Tablet-Based Math Assessment: What Can We Learn from Math Apps?

Gabrielle A. Cayton-Hodges[1*], Gary Feng[1] and Xingyu Pan[2]

[1]Educational Testing Service, Princeton, NJ 08541, USA // [2]Cognitive Science Research Group, University of Michigan, Ann Arbor, MI 48109, USA // gcayton-hodges@ets.org // gfeng@ets.org // xypan@umich.edu

[*]Corresponding author

ABSTRACT

In this report, we describe a survey of mathematics education apps in the Apple App Store, conducted as part of a research project to develop a tablet-based assessment prototype for elementary mathematics. This survey was performed with the goal of understanding the design principles and techniques used in mathematics apps designed for tablets. We focused our reviews on four areas, (1) the quality of mathematical content, (2) feedback and scaffolding, (3) richness of interactions, and (4) adaptability of the applications. These four areas were cultivated from prior research on digital tools in mathematics (e.g., Digital Tools for Algebra Education criteria; Bokhove & Drijvers, 2011), designing principles of learning objects (e.g., Learning Object Evaluation Metric; Kay & Knaack, 2008), as well as quality of mathematics instruction (e.g., Hill et al., 2008). We end with recommendations for tablet assessment design cultivated through this review.

Keywords

Technology, Tablets, Formative assessment, Elementary school, Mathematics

Introduction

The era of tablet computers as wide-spread consumer devices began in 2010 with the announcement of the Apple iPad. In 3 years, the tablet shipment has surpassed either desktop or laptop PCs, and outsold all PCs combined in the fourth quarter of 2013 (IDC, 2013). Tablets see even stronger demands in the education sector (Interactive Education Systems Design, Inc., 2013). In the educational assessment arena, tablet devices have been recognized as an alternative test delivery platform, and various validity issues have been studied (see, for example, Laughlin Davis, Strain-Seymour, & Gay, 2013; Strain-Seymour, Craft, Davis, & Elbom, 2013). How to design educational assessments specifically for tablet computers, however, remains largely an uncharted territory.

Change is imminent, though. Decisions by prominent testing consortia such as PARCC and Smarter Balanced to support touch-screen devices have pushed tablet-based assessments into the spotlight (see Smarter Balanced Assessment Consortium, 2012). High stakes tests are not the only impetus for educators, students, and parents when they order tablets by the thousands (see Blume, 2013); instead, they are drawn to the numerous possibilities that tablets hold, including the idea that tablets will personalize learning and improve student learning outcomes. A key to achieving this goal is in the form of formative assessments to track student learning and provide timely feedback. In sum, tablet-based assessments – both summative and formative – are scheduled to launch shortly. Are we ready?

The answer depends on one's expectations. If the question is whether we are ready for tablet delivery of assessments that were designed for paper and pencil or PCs, the answers, and debates, can be found in various technical analyses and usability studies (e.g., Isabwe, 2012; Kowalski, Kowalski, & Gardner, 2009; Kim, Lim, Choi, & Hahn, 2012; Toby, Ma, Lai, Lin, & Jaciw, 2012; Laughlin Davis et al., 2013; Tsuei, Cho, & Chan., 2013; Strain-Seymour et al., 2013). The consensus seems to be that the technology can be made to work. But if the goal is for tablet-based assessments to be as engaging and interactive as the many apps that students are already enjoying on tablets, we have a long way to go.

This report is a part of a research project under the Cognitively Based Assessment of, for, and as Learning (CBAL) initiative at ETS that aims to build a model for an innovative K–12 assessment system using learning sciences and other related research towards: documenting what students have achieved (of learning); helping to identify how to plan instruction (for learning); and being a worthwhile educational experience in and of itself (as learning, see Bennett, 2010). This survey was inspired by work of the first author to develop a tablet-based assessment prototype for elementary mathematics, namely, students' understanding of fractions and decimals. When we began the project, we found that most published research on the use of tablets in mathematics assessments focused either narrowly on specific features of the tablet technology, such as the usability of digital ink (e.g., Ren & Moriya, 2000; McKnight & Fitton, 2010; Kim et al., 2012), or broadly on the relationship between classroom tablet usage and end-of-semester student learning outcomes (Hieb & Ralston, 2010; Isabwe, 2012; Kowalski et al., 2009). Given the apparent lack of a

systematic review on how to *design* mathematics assessments for tablet computers, we decided to start with a survey of mathematics education apps with the hope of understanding their design principles and techniques. We also looked for potential pitfalls that we hope to avoid in our work. This report is a summary of this qualitative research.

