selected m-learning framework and generated a comprehensive list of rubric subscales, informed by curriculum standards for science, literature on existing mobile app rubrics, instrument development, instructional design and mobile application design. The National Science Education Standards suggest that science education rubric development begin with a clear understanding of the expectations for science literacy (NRC, 1996). Therefore, this initial rubric subscales list was reviewed and adjusted based on individual subscale relationship to the Conceptual Framework for the New K-12 Science Education Standards (NRC, 2012) and the Manitoban Science Senior Years Curriculum Framework (2003). Both American and Canadian science frameworks were used in an effort to expand the applicability of MASS as much as possible. The first version of MASS comprised seven subscales: 1) accuracy, 2) relevance of content, 3) sharing findings, 4) feedback, 5) critical thinking, 6) navigation and 7) cognitive overload. Definitions and descriptions for each criterion were then clarified and weighted using a 4-point response scale: "not-applicable", "unacceptable", "acceptable" and "target."

Table 1
Design cycles, data collection and timeline.

	Activities/data collected	Participants	Timeline
Cycle one	Literature review Curricular standards review Creation of sub-scales and weight	Design team	Jan 2013
Cycle two	Rubric evaluation forms Questionnaire	4 External evaluators Secondary science teachers Texas, U.S. & Manitoba, C.A.	Feb 2013
Cycle three	Rubric evaluation forms Questionnaire Observations	K-12 secondary science teachers Practitioner evaluators	Feb 2013
Cycle four	Comparison of rubric scores Paired samples *t*-test Observations	K-12 secondary science teachers Practitioner evaluators	April 2013

3.2. Design cycle II

During design cycle II, four external evaluators were asked to review MASS Version 1 for clarity, completeness and usability. The four evaluators (two male and two female) were selected on the merits of science teaching experience in grades 5–12, school leadership recommendations, technological proficiency (specifically, comfort with mobile applications), and geographical location (Texas, Georgia and Manitoba, Canada). Evaluators were instructed to provide written feedback on overall rubric clarity, usability and practicality, items that needed to be adjusted or removed, as well as the naming of any criteria they felt might be missing.

The external evaluators suggested improvements in the areas of weight (e.g. removal of terms such as 'relative' or 'obstacle') and science-specific language. All four also suggested that the initial version of the rubric was a bit too large, that descriptors could be summarized and that the last two criteria: 6) navigation and 7) cognitive overload could be combined. Two of the four evaluators challenged the need for a "not-applicable" scale. Finally, two others suggested the need for a criterion measure of student engagement. Modifications based on external evaluator feedback resulted in six subscales: 1) accuracy, 2) relevance of content, 3) sharing findings, 4) feedback, 5) scientific inquiry and practices and 6) usability; and a 3-point response scale: "unacceptable", "acceptable" and "target." Although student engagement is a popular argument for the integration of technology, the designers envisioned the purpose of MASS for initial selection of mobile applications, during lesson-planning. Since the selection of these applications would occur before actual student use, the addition of a student engagement subscale would not serve the eventual purpose of this instrument.

3.3. Design cycle III

The purpose of design cycle III was to establish a clearer and more comprehensive picture of practical usability of the rubric through the eyes of K-12 practitioners. K-12 teacher evaluators for this design phase were selected from a pool of participants involved in the ISOTOPE (Infusing Science Outcomes, Technology Orientations, and Pedagogical Experiences) project. The ISOTOPE Project was a Canadian professional development program designed to promote effective and efficient use of modern technology in 21st century science classrooms. Participants in this year-long program were introduced to re-conceptualized pedagogical approaches and strategies, and pragmatic applications of modern technologies for the modern science classroom. All ISOTOPE participants were selected from northern and rural areas of the Canadian province of Manitoba. Included in this group were 6 middle grades (5–8) teachers, and 18 senior grades (9–12) teachers. Seven were female, and 17 were male. Teaching experience ranged from 1 to 32 years. All participants reported using iPad apps; however only 21 (87%) used them in their science classroom for pedagogical purposes. As a result of their reported pedagogical iPad use in science classrooms, ISOTOPE participants were considered an ideal group for practitioner evaluation of the MASS rubric.

