"An App! An App! My Kingdom for An App": An 18-Month Quest to Determine Whether Apps Support Mathematical Knowledge Building

Kevin Larkin

Abstract In recent years, applications (apps) for iPads are increasingly being used to support learning in primary school contexts. Current estimates put the number of available educational apps at the iTunes store at approximately 500,000. Many of these apps contain mathematical content and purport to improve students' mathematics ability. Despite their availability, overall ease of use, and low price, significant questions remain as to their appropriateness in helping students develop mathematical knowledge. Three quantitative measures, previously used in other research to investigate digital technology use and student learning, were modified to meet the specific demands of evaluating apps. This chapter reports on the findings of a long-term research project that comprehensively reviewed mathematical apps to determine their usefulness for primary school students. It found that although the majority of apps provide little more than *edutainment*, a core group of apps were highly effective in supporting students in their development of mathematics knowledge.

Keywords Mathematics apps · iPad apps · iPad · Primary school mathematics · ICT and mathematics · Digital manipulatives

The Story Thus Far

This chapter is the culmination of an 18-month quest to determine the appropriateness of iPad applications (apps) to support mathematical learning in primary school students. Its purpose is to synthesise the research literature concerning apps and mathematics, and then outline the methodology used to evaluate the appropriateness of 142 apps which, having met initial criteria, were then assessed using three quantitative measures. The outcomes of this chapter include an evaluation of the appropriateness of the apps for developing conceptual, procedural and declarative mathematical knowledge and also an assessment of the validity of using the Haugland Software Evaluation Scale (1999), the Productive Pedagogies

---

K. Larkin (✉)
School of Education and Professional Studies, Griffith University, Gold Coast, Australia
e-mail: k.larkin@griffith.edu.au

T. Lowrie, R. Jorgensen (Zevenbergen) (eds.), *Digital Games and Mathematics Learning*,
Mathematics Education in the Digital Era 4, DOI 10.1007/978-94-017-9517-3_13

Dimensions (2005), and Gee's Learning Principles (2003) in evaluating mathematical apps.

Determining both the number and the quality of the apps at the iTunes store is problematic. Jonas-Dwyer et al. (2012) and Shuler et al. (2012) independently estimate that in 2012 there were approximately 500,000 apps available. Pelton and Francis Pelton (2012) located approximately 4000 mathematics apps and noted that "while some are commendable, almost all of the rest are simple flashcards, numeric procedures, or mobile textbooks and very few currently available apps have engaged best practices by integrating visual models to support sense-making" (p. 4426). As I have shown elsewhere (Larkin 2013), it is difficult to determine the quality of an app based on information available at the iTunes store as it largely consists of marketing for the app. Whilst this is helpful, such information is provided by the developers and is not 100 % accurate. Because of the minimal amount of information available, exacerbated by the existence of significant time demands on teachers (Leong and Chick 2011), it is likely that teachers are unaware of the existence of quality mathematical apps.

Despite the rapid expansion of the use of apps in the educational domain, there is limited research as to their effectiveness in supporting mathematics learning. Some early research in the use of apps on iPods (Kissane 2011) and iPhones (Yuan et al. 2010) have been conducted. Pelton and Francis Pelton (2011, 2012) conducted research which resulted in them creating a range of applications for the iPhone. Attard and Northcote (2011) and Highfield and Goodwin (2013) presented reviews of categories of apps. Calder (this volume) acknowledges this lack of current research and notes that this has contributed to the ad-hoc implementation of iPads in school contexts. This chapter is the first substantive review of iPad apps, to the author's knowledge, that investigates apps claiming to support the mathematical learning of primary-aged students via the use of three quantitative measures.

The chapter has three aims: first, to critique mathematical apps utilising three quantitative measures previously used in academic research (Haugland Software Developmental Scale, Productive Pedagogies, and Gee's Learning Principles); second, to determine whether these scales—used in other research contexts to evaluate web-based software and digital games—are appropriate for critiquing iPad apps; and third, to generate a range of outputs that will be useful in assisting teachers to make informed choices regarding the use of apps for mathematics education in primary school. Prior to outlining the methodology used, and the findings generated, it is useful to briefly examine the extant literature in relation to the use of Information Communication Technology (ICT), particularly software, to support the teaching and learning of mathematics.

By Way of Background

ICT, when used in developmentally appropriate ways (Haugland and Ruiz 2002; Pelton and Francis Pelton 2011), enhances young students' conceptual and procedural knowledge of mathematics leading to the development of higher order thinking in mathematics (Polly 2011) and improves understanding of number recognition, counting, shape recognition, and composition and sorting (Clements and Sarama 2007). However, whilst technology has the potential to enhance mathematics teaching and learning, the use of technology in drill-and-practice activities has been negatively associated with student achievement (Polly 2011). McManis and Gunnewig (2012) claim that technology can assist student learning only if it is *developmentally appropriate*, i.e., "responsive to the ages and developmental levels of the children, to their individual needs and interests, and to their social and cultural contexts" (p. 16).

A key element in developmentally appropriate technology is software that enables students to become active participants in their own learning (Haugland and Ruiz 2002). Many apps appear to have the potential to enhance learning opportunities for young users (Pelton and Francis Pelton 2011) but this potential is often unrealised. Haugland (1999) sounds a word of caution regarding the often explicit message of designers and marketers that the use of software will accelerate children's learning. One outcome of this intention is that the software (largely drill and practice) is often at the skill level of children 2 or 3 years older than the target audience. This results in children becoming frustrated and not using the software; or children using the software for rote learning with the net effect being that "their retention of concepts is poor as is their ability to apply the concepts to off computer activities" (Haugland 1999, p. 245).

The literature presented thus far indicates that ICT is an appropriate tool to support mathematical learning with the proviso that the software utilised is *developmentally appropriate* and that opportunities for learning by discovery and by instruction are present (Scanlon et al. 2005). A key difficulty is making a determination of appropriateness; thus a number of generic criteria for software use have been proposed in the literature (see Ntuli and Kyei-Blankson 2011; Potter et al. n.d.). What remains unclear is how to determine the appropriateness of iPad apps. Early research into iPad use appears to indicate that young children "learn to use the devices quickly, independently, and confidently and explore freely" (McManis and Gunnewig 2012, p. 15), that such devices have lower costs (thus increasing the likelihood of uptake in schools), and that mathematics apps "seem to be ideally positioned to present mathematical models and manipulatives to support mathematical play, exploration and sense-making both in the classroom and at home" (Pelton and Francis Pelton 2011, p. 2200). However, what remains a key consideration is how educators can come to grips with the explosion in available applications and determine the usefulness of the apps for mathematical learning.

How Do We Know If It's Good or Not?