In what follows we will outline the dimensions of mathematical content and interactions used as the basis for review. We then outline the method we used to sample the apps and the evaluation criteria. In the Results section we summarize the main findings from our survey. The Discussion section focuses on the lessons learned from the survey and key take-aways for design of mathematics assessment and instruction applications (i.e., what works and how we can apply these techniques in our tablet-based assessment design).

Dimensions of mathematical review

We focused our review on four dimensions: (1) the quality of mathematical content, (2) feedback and scaffolding, (3) richness of interactions, and (4) adaptability of the applications. These four areas were cultivated from prior research on digital tools in mathematics (e.g., Digital Tools for Algebra Education criteria; Bokhove & Drijvers, 2011), from the evaluation of mobile apps in algebra (Harrison, 2013), and from designing principles of learning objects (e.g., Learning Object Evaluation Metric; Kay & Knaack, 2008), as well as from the quality of mathematics instruction (e.g., Hill et al., 2008). These dimensions are described below.

Mathematical content

We looked specifically at two main dimensions of mathematical content, derived from the Mathematical Quality of Instruction (MQI; see Hill et al., 2008): mathematical accuracy and mathematical richness. Mathematical accuracy is adapted from the MQI dimensions of "error" and "imprecision." Moyer-Packenham, Salkind, and Bolyard (2008) define mathematical accuracy as "the degree to which the mathematical object is faithful to the underlying mathematical properties of that object in the virtual environment" (p. 204). In this case, we're looking at whether the mathematical content in the app is true to its pure mathematical form, and, if not, where does the inaccuracy lie. The richness of mathematics includes two elements: attention to the meaning of mathematical facts and procedures and engagement with mathematical practices (Hill et al., 2008). A mathematically rich experience captures the meaning of the mathematical practices and provides links, explanations, and generalizations.

Feedback and scaffolding

An essential condition for supporting student learning and improving chances of success is to provide feedback and scaffolding (Gibbs & Simpson, 2004; Hattie & Timperley 2007; Nicol & MacFarlane-Dick 2006). We specifically looked at three dimensions: feedback, scaffolding, and opportunity for reflections. The Feedback dimension, derived from Digital Tools for Algebra Education criteria (Bochove & Drijvers, 2010), includes whether any feedback is given, whether the feedback is relevant to the input and the content, the timeliness of the feedback, and the types of feedback (e.g., conceptual, procedural, corrective).

The Scaffolding and the Opportunity for Reflection dimensions were derived from the Learning Object Evaluation Instruction (Haughey & Muirhead, 2005), illustrating "opportunities to extend and expand learning activities to beyond the confines of the object itself". The Scaffolding dimension includes whether the app provides any scaffold to help student learning and, if so, the forms of scaffolds (e.g., hints and guiding questions). The Opportunity for Reflection dimension focuses on whether students have opportunities to think about or explain their thought process. This reflection may occur in the form of verbal or typed explanation, transfer to other problems, or answering questions related to the solution process.

Richness of interactions

Meaningful interactions between students and the apps not only contribute to student learning outcomes (Chan & Black, 2006), but also influence the quality as well as the interpretation of the students' actions. We focused on the

two elements of interactions, derived from the assessment criteria in Digital Tools for Algebra Education (Bokhove & Drijvers, 2010): the modes of interactions and the item types offered.

Modes of interactions may include training, practice test, actual test, and others (Bochove & Drijvers, 2010). Modes are used to organize selective presentation of different attributes within the same app. For example, scaffolding and hints may be available only in training and practice test modes but not in the actual test mode. An app may allow different item types, from static multiple-choice questions to dynamic simulation of real world settings where students interact realistically with data of considerable complexity within a multimedia environment. Different item types allow different opportunities to observe and assess student behavior and reasoning (Gorin, 2006). In this case, we're looking at all interactions present during app use including interactions with the mathematics, the scenario, the digital features, as well as the device.