The data collection process began by asking evaluators ($n = 24$) to record the key features of an iPad application that would contribute to their decision to use or discard a particular app for use in their own science classroom. These comments were annotated by the research team for use in re-design. Afterward, evaluators were given a copy of the MASS rubric, and asked to record comments about MASS in the margins of the rubric itself. Findings from this iteration of the MASS rubric identified that changes to the rubric were necessary to improve its usefulness for teachers. Based on the qualitative feedback collected during this phase, the MASS rubric was modified in two specific ways. First, a "not-applicable" (N/A) column was added. Second, the "scientific inquiry and practices" sub-scale was revised to clarify weighting and description. K-12 evaluators pointed out that the descriptors of scientific practices were too broad. To address this concern, the designers adjusted the sub-scale to more closely reflect language used in curricular standards and frameworks.

3.4. Design cycle IV

Design cycle IV involved the same K-12 teacher evaluators ($n = 24$) from design cycle III in a subsequent session of the ISOTOPE Project. During this cycle, the evaluators critically studied the final MASS rubric that had been revised based on their feedback from design cycle III. Evaluators were encouraged to look through a group of science apps provided for them on individual iPads. The apps were selected to represent a variety of science disciplines and were chosen from the list of most popular science apps as recorded by iTunes sales. Evaluators were instructed to select the one app they would most likely use in their own classroom given their particular science discipline. In terms of app selection, no discipline or app stood out as preferable over the others: CELL STAIN ($n = 6$), VIRTUAL CELL ($n = 6$), ATOMS HD LITE ($n = 5$), END PTE ($n = 4$) and WAVE LAB ($n = 3$). After selecting an app, evaluators then scored their selected app twice: once using the MASS rubric as a scoring tool, and once using the Evaluation Rubric for Mobile Applications (Walker, 2011) as a scoring tool. In order to determine concurrent and discriminant validity, the design team selected four items from the MASS rubric and then paired the scores from these items with scores on four similar items from the ERMA. Two of these pairings featured items in which the MASS rubric was content specific; the other two comparisons involved general mobile attributes that were similarly addressed in both MASS and ERMA. After collecting the scores from both rubrics, a paired-samples *t*-test was conducted to compare the selected pairings. Table 2 summarizes the descriptive and inferential statistics for each pairing.

3.4.1. Pair A – relevance of content vs. curriculum connection

The purpose of this item on the MASS was to measure the degree of alignment between the perceived target audience science learning objectives and the mobile application. The curriculum connection subscale in ERMA determined the degree of alignment between the target skill or concept and the skill or concept addressed in the app. Results indicate a significant correlation between this pairing thus demonstrating concurrent validity as $r = .42$, $p = .04$. Results indicate that evaluators rated the apps significantly higher [$t(1, 23) = 8.41$, $p < .01$] on the MASS score for Relevance of Content ($M = 91.67$, SD = 14.74) than the comparable ERMA score for Curriculum Connection ($M = 64.58$, SD = 14.59). The higher MASS scores indicate the potential for pedagogical flexibility, which was the primary design goal for this instrument. Evaluators were able to score the intention of the mobile apps relative to their specific discipline on MASS, referring directly to the alignment between app focus and curricular standards. Hohlfeld, Ritzhaupt, and Barron (2010) also concluded that pedagogical flexibility allowed teachers to adjust use of an educational technology tool according to curricular and student learning needs, leading to more creative and specialized approaches to technology integration. In contrast, teacher evaluators were unsure what the term "skill" on ERMA meant, as it related to curriculum and

Table 2
Descriptive and inferential results by pairing.

Pairings	*M*	SD	*r*	*t*
Pair A				
Relevance of content	91.67	14.74	.42*	8.41**
Curriculum connection	64.58	14.59		
Pair B				
Scientific inquiry and practice	43.06	37.40	.11	−2.15*
Authenticity	60.42	17.93		
Pair C				
Feedback (MASS)	23.61	34.72	.90**	−4.34**
Feedback (ERMA)	37.50	28.55		
Pair D				
Navigation	84.72	21.93	.72**	2.489*
User friendliness	76.05	23.80		

$^{*}p < .05$. $^{**}p < .01$.

authenticity. Several commented that the apps selected were not necessarily about skill or skill development, but rather for conceptual development, which they regarded as separate items and which may have constricted their perception of app flexibility.