It is necessary initially to outline the difficulties involved in any research involving iTunes apps. I provide a detailed account of the substantial problem of delineating a clear data set in (Larkin 2013, 2014). It is sufficient to indicate here that the initial location of potentially useful apps is a time consuming and imprecise process and it is possible that some appropriate maths apps were therefore not reviewed. At the conclusion of the initial sorting process, there were 142 apps out of an initial pool of 4000, which were subjected to a full qualitative analysis in terms of their relevance to mathematics curricula, their appropriateness for primary school classrooms, and their ability to develop conceptual and procedural mathematical knowledge. Outcomes of this initial process are available at <http://tinyurl.com/ACARA-Apps>.

This chapter outlines the quantitative analysis of the 142 apps using the three quantitative measures indicated earlier. The decision to use these scales was made for several reasons. First, as there are currently no scales specifically designed for the evaluation of maths apps for iPads, scales used in other domains were modified for use in this research. Second, as all three scales have been used in related domains to evaluate software, they provide a mechanism for the later comparison of my findings to previous related research. Finally, as the Haugland Scale emphasises software design with students as the intended end users, and the Productive Pedagogies and Gee Learning Principles emphasise the potential learning afforded by the apps, their combined use provides a balanced review of the apps in terms of technical features, ease of use for students, and their ability to support mathematical learning.

Process One: Haugland Software Developmental Scale (1999)

The Haugland Software Developmental Scale (Haugland 1999)—henceforth referred to in this chapter as the Haugland Scale—is a criterion-based tool used to evaluate the appropriateness of web-based applications and software for use by children (Haugland 1999; Haugland and Ruiz 2002). The scale is based on ten criteria outlined in Table 1.

It is important to note that the Haugland Scale was not designed to evaluate mathematical apps. Consequently, two important modifications were made for this research. First, in order to analyse the data more thoroughly, the ten criteria were grouped into three sub-clusters (child centred, design, and learning). Second, elaborations were added to emphasise the relationship of the apps to mathematics. In scoring the apps, each of the ten criteria is worth one point and each app can thus score between 0–10. The scoring sheet includes a number of sub-indicators for each criterion. For apps to score a 1 for each criterion they must meet all relevant sub-indicators. If they meet 50 % or more of the indicators a score of 0.5 is recorded, and if less than 50 % are met, a score of 0 is recorded. For example, there are three sub-indicators in the Real-World Model criteria (concrete representation, objects function, simple reliable models). If an app demonstrated all three indicators a score of 1 was given, if two of the three indicators were demonstrated a score of 0.5 was given, and if one or none of the indicators were demonstrated a score of 0 was given.

Process Two: Productive Pedagogies

Productive Pedagogies (Atweh and Bland 2005) are criteria that teachers can use to critique their own practices in order to improve educational outcomes for their

Table 1 Adapted Haugland Developmental Software Scale (Haugland 1999) with clusters and elaborations

Cluster	Criteria	Criteria elaboration with links to mathematics
Child centred	Age appropriate	The mathematics concepts taught by the app reflect realistic expectations for the age children for which it was designed
	Child control	When using the app, children decide the flow and direction for the experience, not the device. They are navigators, determining where the experience will lead, and learn the consequences of their choices
	Independence	While adults may need to assist children in loading the application, after this initial guidance and support, children operate the app with minimal adult supervision
	Non-Violence	Violence in apps is of particular concern because children often initiate and control the violence. In addition, the app models appropriate societal values
Design	Clear instructions	Verbal instructions are essential, since even children who are reading text-based instructions navigate with greater success if audio instructions are also provided. Directions are accompanied with visual prompts and/or a help option
	Technical features	The app is colourful with realistic uncluttered graphics, which enable children to focus on the learning objectives. Graphics are animated to help children attend. Whenever possible, children control the animation, learning mathematics through *hands-on* experiences
	Real-world model	The app provides children with concrete representations of objects found in meaningful and mathematically accurate situations or settings. The scale and colour of the objects are realistic, not stereotypical
Learning	Expanding complexity	The app is an exciting world that is easy for children to enter and reflects children's current cognitive, physical, mathematical and language skills. When children use the application a logical, mathematical learning sequence emerges
	Process orientation	Intrinsic motivation; the desire to explore and experiment and discover mathematics motivates children as they use the app, not rewards. The joy of learning is the reward in using the app.
	Transformations	Apps have the unique potential to give children opportunities to change objects and situations over and over, and discover how different mathematical components impact their world

students. They are pluralistic in nature and do not prescribe a single model of pedagogical practice. There are 20 Productive Pedagogies grouped under four dimensions: intellectual quality, supportive classroom environment, connectedness, and recognition of difference (Education Queensland 2004). Although the Productive Pedagogies were designed for Queensland schools, they share much in common with international pedagogy standards such as those proposed by Newmann et al. (1995) to direct pedagogical change in Wisconsin schools. A number of pedagogies (substantive conversation, deep knowledge, connectedness to the real-world) are included in both the Queensland and Wisconsin pedagogies. Fifteen of the 20 Productive Pedagogies (Education Queensland 2004) were used in this research and the key question for each pedagogy was modified to make each relevant to mathematics (see Table 2).

Productive Pedagogies are not used to evaluate how the app might be utilised by a teacher in a teaching context. What is of interest is determining the effectiveness of maths apps in supporting student learning. Therefore, although the Productive Pedagogies refer largely to teaching, under investigation here is how the app encourages students to develop, for instance, deep thinking or self-regulation, or making connections to previous knowledge. Productive Pedagogies have been previously used in the work of Zevenbergen and Lerman (2007) who used them to investigate teacher and student use of interactive whiteboards.

Each of the 56 applications scoring more than 50 % on the Haugland Scale was evaluated using the 15 Productive Pedagogies in Table 2. When evaluating the apps, if there was no evidence of the individual productive pedagogy, a score of 1 was recorded; if a high degree of evidence was present, a 5 was recorded. Consequently, the range of possible scores for the three dimensions was 30, 25 and 20 respectively and the overall range of scores was 15–75. As was the case with the Haugland Scale, Productive Pedagogies were not designed specifically for mathematics research, therefore modifications to the pedagogies were made, guided by the previous research design of Zevenbergen and Lerman (2007) and were also based on the initial qualitative review of the apps reported in Larkin (2013, 2014 in press).

The first modification was to the key question associated with each of the pedagogies. For instance, in the student direction pedagogy, the initial key question did not include any reference to applications but only to students having a say in the direction or outcome of the learning activities. The more substantive change was the decision not to use the recognition of difference dimension. The recognition of difference dimension consists of five pedagogies: cultural knowledge, inclusivity, narrative, group identity, and citizenship. It became very obvious early in the review that the vast majority of apps were scoring very poorly in the recognition of difference dimension (mean score of 5.3/25). This dimension, therefore, did not add anything methodologically in comparisons among the three scales in judging an app's quality. I will articulate more fully, later in the chapter, why apps made minimal attempt to recognise difference.