Scoring and adaptability

Scoring of students' performance is a central component of assessment development. Scores may be used to evaluate student performance and provide real-time feedback for the student, yet the meaning of the scores depends on how the scores are calculated. For instance, the meaning of a score is different when only accuracy is considered as compared with when both accuracy and speed are considered. Scores may also be used as part of a student profile and mastery account to administer appropriate questions (adaptability), while the adaptability will also influence the scoring algorithm. We focus on two elements within this category, namely scoring method and adaptability. Scoring method focuses on which variables were included in score calculation within each app (e.g., correct/incorrect, number of hints needed, time, etc.), while adaptability concerns whether and how each app provides user-dependent content (e.g., on-demand hints vs. on-error hints, adapting to user ability vs. always increasing difficulty, etc.).

Method

Sample

We decided to focus on mathematics education apps on the Apple App Store for two reasons. First, the Apple iOS platform is the most popular tablet platform for the target population (4^{th}- to 5^{th}-grade students, see Mainelli, 2013) and second, the Apple App Store contains many more apps as well as user reviews than other platforms.

We sampled the mathematics education apps on the Apple App Store during the summer of 2013 using the following approach. First, we obtained a list of the top 100 most popular education apps, of which 12 were mathematics oriented. Then we obtained the names of an additional 52 apps that were featured in either the "education collection" for math in iTunes or iTunes Education Spotlight (http://www.apple.com/itunesnews/education/us/).

From this list of 64 apps, decisions to review an app were based on a number of factors: calculator apps were excluded, apps with few downloads and/or low ratings from users were excluded, and apps that were obvious clones of other apps were excluded. Additionally, some apps were offered in series, with each individual app featuring one topic (e.g., algebra, fraction, counting, etc.). For each series, only the most popular app (i.e., the one with the most downloads) was included. Finally, we excluded some paid apps with a high cost for practical reasons of budget limitations. This process resulted in a sample of 16 apps downloaded for review (see Appendix for the list of apps reviewed).

The review process

Selected apps were downloaded and installed on an iPad device and a single researcher reviewed all sixteen apps, allowing 10 - 25 minutes of interaction with each app, depending on the complexity of the design and the variability of interactions available. Additional information regarding each app was collected from user ratings and written comments on the App store. Video reviews of each app were also collected through a keyword search on Youtube.com. A second researcher then reviewed the sixteen apps, along with all notes for general agreement on comments and interactions.

Notes were taken according to the four dimensions outlined in the Introduction. Notes included the positive and negative aspects of the app with regards to each dimension. Because our primary goal is to learn from tablet-based mathematics apps in general rather than to evaluate this particular sample of apps, we did not pursue quantitative analyses of the notes. We will present a qualitative summary of the lessons learned from our interactions with the apps.

Results

Range of mathematics apps

The initial sample of sixty-four applications, as well as the sixteen-app subset reviewed, covered a wide range of mathematics topics, including numbers and operations, algebra, geometrics, and statistics and probability. The target age of the applications ranges from preschool to adults, with most apps focusing on younger populations, namely from preschool to elementary school-aged children.

The apps also differ in genres. Some apps take the form of e-textbooks, providing instructional materials as well as unit-based assessments (e.g., HMH Fuse series, Woot Math). Others take the form of a personal tutor, offering video clips or demonstrations of procedures (e.g., Long Division; Khan Academy). The vast majority of apps, however, take the form of games, in which the player needs to solve mathematics problems to earn points and achieve game goals (e.g., DragonBox Algebra, Motion Math series, Teachley; Addimal Adventure, etc.).

Interestingly, we did not find any apps that claim to be assessment apps, notwithstanding some that are test-prep in nature. On the other hand, both the e-textbook and the game applications typically have built-in assessment elements. Therefore the current review will still provide insights for building tablet-based math assessments.

Evaluation of selected mathematics apps

Mathematical content

When reviewing the content of the sixteen selected apps, we looked specifically at two main dimensions, derived from the Mathematical Quality of Instruction (MQI; see Hill et al., 2008): mathematical accuracy and mathematical richness.

Mathematical accuracy

As stated earlier, mathematical accuracy is adapted from the MQI dimensions of "error" and "imprecision." Moyer-Packenham et al., (2008) define mathematical accuracy as "the degree to which the mathematical object is faithful to the underlying mathematical properties of that object in the virtual environment" (p. 204).