3.4.2. Pair B – scientific inquiry and practice vs. authenticity

The intent of the Scientific Inquiry and Practice item on the MASS is to rate opportunities for scientific inquiry and practices through data collection, observation, experiences, reflection, reasoning, and communication. The MASS item was paired with the Authenticity item of ERMA that is intended to measure the relationship between targeted skills and learning activities. Because the MASS item is specific in nature and the ERMA item is general in nature, this pairing was chosen to test discriminant validity. Statistical analyses demonstrate that these items do, in fact, diverge ($r = .11$, $p = .60$) with participants rating the app significantly higher [$t(1, 23) = -2.15$, $p = .04$] on the ERMA item for Authenticity ($M = 60.42$, SD $= 17.93$) than the MASS item for Scientific Inquiry and Practice ($M = 43.06$, SD $= 37.40$). ERMA's Authenticity descriptor asks evaluators to identify the presence or absence of a real world or problem-based learning environment. However, the description for Scientific Inquiry and Practice lists the actions students should be able to conduct within this environment: observation, experience, reflection, reasoning and communication. As a result, the specificity of MASS's language in this subscale may have encouraged more rigorous rating on the part of the teacher-evaluators. The results of this pairing are substantiated by considerable precedence for the inclusion of clearly stated science standards in instruments used to evaluate materials, activities and student work in the K-12 science classroom (Luft, 1999). The use of scientific inquiry standards enables science teachers to use these instructional materials, such as a mobile app, to set curricular expectations and monitor student mastery of science concepts (Jensen, 1996; Lundberg, 1997).

3.4.3. Pair C – feedback vs. feedback

The purpose of this item on the MASS is to provide a score for the existence and quality of feedback in terms of specificity, detail, relevance, and delivery at the point of need. The Feedback subscale on ERMA scores feedback as it relates to improved student performance, and re-teaching. Results indicate very strong correlation between this pairing, $r = .90$, $p < .01$. This result was expected as categories on both rubrics are general and absent of content specificity. Despite concurrent validity between the two items, participants rated apps significantly higher [$t(1, 23) = -4.34$, $p < .01$] on the ERMA Feedback item ($M = 37.50$, SD $= 28.55$) as compared to the MASS Feedback item ($M = 23.61$, SD $= 34.72$). This result may be due to the feedback description on the MASS being more detailed, and thus susceptible to greater scrutiny by the evaluators. As with Pairing B, ERMA simply asks the evaluator to determine the presence or absence of feedback, whereas MASS asks the evaluator to determine the quality and timing of that feedback. In a meta-analysis of seventy five research articles on rubric reliability and validity, Jonsson and Svingby (2007) determined that topic-specific and detailed categories are likely to result in more generalizable and dependable scores when compared to their more general counterparts, a pattern present in the results of Pair C.

3.4.4. Pair D – navigation vs. user friendliness

The aim of this item on the MASS is to score the ease of navigation of the app in terms of design, menus, buttons, functions layouts, and age and audience appropriateness of graphics. On ERMA, User Friendliness refers to the ease with which a student can interact with the app with or without teacher guidance. Results indicate a very strong correlation and concurrent validity of this pairing, $r = .72$, $p < .01$ with participants rating the apps significantly higher [$t(1, 23) = 2.49$, $p = .02$] on the Navigation item of MASS ($M = 84.7$, SD $= 21.9$) than the User Friendliness item of ERMA ($M = 76.0.5$, SD $= 23.8$). Strong correlation in this pairing is not surprising as navigation is a common area of app designer focus (mHIMSS, 2012). Even so, there is a significant difference in the way MASS and ERMA ask the teacher to evaluate navigation and user friendliness. While MASS targets the navigation elements themselves, ERMA measures a student's interaction with the app – an interaction the teacher cannot fully and accurately predict when lesson planning and choosing mobile apps. Evaluators may have scored this item lower on ERMA because ERMA asked them to guess how a student would navigate the app, while MASS asked them to evaluate items present before the lesson occurred. Although, student feedback and interaction with instructional materials and tools is a key component in promoting student engagement and academic growth, these behaviors cannot be replaced with supposition and should, instead, comprise the pedagogical cycle of planning, teaching, evaluating, reflecting and re-teaching (Ertmer et al., 2012; Jonsson & Svingby, 2007).