Table 2 Productive Pedagogies and Key Questions (adapted from Classroom Observation Booklet by New Basics Branch and the Queensland School Reform Longitudinal Study [QSRLS] commissioned by Education Queensland)

Dimensions	Productive pedagogy	Key question
Intellectual quality	Higher order thinking	Are students using higher order thinking operations whilst engaging with this app?
	Deep knowledge	Does the app cover mathematical content in any depth, detail or level of specificity?
	Deep understanding	Are the students required to demonstrate a deep understanding of concepts or ideas in completing activities associated with this app?
	Substantive conversation	Is there opportunity for dialogue between the student and the app in order to create or negotiate understanding of mathematics content?
	Knowledge as problematic	Are students critiquing ideas and knowledge presented via this app?
	Metalanguage	Are aspects of mathematics language being foregrounded in this app?
Supportive environment	Student direction	Do students have any say in the pace, direction or outcome of the app?
	Social support	Does the app provide a supportive, positive learning context?
	Academic engagement	Does the app encourage student engagement and on-task behaviours?
	Explicit Quality Performance Criteria	Are criteria for student success in the app made explicit?
	Self-regulation	Is the direction of students' activity implicit and self-regulatory?
Connectedness	Knowledge integration	Does the app draw on knowledge from a range of mathematical domains?
	Background knowledge	Does the app scaffold early learning outcomes in the activity with those developed later in the activity?
	Connectedness to the world	Do activities within the app connect with students real-world mathematical experiences?
	Problem-based curriculum	Does the app encourage authentic mathematical problem solving and reasoning?

Process Three: Gee's Learning Principles (2003)

Gee (2003) established a set of 36 principles that underpin learning in digital environments. The use of Gee's work in reviewing digital games is based on the premise that "digital games are user-centred; they can promote challenges, co-operation, engagement, and the development of problem solving strategies" Gros (2007, p. 23). According to Jorgensen and Lowrie (2012), these 36 principles are drawn from three discourses (situated cognition, new literacy studies, and connectionism) and provide a "comprehensive account of the possibilities of games to create exciting and engaging learning opportunities" (p. 379). Table 3 indicates the ten learning principles selected for the evaluation of the apps and includes a modified definition for each.

For this research, the number of principles was reduced from 36 to ten for conceptual and methodological reasons. Based on the experience of the earlier evaluations, it was clear that many of the principles were not applicable for evaluating apps. For example, the self-knowledge principle indicates that learners learn about themselves in a virtual world. None of the apps develop virtual worlds and thus this principle is redundant. Also non-applicable, due to the already noted lack of concern with recognition of difference, were two principles related to cultural knowledge. In terms of methodology, it was considered a cumbersome process to use 36 principles in evaluating apps. In addition, previous research by Jorgensen and Lowrie (2011, 2012) indicated significant overlap on many of the principles. In determining the ten principles to use in this research, I was guided by the work of Jorgensen and Lowrie as well as my earlier qualitative experience of evaluating the apps.

Internal Reliability of the Three Quantitative Measures

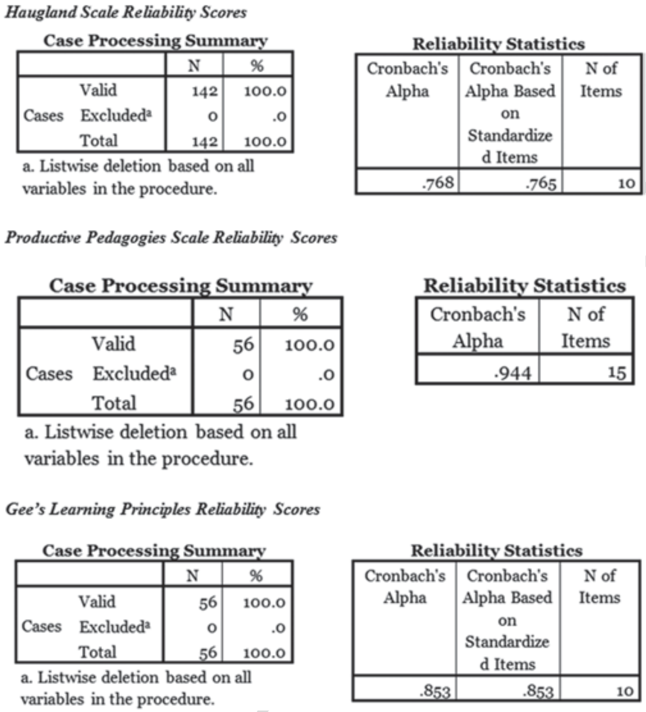
In order to determine the reliability of the three quantitative m easures used in this research, a Cronbach alpha—α was generated. The three individual α- scores are presented in Table 4.

In social science research it is generally accepted that Cronbach alpha scores greater than 0.7 indicate a high degree of internal consistency (Muijs 2011). Briefly, Cronbach's alpha is concerned with the homogeneity of the items that make up the scale or how well the items hang together. In this particular case, we can view alpha in terms of the apps consistency of rating (be that high or low) across each of the domains (Haugland Criteria, Productive Pedagogies, and Gee's Learning Principles). From the data presented here, there is a high degree of confidence that the three scales are internally consistent, and thus we can be confident in their reliability to determine the quality of an app. I address issues of their validity in determining the quality of apps later in the chapter.

Table 3 Selected Learning Principles with modified definitions. (Adapted from Gee 2003)

Principle	Modified definition
Active, critical learning	All aspects of the app environment (including ways in which the semiotic domain is designed and presented) are set up to encourage active and critical, not passive, learning
Semiotic	Learning about and coming to appreciate interrelations within and across multiple sign systems (e.g., images, words, actions, symbols, artefacts) as a complex system is core to the learning experience
Achievement	For learners of all levels of skill there are intrinsic rewards from the beginning, customised to each learner's level, effort, and growing mastery, and signalling the learner's ongoing achievements
Regime of competence	The learner operates within, but at the outer edge, of his/her level of competence so that there is both safety and challenge
Probing	Learning is a cycle of probing the world (doing something); reflecting in and on this action and, on this basis, forming a hypothesis; re-probing the world to test this hypothesis; and then accepting or rethinking the hypothesis
Multiple routes	There are many ways to complete the app, each of which caters for the strengths and interests of the learner
Situated meaning	The meanings of signs (e.g., words, actions, artefacts, symbols, texts) are situated in embodied experiences. Meanings are not general or decontextualised. Whatever generality meanings come to have is discovered bottom up via embodied experiences
Practice	Learners get lots and lots of practice in a context where the practice is not boring (i.e., in an environment that is compelling to learners on their own terms and where the learners experience ongoing success). They spend lots of time on task
Discovery	Overt telling is kept to a well-thought-out minimum, allowing ample opportunity for the learner to experiment and make discoveries
Transfer	Learners are given ample opportunity to practice, and support for, transferring what they have learned earlier to later problems, including problems that require adapting and transforming that earlier learning

Table 4 Cronbach alpha reliability scores for the three scales



Haugland Scale Reliability Scores

Case Processing Summary

		N	%
Cases	Valid	142	100.0
	Excluded[a]	0	.0
	Total	142	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.768	.765	10

Productive Pedagogies Scale Reliability Scores

Case Processing Summary

		N	%
Cases	Valid	56	100.0
	Excluded[a]	0	.0
	Total	56	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.944	15

Gee's Learning Principles Reliability Scores

Case Processing Summary

		N	%
Cases	Valid	56	100.0
	Excluded[a]	0	.0
	Total	56	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.853	.853	10

The News Is…

The following section will briefly recap the findings from the qualitative review (as it contributes to the discussion of app quality) before examining in detail the findings from the various quantitative measures.