The mathematics apps appear mostly accurate in mathematical content, but sometimes conscious design decisions are made to sacrifice mathematical accuracy for ease of use or to match the user expectations. For example, *DragonBox Algebra*, is designed to teach students how to solve for the value of an unknown. Certain valid mathematical solutions are not possible, however, depending on the level of the game, presumably to eliminate distractions for students performing at lower levels. For example, sometimes both sides of the equation can be multiplied by the same number, but not divided. Another type of example can be seen in the problem: $x \times x = x/3$. The user is expected to divide both sides of the equation by x, obtaining the solution $x = 1/3$. This, however, makes it impossible to get the alternative [valid] solution of $x = 0$. These and other examples may generate or reinforce mathematical misunderstandings.

Mathematical richness

The richness of mathematics includes two elements: attention to the meaning of mathematical facts and procedures and engagement with mathematical practices (Hill et al., 2008).

Meaning making includes explanations of mathematical ideas and drawing connections among different mathematical representations. Most applications are doing poorly in this aspect. They typically focus on the retrieval of mathematical facts using simple response item types (such as numeric response consisting of a single digit) with automatically generated items. They may also focus on how to carry out procedures (e.g., fraction addition with different denominators, long division, etc.). We did not find any apps in our review that engaged users in activities where they need to explain or reflect on why a procedure works or why certain strategies do not work.

As for drawing connections among different representations, apps vary greatly. Some apps offer only Arabic numerals and algebraic expressions (e.g., *TouchyMath* by Joel Martinez), while others present multiple representations without explicitly drawing connections between them. For example, *DragonBox Algebra* offers graphic demonstration of the additive inverse relationship. An object icon and its color-inverted counterpart will cancel each other once they are put on top of each other. It also offers numeric expressions of additive inverse relationships, such as $5 + (-5) = 0$. No explicit connections, however, are drawn between the graphic and the numerical representation. This is part of the design philosophy of *DragonBox Algebra*, which claims to teach rules of algebra surreptitiously in a game without explicit references to numbers and variables. In contrast to the above, some apps focus explicitly on mappings among different mathematical representations. For example, in *Motion Math HD: fraction*, users need to identify identical fractions represented in different forms, including area models, number line models, fractions, percentages, and decimals.

Overall, most apps do poorly on engaging users in meaningful mathematical practices. Mathematical practices include the presence of multiple solution methods, the development of mathematical generalizations from specific examples, and the fluent and precise use of mathematical language (Hill et al., 2008). In our search, we were unable to find any apps on the Apple App store that require users to apply multiple strategies to solve a problem. Efficiency of the solution is usually evaluated through the speed of users' responses. But some apps adopt a different approach and evaluate the efficiency of the solution by the number of moves used to solve a problem. For example, *DragonBox Algebra* has set a maximum number of moves for each item. The user will have to isolate the unknown variable on one side of the equation before he/she runs out of moves. In this way, users need to solve each problem in a relatively efficient way, rather than through a trial-and-error approach.

In sum, our impression is that most apps we reviewed provide relatively accurate mathematical content, yet, in one way or another, most apps are inadequate in the mathematical richness dimension. This may be due in part to limitations of the genre of the mathematics apps – in a game that emphasizes speed, there is often little room for multiple solutions or reflections. In other cases the narrow focus may be a design decision made implicitly (e.g., the author may be unaware of the richness) or explicitly (e.g., *DragonBox Algebra*, which insists on not making explicit connections between in-game actions and algebraic operations).

Feedback and scaffolding

Although opportunities for reflection are few and far between in current mathematics education apps, feedback and scaffolding are available on most of them, taking on different forms.