4. Conclusion

The Mobile App Selection for Science (MASS) rubric (see Appendix A) attempts to provide a framework for the evaluation of science education applications, using a common language structure associated with lab-based technologies and scientific tools. Through four design cycles, this instrument was developed to aid secondary science teachers in the critical exploration and evaluation of individual digital applications towards pedagogical suitability for curricular integration and support. Qualitative data collected during cycles II and III suggest that teacher evaluators continually referred to pedagogical elements, asking how the app uniquely contributed to student learning (Kearney et al., 2012; Ting, 2012). Quantitative data collected in cycle IV further strengthened the argument for content-specific language combined with descriptors tied to curricular standards. Data collected during this cycle indicated a greater level of teacher scrutiny of the pedagogical implications of the chosen mobile app versus technical bells and whistles. The m-learning framework informed the design of this tool and influenced its detailed description of pedagogical elements (Kearney et al., 2012). By focusing our design on the presence of these elements, we developed an instrument that encourages users to place a higher importance on technology integration within the context of the learning environment as a whole.

The use of language in the criteria descriptions that reflects curricular frameworks and standards for 5th–12th grade science, while specific enough to provide a clear benchmark for aligning the use of an app to student learning outcomes; is broad enough to allow for pedagogical flexibility in both choice of scientific topic and choice of technology-supported activity. The content-specificity of MASS is also visible in its description of scientific behavior, reflecting the m-learning framework's emphases on identification of tools that enable *situatedness*, authentic tasks which students complete in actual scientific practice. Thus, the design of MASS supports a more rigorous analysis of the scientific activities and behaviors possible with the selected mobile apps. Two categories included in MASS that were not present in other rubrics, "Accuracy" and "Shared Findings," contribute to its use as a content-specific rubric. Commonly held misconceptions about scientific phenomena are oftentimes more easily dispelled with the use of visualizations and representations that challenge students' knowledge constructs (Green, Chassereau, Kennedy, & Schriver, 2013; Stein, Larrabee, & Barman, 2008). Therefore, accurate scientific information and visualizations, evaluated in the "Accuracy" category, are essential in 5th–12th grade science mobile applications. The "Shared Findings" category addresses a major practice of science as identified by the National Science Teachers Association: "to communicate ideas and the results of inquiry – orally; in writing; with the use of tables, diagrams, graphs and equations; and by engaging in extended discussion with peers" (Bybee, 2011, p. 10).

Nevertheless, the results of this design process are limited to the evaluator population chosen. Further study limitation is present in that evaluators reviewed four out of six categories when comparing MASS and ERMA. Therefore, we plan to conduct further research on the strength of MASS as a content-specific mobile app rubric which addresses these two categories, as well as its use with a larger and more diverse teacher population. Other phases of research on the applicability and practicality of MASS must also explore its use with teachers who are not as comfortable with mobile applications as the teacher evaluators involved in MASS's design. The role of content-specificity in teacher evaluation of mobile apps may be affected by teacher comfort and knowledge of mobile technologies. These continued investigations are crucial to the refinement of tools such as MASS and the role these tools play in critical evaluation of technology tools and applications for K-12 science education.