Qualitative Analysis

I have accounted for the qualitative review of the apps that are discussed here elsewhere (Larkin 2013, 2014 in press). In this chapter, the focus here is on considering the types of mathematical knowledge (Miller and Hudson 2007) developed by the apps. Conceptual knowledge involves a deep understanding related to the meaning of mathematics. Procedural knowledge is the ability to follow a set of sequential steps to solve a mathematical task. Declarative knowledge is information that students retrieve from memory without hesitation.

Table 5 is a summary of the number of apps supporting the development of conceptual, procedural, or declarative knowledge, or a combination of them. In percentage terms, 44.4% of the apps developed only declarative knowledge and

Table 5 Total number and percentage of apps developing differing forms of mathematics knowledge

Type of knowledge	Number of apps ($n=142$)	Percentage* *(to nearest 0.1)*
Declarative	63	44.4
Procedural	42	29.6
Conceptual	12	9.9
Both conceptual and procedural	14	8.5
Both conceptual and declarative	2	1.4
Both procedural and declarative	7	4.9
All three knowledge types	2	1.4

52.1 % developed a combination of declarative and other types of knowledge. These percentages reflect findings in relation to iPod and iPhone mathematical apps (Kissane 2011; Pelton and Francis Pelton 2011). It is not suggested that declarative knowledge is of itself a negative as declarative knowledge provides an important foundation for procedural knowledge with the student accessing facts to complete a task (Bottge in Miller and Hudson 2007). What is suggested by the data is that declarative knowledge is overemphasised and that conceptual, and to a lesser degree, procedural knowledge is undervalued in mathematical apps. iPad software appears to be able to support manipulative devices and pictorial representations useful for conceptual development, and to provide sequential scaffolding for procedural mastery, yet most of the apps do not do so.

Quantitative Analysis

In order to get an overall sense of the data from the three scales, basic descriptive data on the three quantitative measures is provided in Table 6.

As an initial observation, a wide range of scores was recorded for each of the three measures. This clearly indicates that, even after the initial reduction of the

Table 6 Descriptive statistics of the three measures

Measure	Haugland scale	Productive pedagogies	Gee learning principles
N of cases	142	56	56
Possible total	10	75	50
Minimum	1	29	12
Maximum	10	71	45
Range	9	42	33
Mean	5	54.8	23.8
Std deviation	1.99	10.64	7.40

4000 apps to just 142, and the subsequent further reduction of apps from 142 to 56, that there is still a very wide discrepancy in the quality of the apps. It is not the case that there are only one or two high or low outliers in the data, as the large standard deviations for each of scales indicates that there was a consistent and large spread of quality across the app range. In terms of mean scores for each of the measures, it is evident that the apps scored most highly according to the Productive Pedagogies (73 %), attained a low pass mark on the Haugland Scale (50 %), and performed poorly on the Gee Learning Principles (43 %).

This initial statistical data supports the findings of the descriptive review where it was very clear that although there were some very strong apps, most were very poor quality. Once again, this comment takes into account only the 142 apps that were considered worthy of substantive examination and suggests that the vast majority of apps do not support the development of mathematical knowledge.

Process One: Haugland Scale

Table 7 indicates the top 20 apps according to the Haugland Scale; however, to indicate the results of all 142 apps, overall mean scores have been included.

The data indicates that the apps were strongest in the child-centred cluster (2.96/4) but weak in the other two clusters (design 1.35/3; learning 0.69/3). In terms of some of the individual criteria, the apps were quite strong on (independence 0.85/1; and non-violence 0.89/1) but extremely poor on (expanding complexity 0.25/1 and transformations 0.16/1). As the Haugland Scale is not specific to mathematics but designed to determine appropriateness of software for children, it was used in this research as a first cut measure of the apps. I took the research decision that if an app could not meet the three core demands of the Haugland Scale then the app was not appropriate to use. In essence, the Haugland Scale is very useful in weeding out the poor applications, and any app that did not score more than five was excluded from further investigation. Only 56 of the 142 apps initially evaluated using the Haugland Scale were considered appropriate for any further investigation. However, a score over 50 % does not in itself provide sufficient information for a decision to be made regarding the app's appropriateness in terms of developing mathematical knowledge. Consequently, two different quantitative measures were used to further evaluate the quality of the 56 remaining apps.

Process Two: Productive Pedagogies

The second quantitative evaluation procedure utilised 15 Productive Pedagogies. I have previously indicated that the recognition of difference dimension, with a mean score of 5.3/25, will not be used in further comparison of the apps. There are a number of reasons for the overall poor scores in this dimension. First, there is a global market for the apps and therefore any customisation for specific cultural

Table 7 Haugland Scale: Top 20 apps

Clusters on Haugland Scale	Child	Design	Learning	Total
Area of Rectangles	4	3	3	10
Miracle Learning for Calculation	4	2.5	3	9.5
Early Numbers: Maths Wizard Counting	3.5	2.5	3	9
Hands-On Equations—Lite	3.5	3	2.5	9
I See! Math 1	4	2.5	2.5	9
Mathemagica—Kids Math	4	2.5	2.5	9
Common Core Number	4	2	3	9
Adding Beads	4	2	2.5	8.5
Find and Count	4	2.5	2	8.5
Learn Numbers: Learn2Count	4	2.5	2	8.5
Math Dream	4	1.5	3	8.5
Maths Skill Builders—Primary	3.5	3	2	8.5
Fact Families—Add/Subtract	4	2	2	8
Friends of Ten	4	2.5	1.5	8
Hands-on Maths Attribute Blocks	4	2	2	8
Learn Math 1 (Mondiso)—Add/Subtract	4	2.5	1.5	8
Marble Math Junior	4	2.5	1.5	8
Time Math Free	3.5	2.5	2	8
123 Counting Fun Lite	4	2	1.5	7.5
Hands-on maths Number Sense	3	2	2.5	7.5
Overall mean for 142 apps	*2.96*	*1.35*	*0.69*	*5.01*

groups is problematic. Second, the lack of recognition of difference relates to the notion of *gamification*, discussed by Bossomaier (this volume) who suggests that in many cases, game-like elements are superficially added to digital resources in order for them to mimic games. For example, a rote learning app may be *gamified* by providing an opportunity for users to play a game if they score more than 18/20 in the maths tasks. Unfortunately, the *gamification* of the apps further minimises any potential recognition of difference as the rewards are generic in nature. Regardless of the exact reason, it is clear that most apps do not cater for diversity.