Relevant and timely feedback

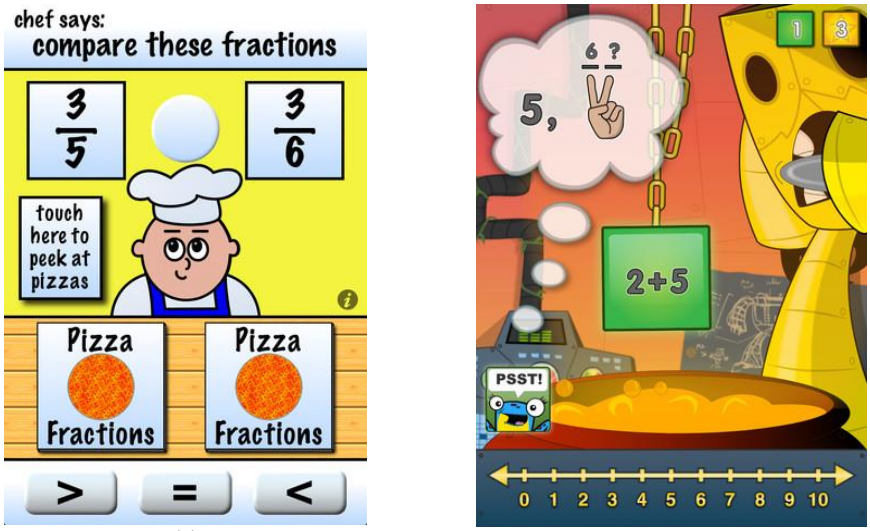
Feedback on user performance can take three forms: (a) It can give *status feedback* on the problem being solved or on the mathematical objects being examined; (b) it can be *corrective feedback*, letting the user know a mistake has been made and guiding the user to correct the mistake; (c) it can be *conceptual feedback*, asking the user questions that cause the user to reconsider his or her perceptions of the objects.

A majority of the apps explored offer corrective feedback upon finishing an item. For example, *Algebra Touch* by Regular Berry Software LLC clearly indicates when an expression is fully simplified or an equation is solved. Invalid actions are immediately identified visually, audibly, and tactilely as parts of the equation vibrate to indicate where and why the action was invalid. As for status feedback, a score board/bar is present in most apps to provide information about a user's overall performance so far. Users may also choose to go to a status board to review their performance on each item.

Compared with the corrective and status feedback, conceptual feedback is less common. For example, in *DragonBox Algebra*, invalid actions are visibly and audibly identified (although as discussed before, some otherwise valid actions are not allowed), but reasons for their invalidity are not given. On the other hand, some apps provide conceptual feedback during the user's problem-solving experience. For example, *BuzzMath Middle School* will offer students a complete problem-solving procedure and explanation when they answer questions incorrectly. Some apps will not only provide feedback on what is the right solution and why it is right, but also provide feedback on why the given solution is wrong. In *Touchymath* by Joel Martinez, when invalid actions are attempted, the app highlights terms or operations to indicate why the actions were invalid. When a user submits a solution that is not yet simplified, the app will indicate what parts of the solution can still be simplified. Conceptual feedback can also take the form of video lectures. For example, *Khan Academy*, which works as a video library and an extension of its website, syncs user activity across the platform, and links the user to video lectures of topics on which there was poor performance.

Scaffolding

Some applications offer scaffolding when students have consistently experienced difficulties. The scaffolding is typically in the form of hints or guiding questions that are offered when students demonstrate evidence of difficulty in problem solving, such as prolonged reaction time, or an incorrect first attempt. Hints are also offered when students request them (on-demand hints), regardless of their current performance. An example of an on-demand hint can be seen in *Pizza Fractions* by Brian West, in which students can choose to "peek at pizzas" to see the area model of the magnitude of fractions before they indicate their choice (Figure 1a). On-demand hints may not be tailored towards the specific item. For example, in *Numerosity: Play with multiplications*, a user's request for a hint will activate a generic multiplication table. An example of as-needed hints comes from *Teachley: Addimal Adventure* by Teachley (Figure 1b). As shown in the picture, the block 2+5 is hanging by two chains. One chain is cut if no response is given after 2 seconds. Once one chain is cut, the hint for that problem jumps out. The hint on this picture suggests that the user should apply the "count on" strategy; for users unfamiliar with this strategy, a tutorial is available outside of the gaming interface.



(a) (b)

Figure 1. Example of hints in Pizza Fractions (a) and Teachley: Addimal Adventure (b)

Opportunities for reflection

The apps we reviewed rarely promoted reflections through sense-making questions to stimulate user's summarizing of mathematical rules. Some apps, however, do make available a complete history of users' problem-solving actions at the end of the task (e.g., *TouchyMath*). The purpose of this is to afford students an opportunity to review their

problem-solving process. Without an explicit benefit or gain in the context of the game though, users are likely to skip this step and proceed to the next task.