Appendix A. Mobile App Selection for Science (MASS) Rubric

	0 Not Applicable	1	2	3
Accuracy		Contains some inaccuracy in scientific content. Graphics may be misleading. Inaccurate representation of experimental procedures and measurements.	Content scientifically accurate but limited in scope. Graphics promote limited understanding of science content. Accurate but incomplete representation of experimental procedures and measurements.	Content scientifically accurate. Graphics promote understanding of science content. Accurate and complete representation of experimental procedures and measurements.
Relevance of Content		Loosely aligned with science learning objectives. Does not directly address science literacy. Somewhat current on accepted scientific practices, ideas and discoveries.	Appropriately aligned with science learning objectives. Supports science literacy. Current on accepted scientific practices, ideas and discoveries.	Closely aligned and connected with science learning objectives. Supports and enhances science literacy. Accepted scientific practices, ideas and discoveries current through frequent updates.
Sharing Findings		Findings can be shared and/ or exported through limited means (e.g. email text only).	Findings can be shared and/ or exported through multiple applications (e.g. Dropbox, Google Docs) and/ or social media platforms (e.g. Edmodo).	Findings can be shared and/ or exported through multiple applications including documents, other applications, social media platforms and email.
Feedback		Provides minimal feedback. May not be specific or detailed. May not be provided at point of need.	Provides relevant feedback that may not necessarily be meaningful, detailed or specific. Feedback is provided at point of need.	Provides feedback in response that is meaningful, specific, detailed, relevant and provided at the point of need.
Scientific Inquiry and Practices		Limited opportunities for increased scientific inquiry and practices. Severely limits ability to gather information through observation, experience, reflection, reasoning and communication.	Sufficient opportunities for increased scientific inquiry and practices. May not allow one or more of the following: information gathering through observation, experience, reflection, reasoning and communication.	Multiple opportunities for increased scientific inquiry and practices by allowing information to be gathered through observation, experience, reflection, reasoning and communication.
Navigation		Navigation is a challenge. Design and layout (e.g. menus, buttons, functions) are consistent but relatively confusing. Graphics not age appropriate.	Somewhat easy to navigate. Consistent design and layout (e.g. menus, buttons, functions). Graphics are age appropriate.	Easy to navigate. Consistent design (e.g. menus, buttons, functions) and layout. Graphics are age appropriate.

References

Ackermann, E. (2009). *Piaget's constructivism, papert's constructionism: What's the difference?* Future of Learning Group: MIT Media Laboratory.

Ahmed, S., & Parsons, D. (2013). Abductive science inquiry using mobile devices in the classroom. *Computers & Education, 63*, 62–72.

Banister, S. (2010). Integrating the iPod touch in k-12 education: visions and vices. *Computers in the Schools, 27*(2), 121–131.

Brand, J., & Kinash, S. (2010). Pad-agogy: a quasi-experimental and ethnographic pilot test of the iPad in a blended mobile environment. In *Paper presented at the 27th annual conference of the Australian Society for Computers in Learning in Tertiary Education (ASCILITE), Sydney, Australia*. Retrieved February 5, 2013, from http://works.bepress.com/jeff_brand/18.

Bybee, R. W. (2011). Scientific and engineering practices in the k-12 classroom: understanding a framework for k-12 science education. *The Science Teacher, 78*(9), 10–16.

Carr, J. M. (2012). Does math achievement h'APP'en when iPads and game-based learning are incorporated into fifth-grade mathematics instruction? *Journal of Information Technology Education: Research, 11*, 269–286.

Castelluccio, M. (2010). The table at work. *Strategic Finance, 92*(5), 59–60.

Chen, S., Kao, T., & Sheu, J. (2003). A mobile learning system for scaffolding bird watching learning. *Journal of Computer Assisted Learning, 19*, 347–359.

Clark, R. E. (2005). Five common but questionable principles of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: a critical relationship. *Computers & Education, 59*, 423–435.

Green, L. S., Chassereau, K., Kennedy, K., & Schriver, M. (2013). Where technology and science collide: a co-teaching experience between middle grades science methods and instructional technology faculty. *Journal of Technology and Teacher Education, 21*(4), 385–408.

Guba, E., & Lincoln, Y. (1988). Do inquiry paradigms imply inquiry methodologies? In D. M. Fettermen (Ed.), *Qualitative approaches to evaluation in education: The silent scientific revolution* (pp. 89–115) New York: Praeger.

Hill, R. A. (2011). Mobile digital devices. *Teacher Librarian, 39*(1), 22–26.

mHIMSS. (2012). mHIMSS app usability workgroup. *Healthcare Information and Management Systems Society*. Retrieved February 1, 2013, from http://www.mhimss.org.

Hohlfeld, T. N., Ritzhaupt, A. D., & Barron, A. E. (2010). Development and validation of the student tool for technology literacy (ST^2L). *Journal of Research on Technology in Education, 42*(4), 361–389.

Huang, Y.-M., Lin, Y.-T., & Cheng, S.-C. (2010). Effectiveness of a mobile plant learning system in a science curriculum in Taiwanese elementary education. *Computers & Education, 54*, 47–58.

Hwang, G.-J., & Chang, H.-F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*, 1023–1031.

Jensen, K. (1996). Effective rubric design: making the most of this powerful assessment tool. *The Science Teacher, 64*(5), 34–37.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. *Educational Researcher, 33*(7), 14–26.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review, 2*, 130–144.