Table 8 provides summary data based on the three dimensions of intellectual quality, supportive environment, and connectedness. Although 56 apps were reviewed using these dimensions, data in this table reports only on apps scoring more than 37.5 (50 %).

An examination of the data provided in Table 8 indicates that 39 apps met the greater than 50 % criteria. The mean score for these apps was 49.2/75 (66 %). This percentage score was reasonably consistent across the three dimensions of intellectual quality (63 %), supportive environment (69 %), and connectedness (65 %). This again supports the claim made earlier that the Productive Pedagogies are internally consistent to a high degree. In terms of intellectual quality, three of the six criteria scored highly (deep knowledge 4/5, understanding 3.9/5, and metalanguage 3.7/5). The apps may have scored highly on these criteria because many apps are designed as knowledge generators, at least in terms of declarative knowledge, and to a lesser degree procedural knowledge. So the clear design intent of the apps was, for example, that students improve their multiplication facts or practice the division algorithm. In developing this type of knowledge, most of the apps used appropriate mathematical language and this accounts for the high score on the metalanguage criteria. In contrast, apps scored quite poorly on substantive conversation (1.4/5) and knowledge as problematic (2.2/5). Substantive conversation scores lowly by virtue of the fact that it is a substantial coding challenge for designers to cater for user interaction. The low score also relates to the overarching issue of diversity, and there is no acknowledgement in the apps that a conversation might be required with the user to better tailor the apps for their individual experience. The low score on the knowledge is problematic criteria correlates with low scores in terms of the connectedness dimension and will be discussed later in this analysis.

The apps score consistently across the supportive environment dimension with a mean score of close to 3.5/5 for four of the five criteria. This is not surprising as the apps are designed for independent use by young children, primarily in the home environment. Consequently, there are scaffolds in place to assist students in the design of the apps. The one dimension in this section, which scored slightly lower than the others, was self-regulation with a mean score of 3.1/5. This can be explained by the observation that self-regulation correlates with catering for individuals, and it has already been established that apps do not cater for this high degree of individuality.

There was a little more variation in the scores for the four criteria comprising the connectedness dimension. Two criteria scored well: connectedness to the real-world (3.7/5) and background knowledge (3.5/5); however the remaining two criteria scored poorly: knowledge integration (2.9/5) and problem-based learning (2.9/5). It

Table 8 Productive Pedagogies results for apps scoring 50 % or more overall

Productive pedagogies themes	IQ	SE	C	TOTAL
Possible maximum	*30*	*25*	*20*	*75*
Mathemagica	28	23	20	71
Area of Rectangles	28	22	16	66
Early Numbers: Maths Wizard Counting	24	22	14	60
Marble Math Junior	22	21	17	60
Miracle Learning for Calculation	24	23	12	59
Math Galaxy Fractions	21	21	17	59
Math Model	24	21	14	59
I See! Math 1	23	19	16	58
Find and Count	21	23	12	56
Common Core Number and Operations	21	19	16	56
Hands-On Maths Number Sense	23	18	15	56
Learn Numbers: Learn2Count	19	22	15	56
Hands-On Equations	20	20	14	54
Friends of Ten	20	19	14	53
Adding Beads	21	17	14	52
Learn Math 1 (Mondiso)	19	19	13	51
Statistics!!!	20	17	14	51
Math Dream	21	19	10	50
Maths Skill Builders—Primary	18	18	14	50
Fun Count App	19	17	12	48
Patterns, Colors and Shapes	17	17	14	48
123 Counting Fun	18	17	12	47
Fact Families—Add/Subtract	18	15	12	45

Table 8 (continued)

Productive pedagogies themes	IQ	SE	C	TOTAL
Visual Math 1 & 2	18	14	12	44
Astromat Lite	16	18	9	43
Hands-on Maths Attribute Blocks	17	14	12	43
Math Grade One	17	15	11	43
Middle School Math	17	15	11	43
Kindergarten Math	15	17	10	42
Base Ten Number blocks	18	13	10	41
Abby Adventure Winter Maths	13	17	10	40
Adventure Basic School Math	13	17	10	40
Column + - */	15	13	12	40
Number Skills	17	15	8	40
Red Dragonfly Mathematics Booklet	17	10	13	40
Telling Time HD	16	14	10	40
Time Math Free	14	14	12	40
Know Your Maths Facts Free	14	13	12	39
Probability Tools	16	9	14	39
Mean of apps scoring above 50 %	19.0	17.4	12.9	49.2
Std deviation	3.7	3.5	2.5	8.5

has already been established that apps are strong as declarative knowledge generators so it makes sense that the app designers take some cognisance of what children already know and build upon that throughout the apps by connecting with the real-world experience of the children (at least in a generic sense). The lack of knowledge integration and problem-based learning in the apps reflects the fact that most apps are designed as stand-alone apps targeting a particular type of knowledge or content area (e.g., adding common fractions, subtraction of two-digit numbers). There were very few apps that went deeper than this to connect different content areas in mathematics (e.g., common, decimal, percentage and proportional reasoning knowledge). This of course may relate to limitations with the available coding software on iPads. I suggest, however, it is more likely due to the desire of the designers to quickly develop and sell high volumes of a product in the one to two dollar range and the associated unwillingness to invest time and money into the development of a more substantive product for which there may be only a limited market. In summary, the findings from the Productive Pedagogies dimension review indicate that 12 of the 56 apps scored 75 % or higher overall and can be confidently recommended for use with primary-aged students. A further 11 apps scored between 60 and 75 % overall and thus have some worth. The remaining 33 apps have only limited use in developing mathematical knowledge.

Process Three: Gee's Learning Principles

The final quantitative measure used to evaluate the apps was the selected Gee's Learning Principles, henceforth referred to as GLP, and only 24 of the 56 apps evaluated scored more than 50 % (see Table 9). In scoring the apps with this scale, each app could score from 1 (no evidence) to 5 (very strong evidence).

As was the case with the previous two scales, it is very clear that, according to GLP scores, there is a wide range in the quality of the apps. This applies across the 56 apps that were reviewed using these principles but is also evident in the 24 apps that scored more than 50 %. This again clearly indicates that there is a wide gulf between quality apps, however they are measured, and the majority of apps. The two principles that scored most highly across these 24 apps are the semiotic principle (mean score of 4/5) and the active learning principle (3.96/5). From the data already analysed using the Productive Pedagogies, this is not particularly surprising as the semiotic principle relates very closely to the metalanguage criteria, which scored highly. Likewise the observation of students being actively involved in their learning correlates with high scores on the academic engagement pedagogy as the apps are designed to support and encourage student learning.