Richness of interactions

The number of interactions available sets the limit for the possible user actions while using an app. Available interactions also influence how students approach a task. For example, different modes (i.e., studying, practice, or test) would influence students' strategy choice as well as their level of engagement (Roediger, Agarwal, McDaniel, & McDermott, 2011; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011). Moreover, item types (e.g., multiple choice, short answers, essays) also influence how students approach tasks (Butler and Roediger, 2007).

Modes of interactions

By modes of interactions we mean phases of a game or an application that elicit distinct patterns of user interactions, for examples, tutorials, practices, competitions or main task sets, reviews or scoreboards, etc. Each mode has a different learning goal.

Figure 2. Example set up of practice questions in *DragonBox Algebra*

Some apps offer practice questions via a game tutorial for students to get familiar with the types of interactions that will be used in the game. For example, *DragonBox Algebra* provides a tutorial to introduce the set-up of the screen, the different elements of the game (e.g., the "zero" card, the "one" card, the "dragon" card), as well as the goal of the game (i.e., isolate the "dragon" box on one side; see Figure 2). Users need to follow the instruction to finish a task before they start freely interacting with the game.

The practice mode is also used before or in mid-game, typically when students have difficulty with certain problems. For example, in *Hands-on Equation*, students may choose to watch a short video-clip for help. After watching the video, they need to finish two practice problems before getting back to the "exercise" interface.

Item types

In most of the apps there are fairly clear units of interactions that are analogous to test items in an assessment. They may be in the form of arithmetic problems to be solved one after another, or challenges to be passed in order to advance to the next level.

A great variety of items are available in these apps, although constructed response items are uncommon. Most items are variations of multiple-choice questions. Some items vary the content of the choices; instead of numbers, equations, or texts, choices may include pictures, graphs, or even video clips. Variations are also included in the layout of the choices. Instead of having a finite number of choices horizontally or vertically aligned, the number of choices may be infinite, (e.g., locations on a number line). Moreover, variations may also happen with the choice-making actions. Some choice-making movement involves novel interactions other than tapping. For example, in the *Motion Math* series, students need to tilt the tablet for the object to fall on the choice area located on the two lower corners of the screen. Drag-and-drop actions are also used in multiple-choice questions, where users need to drag the selected choices to desired locations.

Some apps use a combination of grid-in and multiple-choice format in presenting their questions. For example, in *Numerosity: Play with Multiplication!*, users may encounter vertical multiplication of multi-digit numbers. They need to select the number for the units first. The choices will refresh after that and they can select the number for tens (see Figure 3), then hundreds and thousands, if applicable.

Figure 3. Example of combining grid in with multiple-choice in *Numerosity: Play with Multiplication!*

Constructed response items usually involve interactions with virtual manipulatives, such as fraction strips and place value blocks. Items may require students to use manipulatives to represent mathematical entities. For example, in *Hands on Math: Base Ten Blocks*, users can drag unit blocks, ten blocks, and hundred blocks into virtual "trays" to represent multi-digit numbers. Items may also require students to use manipulatives to represent mathematical operations. For example, Figure 4 comes from *Woot Math* by Nimbee LLC. To solve this problem, users need to drag corresponding unit fraction "slices" into the "fraction circle", in this case 3 "1/10" slices and 2 "1/5" slices, to show the process of 3/10 + 2/5. Moreover, users will also need to drag 2 "1/10" slices on top of each "1/5" slice to simulate the process of finding equivalent fractions.

Handwritten input was rare among the reviewed apps. When it was used, users' handwritten input may be left "as is" (e.g., *Woot Math*) or translated into text input for further processing (e.g., *FluidMath*, *Todo K-2 Math Practice*). Most of the apps that did allow handwritten input were usually interactive graphing tools. For example, *FluidMath* can translate a user's handwritten input into an equation, then sketch a plot of the graph of that equation, including multiple "objects" on the same plot. Figure 5 shows a sketched ellipse, line, and parabola that were translated into a single graph, with the simultaneous presentation of their algebraic expressions. When the user changes the location or shape of the objects on the coordinate plane, the algebraic expression changes accordingly and vice versa.