Kaput, J., & Hegedus, S. (2002). Exploiting classroom connectivity by aggregating student constructions to create new learning opportunities. In *Paper presented at the 26th conference of the International Group for the Psychology of Mathematics Education*. Norwich, UK.

Kearney, M., Schuck, S., Burden, K., & Aubusson, P. (2012). Viewing mobile learning from a pedagogical perspective. *Research in Learning Technology, 20*, 1–17.

Li, S. C., & Pow, J. C. (2011). Affordance of deep infusion of the one-to-one tablet-PCs into and beyond the classroom. *International Journal of Instructional Media, 38*(4), 319–326.

Luft, J. (1999). Rubrics: design and use in science teacher education. *Journal of Science Teacher Education, 10*(2), 107–121.

Lundberg, R. (1997). Student-generated assessment. *The Science Teacher, 64*(1), 50–53.

McLoughlin, C., & Lee, M. (2008). The 3 P's of pedagogy for the networked society: personalization, participation, and productivity. *International Journal of Teaching and Learning in Higher Education, 20*(1), 10–27.

Murray, O., & Olcese, N. (2011). Teaching and learning with iPads, ready or not? *TechTrends, 55*(6), 42–48.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academies Press.

National Research Council. (2012). *A framework for k-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

Onwuegbuzie, A. J., & Leech, N. L. (2004). On becoming a pragmatic researcher: the importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology, 8*(5), 375–387.

Park, J., Parsons, D., & Ryu, H. (2010). To flow and not to freeze: applying flow experience to mobile learning. *IEEE Transactions on Learning Technologies, 3*(1), 56–67.

Peng, H., Su, Y.-J., Chou, C., & Tsai, C.-C. (2009). Ubiquitous knowledge construction: mobile learning re-defined and a conceptual framework. *Innovations in Education & Teaching International, 46*, 171–183.

Radinsky, J., Bouillion, L., Lento, E. M., & Gomez, L. M. (2001). Mutual benefit partnership: a curricular design for authenticity. *Journal of Curriculum Studies, 33*(4), 405–430.

Roschelle, J., Penuel, W., Yarnall, L., Shechtman, N., & Tater, D. (2005). Handheld tools that 'informate' assessment of student learning in science: a requirements analysis. *Journal of Computer Assisted Learning, 21*, 190–203.

Schrock, K. (2013). *iPads in the classroom*. Retrieved February 18, 2013, from http://www.schrockguide.net/ipads-in-the-classroom.html.

Sha, L., Looi, C.-K., Chen, W., & Zhang, B. H. (2012). Understanding mobile learning from the perspective of self-regulated learning. *Journal of Computer Assisted Learning, 28*, 366–378.

Stein, M., Larrabee, T. G., & Barman, C. R. (2008). A study of common beliefs and misconceptions in physical science. *Journal of Elementary Science Education, 20*(2), 1–11.

Stroup, W. M. (2002). Instantiating seeing mathematics structuring the social sphere (MS3): updating generative teaching and learning for networked mathematics and science classrooms. In *Paper presented at the International Conference of the Learning Sciences*. Seattle, WA.

Sung, E., & Mayer, R. E. (2013). Online multimedia learning with mobile devices and desktop computers: an experimental test of Clark's methods-not-media hypothesis. *Computers in Human Behavior, 29*, 639–647.

Ting, Y.-L. (2012). The pitfalls of mobile devices in learning: a different view and implications for pedagogical design. *Journal of Educational Computing Research, 46*(2), 119–134.

Tinker, R. F., & Krajcik, J. (Eds.). (2001). *Portable technologies: Science learning in context*. New York: Kluwer Academic Press.

Traxler, J. (2007). Current state of mobile learning. *International Review of Research in Open and Distance Learning, 8*(2), 1–10.

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

de Waal, C. (2001). *On pierce*. Bedmont, CA: Wadsworth.

Walker, H. (2011). Evaluating the effectiveness of apps for mobile devices. *Journal of Special Education Technology, 26*(4), 59–63.

Yarnall, L., Penuel, W. R., Ravitz, J., Murray, G., Means, B., & Broom, M. (2003). Portable assessment authoring: using handheld technology to assess collaborative inquiry. *Education, Communication & Information, 3*, 7–55.