The type of active learning evident in many apps is, however, different to that envisaged by Gee (2003) in the gaming environment context where the activity that users demonstrate is oriented towards a range of narrative goals. In many instances in using apps, student activity is solely related to completing a level in order to receive a non-related reward. In addition, the fact that the type of active learning

Table 9 Gee Learning Principles: Results for apps scoring 50 % or more overall

Gee learning principles	ACL	SEM	ACH	COMP	PROB	MULT R	SIT MEM	PRAC	DISC	TRANS	TOTAL
Math galaxy fractions	5	5	4	5	4	5	4	4	4	5	45
Hands-on maths attribute blocks	5	5	4	4	4	3	4	3	5	5	42
Area of rectangles	5	5	3	3	4	3	3	3	4	4	37
Mathemagica	4	3	3	5	2	5	3	5	2	4	36
Marble Math Junior	4	4	4	4	3	4	3	3	3	2	34
Adding Beads	4	3	4	4	2	4	2	3	4	3	33
Maths Skill Builders	3	4	3	4	2	4	3	4	2	4	33
Friends of Ten	4	5	3	2	2	1	5	3	3	4	32
I See! Math 1	4	5	4	3	2	4	3	3	2	2	32
Miracle Learning for Calculation	4	5	3	4	2	2	3	4	3	2	32
Statistics!!!	5	5	2	4	2	4	2	2	2	4	32
Tens Frame	4	2	2	4	4	4	3	4	4	1	32
Math Dream	5	4	1	2	4	2	5	1	4	2	30
Hands-On Maths Number Sense	4	5	2	2	2	2	3	3	3	2	28
Probability Tools	4	2	1	2	5	1	3	1	5	4	28
Solids Elementary HD	5	3	1	1	4	3	3	1	5	2	28
Find and Count	4	4	2	2	1	2	5	3	3	1	27
Geometry 4 Kids	3	4	2	1	1	2	4	3	3	4	27
Middle School Math	3	3	3	2	2	4	1	4	3	2	27
Early Numbers: Wizard Counting	4	4	3	2	1	3	1	2	2	4	26
Toddler Counting 123	3	4	3	2	2	2	5	2	2	1	26
123 Counting Fun	3	4	3	2	1	3	2	3	2	2	25

Table 9 (continued)

Gee learning principles	ACL	SEM	ACH	COMP	PROB	MULT R	SIT MEM	PRAC	DISC	TRANS	TOTAL
Math Grade One	3	4	3	2	2	4	2	1	3	1	25
Visual Math 1 & 2	3	4	3	3	2	2	3	2	2	1	25
Mean	*3.96*	*4.00*	*2.75*	*2.88*	*2.50*	*3.04*	*3.13*	*2.79*	*3.13*	*2.75*	*30.91*
Std deviation	*0.75*	*0.93*	*0.94*	*1.19*	*1.18*	*1.16*	*1.15*	*1.10*	*1.03*	*1.36*	*5.25*

encouraged is mainly declarative knowledge, remains problematic. For example, in the Maths Alien app, the reward for passing a level is shooting alien ships, and in Monty's Quest the user gets to help a mouse push cheese up a hill as a reward for solving division problems. It is likewise unsurprising that the achievement principle (0.9/5), the probing principle (1.2/5), and the transfer principle (1.4/5) scored poorly. Higher scores can only be attained in these principles via a high level of customisation and the longer-term development of complex game narratives. Both of these dimensions are almost non-existent in apps. These scores mirror the findings of the knowledge integration and problem solving pedagogies, and again indicate that apps make little to no attempt to recognise any diversity in the end users.

A second point to consider is why so many of the apps failed to score more than 50 % on the GLP. Firstly, and most obviously, the apps are not designed like games. The types of games that were reviewed in Gee's initial research, and also, to some degree, those used in the work of Jorgensen and Lowrie (2011, 2012), were full-featured video or digital games. These games had the opportunity to develop narratives and often offer multiple routes for solutions. Most apps are not designed in this fashion as their focus is *gamification* where, as indicated previously a simple rote learning activity is enhanced via the use of minor game elements. So the genre of many of the apps could be labelled *edutainment* where mathematics knowledge may develop as a consequence of students striving for the gameplay at the end of the task (see Nansen et al. 2012). The apps are also designed to promote *fun* learning and this facet is heavily promoted in marketing to parents. Finally, as indicated earlier, cost is a clear factor in the creation of these apps with the developers hoping for large sale volumes rather than making a serious investment into the creation of a quality game.

At the conclusion of the quantitative analyses, the question remains whether the failure of many apps to score highly according to GLP is an indication that the apps are poor, or an indication case that GLP is evaluating qualities that are not necessary in a quality mathematical app. An alternative hypothesis is that GLP are in fact evaluating quality in the apps; however, they are doing so using a stricter measure of quality than is the case with the Productive Pedagogies. Therefore, is it the case that both scales measure similar aspects of quality apps, with the GLP measuring that quality at a higher level of compliance than the Productive Pedagogies? The final contribution of this chapter is to determine whether the Productive Pedagogies and the GLP are both measuring similar quality in the apps, but are doing so in a different fashion. If the answer is "Yes" to this question, then teachers can use either of these measures and be confident in determining the quality of an app.

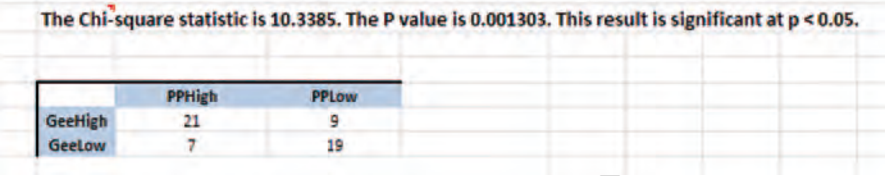
Correlation Analysis

In order to determine the correlation between the apps determined as being of high quality using the Productive Pedagogies and those determined using GLP, a Spearman's Ranked correlation on the two variables was performed. This data is presented in Table 10.

Table 10 Spearman’s ranked correlation for Gee learning principles and productive pedagogies

	Gee total	Intel qual	Sup envir	Connect	PP total
Gee total	1.000				
Intel qual	0.602	1.000			
Sup envir	0.415	0.783	1.000		
Connect	0.568	0.842	0.683	1.000	
Pp total	0.551	0.950	0.903	0.875	1.000

Table 11 Chi-square and *p* values for correlation



The Chi-square statistic is 10.3385. The P value is 0.001303. This result is significant at $p < 0.05$.

	PPHigh	PPLow
GeeHigh	21	9
GeeLow	7	19

Ma and Kishor (1997) suggest that correlations ranging from 0.20 to 0.40 can be considered practically meaningful in behavioural sciences and indicate that “a correlation of 0.30 is actually equivalent to an increase of 30 % in the success rate of an intervention” (p. 27). The correlation coefficient (rho) between the variables shown in Table 10 indicate a moderate positive correlation between the Productive Pedagogies and GLP scores overall, and very high correlations between the three Productive Pedagogies dimensions. This is not surprising given the high Cronbach alpha scores reported earlier. The correlation analysis suggests that although the scales are determining quality using different criteria, both scales are delineating similar apps as being of high quality. This is significant as it answers “Yes” to the question posed earlier, inferring that teachers can be confident in using either measure to assist them in determining the quality of an app.