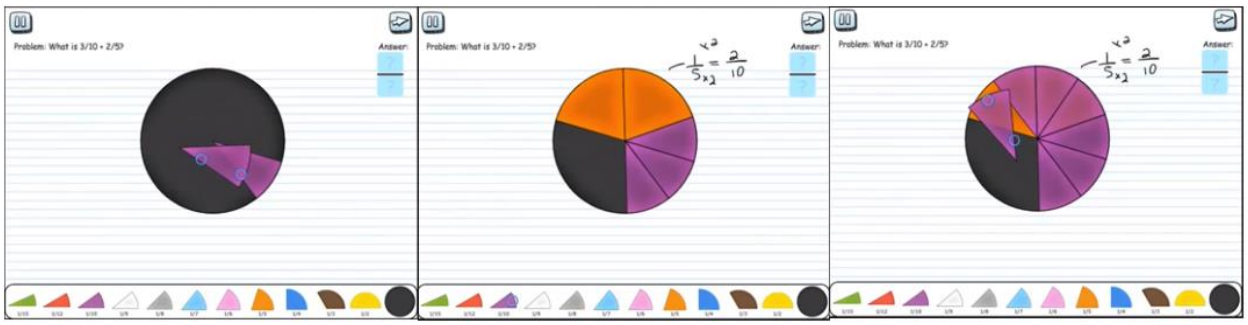


Figure 4. Use of virtual manipulations in Woot Math

In sum, the richness of interactions we found in the mathematics apps is far greater than that of typical educational assessments, including the latest technology-enhanced items. Not only do tablets afford more varieties of item presentations and responses (e.g., intuitive gesture commands, accelerometer functionality, potential for handwritten responses, ability to include drawings to accompany text responses), they also create opportunities to collect data on response processes that are typically not present in traditional assessments (e.g., recording the order in which a diagram is drawn, attempts/resets on a problem before submitting the answer).

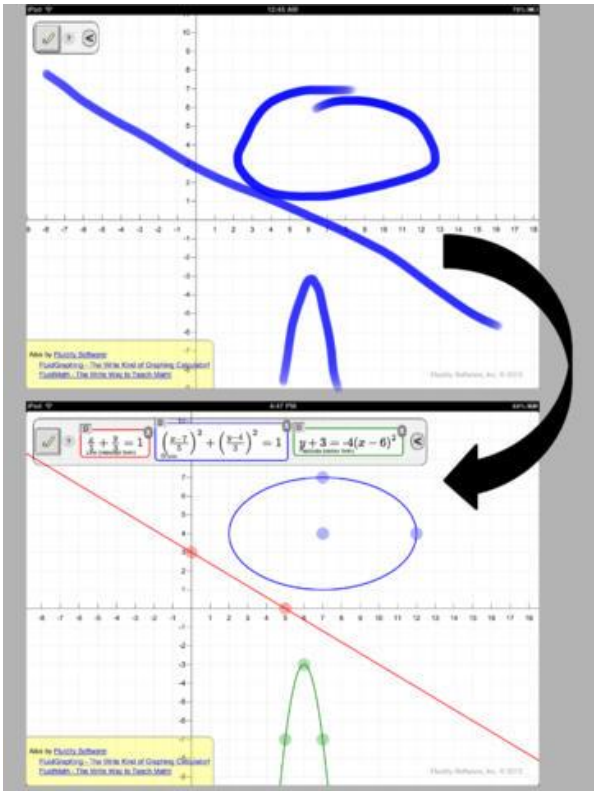


Figure 5. Sketch to graph recognition function in *FluidMath*

Scoring and adaptability

Scoring of student performance is a central component of assessment development. Scores can be used not only to evaluate student performance, but also to provide real time feedback for students in various ways. Lower scores on a particular topic indicate more difficulty experienced and the need for additional work on that topic. Scores may also be used to determine what items may best fit a student's current level of competence so as to keep the student engaged during assessments.

Scoring

Most apps mark a response as either right or wrong, and use percent correct as an indicator of user's overall performance. Speed may also be considered in scoring. For example, in *Math Champ Challenge*, the points one may receive on an item decrease with time. Item difficulty is typically not considered in scoring, so that all items are weighted equally in calculating one's final score. Scores in these apps are only meaningful within the apps themselves; no apps in the sample we reviewed claim any external validity of their scores, either as evidence of competency or as a basis for decisions to be made outside of the virtual world of the app. This is a critical difference from educational assessments.