In terms of this particular research, given that I have measured the apps using both scales, it is reasonable to conclude that the combined distance from the median scores on Productive Pedagogies and GLP will provide a very accurate measure of the quality of the apps reviewed. Table 11 indicates the location of the 56 apps according to a measure of distance from the median score using a Chi-square measure.

It is evident from Table 11 that of the 56 apps evaluated, 21 scored above the median scores in both Productive Pedagogies and GLP; 7 were above on Productive Pedagogies but below on GLP; 9 were below on Productive Pedagogies but above on GLP; and 19 were below on both measures. There were other apps which scored above the median overall; however, these were not included as they scored below the median in one of the two individual measures. Table 12 provides further information on the 21 apps that scored above median values for both measures.

Based on this median data, I am confident in reporting that any of the apps listed in Table 12 are very useful in assisting students to develop mathematical knowledge in primary school contexts. In addition, although the Productive Pedagogies dimensions are more closely related to school classrooms and therefore easier to use

Table 12 Median scores for Productive Pedagogies & Gee and overall combined median score

Apps	PP Total	Median	Median dev	High low	Gee Total	Median	Median dev	High low	Med dev sum
Mathemagica	71	42.50	28.50	1	36	22	14	1	42.50
Math galaxy fractions	59	42.50	16.50	1	45	22	23	1	39.50
Area of rectangles	66	42.50	23.50	1	37	22	15	1	38.50
Marble math junior	60	42.50	17.50	1	34	22	12	1	29.50
Miracle learning for calculation	59	42.50	16.50	1	32	22	10	1	26.50
I see! Math 1	58	42.50	15.50	1	32	22	10	1	25.50
Early numbers: maths wizard	60	42.50	17.50	1	26	22	4	1	21.50
Friends of ten	53	42.50	10.50	1	32	22	10	1	20.50
Adding beads	52	42.50	9.50	1	33	22	11	1	20.50
Hands-on maths attribute blocks	43	42.50	0.50	1	42	22	20	1	20.50
Hands-on maths number sense	56	42.50	13.50	1	28	22	6	1	19.50
Find and count	56	42.50	13.50	1	27	22	5	1	18.50
Statistics!!!	51	42.50	8.50	1	32	22	10	1	18.50
Maths skill builders	50	42.50	7.50	1	33	22	11	1	18.50
Math model	59	42.50	16.50	1	23	22	1	1	17.50
Learn numbers: learn2count	56	42.50	13.50	1	24	22	2	1	15.50
Math dream	50	42.50	7.50	1	30	22	8	1	15.50
Hands-on equations	54	42.50	11.50	1	22	22	0	1	11.50
123 counting fun	47	42.50	4.50	1	25	22	3	1	7.50
Middle school math	43	42.50	0.50	0	27	22	5	1	5.50
Visual math 1 & 2	44	42.50	1.50	0	25	22	3	1	4.50

for teachers, either measure will determine an app's quality. The apps scored more highly on the Productive Pedagogies, perhaps because this measure is designed for more formal educational contexts. In addition, the apps are not designed as video games, but rather as small-scale, content-specific learning applications and thus have not scored as well on a scale designed to evaluate video games.

So Do Apps Cut the Mustard?

The process outlined in this chapter has established that there is a large discrepancy in the quality of apps available at the App Store, with many of limited to no use at all in terms of mathematical learning. However, I do not wish to be a 'prophet of doom' in relation to their use by children. Despite the fact that many apps are marketed with glib promises of accelerating student learning or making learning fun, and the observation that many apps clearly do not meet the criteria of *serious* digital games as suggested by Bossomaier (this volume) and Beavis (this volume), the apps listed in Table 12 certainly do 'cut the mustard' and are highly innovative in terms of supporting mathematical knowledge. The three tools used for moderating the quality of the apps have shown that there is high intellectual quality in the final cut of the apps. This process has highlighted the potential of these apps to promote deep learning across a number of areas in mathematics.

Three apps in particular, as evidenced by the median scores, are clearly exceptional. Mathemagica is an innovative application, using a range of digital images and sounds, to provide an engaging experience for students in developing a diverse range of concepts including Number, Place value, the null property, and order of operations. Although a different style of app, Area of Rectangles combines a range of activities, utilising similar technology evident in full-scale virtual manipulatives websites such as Illuminations or the National Library of Virtual Manipulatives (NLVM) to develop conceptual, procedural, and declarative knowledge. Finally, Maths Galaxy Fun uses a range of pictorial representations to develop conceptual understanding of fractions. In addition, students can complete a range of step-by-step tutorials and thus be in control of their learning. What these apps demonstrate is that the iPad, as a technological tool, is a capable platform for the delivery of quality mathematical apps for primary-aged children. Given this, and pending future confirmation by classroom teachers and students, I am confident at this stage to answer "Yes" to the implied question rhetorically posed in the title. There are apps currently available that support the development of mathematical knowledge. What is critical in terms of the plethora of apps that are on the market is that caution and care must be enacted if quality learning (as opposed to rote activities) is the desired outcome. The quantitative measures described and implemented in this paper have shown that there are means by which educators are able to discern quality apps for mathematics learning.

What is of greater concern, given the vast number of available apps, is the issue of how teachers, under significant time pressures, can accurately determine whether an application will develop the type of deep learning evidenced in the selection of apps recommended in Table 12. I have argued in this chapter and elsewhere that teachers cannot rely solely on information from the iTunes store. Qualitative reviews of apps—such as the one I conducted in 2013, or those available at a range of educational websites—are useful but limited by a number of factors, e.g., assessor subjectivity, range of access to apps, and most significantly, the reviews become quickly dated as existing apps are deleted or updated and new apps become available. The quantitative measures utilised in this research bring a high degree of academic rigour to the evaluation of apps. With a high degree of confidence, the data indicate that an app scoring highly on either measure is indeed a quality app capable of supporting deep learning. The measures are internally consistent, relatively simple to use, assess the types of knowledge that teachers expect their students to develop, and more objective than current qualitative measures, and the determination of quality is more easily communicated to colleagues via numerical scores.

This chapter has demonstrated that the use of Productive Pedagogies and Gee's Learning Principles, either together or independently, given the degree of positive correlation, is a means by which teachers (educators or even parents) can confidently identify whether or not an app will support deep, connected learning beyond the normal confines of formal schooling. This is a fundamental and profound shift in approach to the evaluation of apps. Until now, rigorous quantitative measures for evaluating apps have been unavailable, and teachers have been obliged to rely on largely anecdotal and often prejudiced accounts of educational quality. Using the measures proposed in this chapter positions teachers as educational leaders, confident in selecting and using apps with their students that will enhance deep and connected learning. This confidence, when communicated to colleagues, will encourage those who may have been reluctant to use iPad apps in their mathematical practice, to likewise engage with the technology and thus offer an enhanced range of learning opportunities for their students. The use of the quantitative measures designed in this research is a substantive contribution to overcoming the difficult problem of sorting the "wheat from the chaff" in terms of mathematical apps.

This chapter has reported on the process by which the author evaluated a range of mathematical apps. The next phase of the research is currently underway and involves groups of teachers, at a range of schools, using the quantitative measures to evaluate apps. This process will assist in the fine-tuning of the evaluative measures discussed in this chapter and will also enhance the quality of the current reviews available to classroom practitioners. Future research will then involve the use of some of the recommended apps with primary school students to begin to measure the impact of app use on student learning.

Acknowledgements I would like to acknowledge the contribution of Dr. Todd Milford with some of the data analysis presented in this chapter.

References

Attard, C., & Northcote, M. (2011). Teaching with technology: mathematics on the move: using mobile technologies to support student learning (Part 1). *Australian Primary Mathematics Classroom, 16*(4), 29–31.

Atweh, B., & Bland, D. (2005). Mathematics through/for understanding social life: Productive pedagogies meets critical mathematics. Paper presented at the *Mathematic Education and Society Proceedings of the 4th International Conference*. Brisbane.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the building blocks project. *Journal for Research in Mathematics Education, 38*(2), 136–63.

Education Queensland. (2004). *Productive pedagogies: Classroom reflection manual* [website]. State of Queensland (Department of Education). http://education.qld.gov.au/public_media/reports/curriculum-framework/productive-pedagogies/.html. Accessed 21 Dec 2012.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Gros, B. (2007). Digital games in education: The design of games-based learning environments. *Journal of Research on Technology in Education, 40*(1), 23–38.

Haugland, S. (1999). Computers and young children: The newest software that meets the developmental needs of young children. *Early Childhood Education Journal, 26*(4), 245–254.

Haugland, S. W., & Ruiz, E. A. (2002). Computers and young children empowering children with technology: Outstanding developmental software for 2002. *Early Childhood Education Journal, 30*(2).

Highfield & Goodwin. (2013).

Jonas-Dwyer, D. R. D., Clark, C., Celenza, A., & Siddiqui, Z. S. (2012). Evaluating apps for learning and teaching. *International Journal of Emerging Technologies in Learning, 7*(1), 54–57. doi:10.3991/ijet.v7i1.1901.

Jorgensen, R., & Lowrie, T. (2011). Digital games: Creating new opportunities for mathematics learning. Paper presented at the *Mathematics: Traditions and [New] Practices* (Proceedings of the 34th annual conference of the Mathematics Education Research Group of Australasia and the Australian Association of Mathematics Teachers). Alice Springs.

Jorgensen, R., & Lowrie, T. (2012). Digital games for learning mathematics: Possibilities and limitations. In J. Dindyal, L. P. Cheng, & S. F. Ng (Eds.), *Mathematics education: Expanding horizons* (Proceedings of the 35th annual conference of the Mathematics Education Research Group of Australasia, pp. 378–384). Singapore: MERGA.

Kissane, B. (2011). Mathematics education and the iPod touch. Paper presented at the *Mathematics: Traditions and [New] Practices*. (Proceedings of the 34th annual conference of the Mathematics Education Research Group of Australasia) and the Australian Association of Mathematics Teachers). Alice Springs.

Larkin, K. (2013).

Larkin, K. (2014 in press).

Leong, Y. H., & Chick, H. L. (2011). Time pressure and instructional choices when teaching mathematics. *Mathematics Education Research Journal, 23*(3), 347–362. doi:I0.I007/s13394-011-0019-y.

Ma & Kishor. (1997).

McManis, L. D., & Gunnewig, S. B. (2012). Finding the education in educational technology with early learners. *Young Children*, 14–24.

Miller, S. P., & Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learning Disabilities Research & Practice, 22*(1), 47–57.

Muijs, D. (2011). *Doing quantitative research in education with SPSS* (2nd ed.). London: Sage.

Nansen, B., Chakraborty, K., Gibbs, L., Vetere, F., & MacDougall, C. (2012). 'You do the math': Mathletics and the play of online learning. *New Media & Society, 14*(7), 1216–1235. doi:10.1177/1461444812442926.

Newmann, F., Marks, H., & Gamoran, A. (1995). Authentic pedagogy: Standards that boost student performance. *Center on Organization and Restructuring of Schools*. Madison. Wisconsin.

Ntuli, E., & Kyei-Blankson, L. (2011). Teacher criteria for evaluating and selecting developmentally appropriate computer software. *Journal of Educational Multimedia and Hypermedia, 20*(2), 179–193, 1–52.

Pelton, T., & Francis Pelton, L. (2011). Design principles for making meaningful mathematics apps. Paper presented at the *Proceedings of Society for Information Technology & Teacher Education International Conference 2011*. Chesapeake.

Pelton, T., & Francis Pelton, L. (2012). Building mobile apps to support sense-making in mathematics. Paper presented at the *Proceedings of Society for Information Technology & Teacher Education International Conference 2012*. Chesapeake.

Polly, D. (2011). Examining how the enactment of tpack varies across grade levels in mathematics. *Journal of Computers in Mathematics and Science Teaching, 30*(1), 37–59.

Potter, J., Johanson, J., & Hutinger, P. (n.d.). *Creative software can extend children's expressiveness*. http://www.wiu.edu/users/mimacp/wiu/articles/software4.html. Accessed 12 Dec 2012

Scanlon, M., Buckingham, D., & Burn, A. (2005). Motivating maths? Digital games and mathematical learning. *Technology, Pedagogy and Education, 14*(1), 127–140.

Shuler, C., Levine, Z., & Ree, J. (2012). *iLearn II: An analysis of the education category of apple's app store*. New York: The Joan Ganz Cooney Center at Sesame Workshop.

Young, M., Slota, S., Cutter, A., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming in education. *Review of Educational Research, 82* (1), 61–89. doi:10.3102/0034654312436980.

Yuan, T., Chae, H. S., Nantwi, G., & Garg, P. (2010). An Examination of 300 iPhone Math Applications: Implications for the Development of an Educational Math Game. Paper presented at the *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*. Chesapeake: AACE.

Zevenbergen, R., & Lerman, S. (2007). Pedagogy and interactive whiteboards: Using an activity theory approach to understand tensions in practice. *Mathematics: Essential Research, Essential Practice, 2*, 853–862.

Kevin Larkin is a Lecturer in Mathematics Education in the School of Education and Professional Studies at Griffith University in Australia. Kevin's research interests and expertise include student attitudes to mathematics in primary school, the experience of transition into high school and its implications for mathematics education, mathematics education for pre-service teachers in blended and online contexts, and the nexus between digital technologies and mathematics education. He is involved in four different research teams investigating these research interests, including the $AUD 3.2 million research project funded by the Office for Learning and Teaching, *Step up! Transforming Mathematics and Science Pre-service Secondary Teacher Education in Queensland*.