Adaptability

Not all applications adapt to user responses. The ones that do typically use one of three ways to adapt the difficulty of items to user's ability: The first approach is to let the user select a difficulty level. For example, *Middle School Math HD* by Interactive Elementary allows a user to pick between one of the three difficulty levels: easy, medium, and hard, before the user starts the game play. A second approach is to offer items of increasing difficulty, while successful completion of an "episode" is required to unlock the next one (e.g., *Fraction Planet* by Playpower labs). The last approach involves the dynamic presentation of questions. Users still need to finish the current task before they can unlock the next task, but which task will be presented next is determined by their performance on the current task (e.g., *Woot Math*).

In addition to the above notes, we also observed various usability issues in the mathematics apps. Instead of enumerating them here, we shall discuss ways to avoid them in the context of assessment design in the next Section.

Discussion

We conducted this survey of mathematics apps with the intent to learn what we should (and should not) do in developing innovative mathematics assessments on tablet devices. While the landscape of mathematics apps may seem vast, there are only a few categories of gameplay, interaction, and adaptability that are currently at play. These features may influence the how students engage in problem solving and learning. They also have an impact on how we can interpret students' interactions with the interface as well as their final answers. These are critical concerns as we design the next generation tablet-based mathematics assessments. In what follows we will discuss some of the lessons learned and how they may apply to tablet-based assessment design. Many of these recommendations may apply to all Technology Based Assessment (TBA) but we leave the recommendations as specific to tablets since the interactive nature of the tablet interface is fundamentally different from a computer and students have been shown to interact with tablets and computers in different ways and with different results (see Abrams, Davoli, Du, Knapp, & Paull, 2008; Davoli, Du, Montana, Garverick, & Abrams, 2010; Davoli & Brockmole, 2012).

Mathematical content representations

By design or by accident, many mathematics apps we reviewed have inadequacies in representing the accuracy and richness of the mathematics content. While this may be forgiven in games and informal learning environments, such flaws may create systematic errors in scoring and interpretations when it comes to assessments. Fortunately, in the educational assessment field, there is a rigorous review process in place to guarantee the content accuracy, which works well for familiar item types.

One challenge we are likely to grapple with, though, is how to continue this strong tradition as assessment tasks become increasingly complex, interactive, and game-like. We observed that in some cases when a mathematics problem is operationalized in a game-like setting, where mathematical concepts map to manipulatable objects on the tablet, valid mathematical operations do not always translate to legitimate physical operations on the tablets, such as the issue described earlier in *DragonBox*, where the allowed actions preclude obtaining one of two valid answers. This may not be a concern for gameplay, but it could undermine the validity of an assessment. Thus one lesson learned is: *In creating an interactive game-like assessment item, be very careful about how mathematical concepts*

and operations map to objects and actions in the virtual world; thoroughly review the mapping early in the task design stage.

Interactive items

One indisputable attraction of these apps is that they provide a wide range of rich and engaging interactions around mathematics content. This is what users – students, parents, and teachers – come to expect of an "iPad experience." This, undoubtedly, is also the expectation they will bring to tablet-based assessments. The bar is set very high, and the assessment industry has much ground to cover to catch up with the apps. However, at the same time, the level of rigor needed is much higher for assessments than low/no-stakes games and certain liberties taken with content for the sake of interactivity is not acceptable for high quality assessments.

We take high quality interactive item types as a given for any serious attempts to create tablet-based mathematics assessments; they are what students have come to expect of all things tablet, and, when done right, they engage students in mathematical thinking. We nevertheless advise against creating interactive items only for the sake of engagement. As we pointed out earlier, without a clear focus and a meticulous execution, it is easy to lose sight of the construct and end up developing a fun but uninterpretable distraction.

From an assessment point of view, the biggest value of interactive tasks is that they provide data on the intermediate steps and strategies students use to solve problems. It is often hard to infer why a student answered an item incorrectly based on the final answer alone. Having a complete record of the problem-solving process helps in determining the cause of the error. Such information can help to interpret student performance in an assessment, or to determine scaffolding or instructions one might need. We provide three illustrations:

- **Final products and response processes.** Most applications use students' answering action to calculate scores. Besides accuracy and speed, students' change of answers can also shed light on their strategy use and knowledge. For example, higher-ability students are less likely to change their answers, although they are more likely to benefit more from their answer change (McMorris et al., 1991). Recording the sequence of answering action may also reveal the context in which students change their answers, which may help identifying the cognitive process involved in such changes. For example, changing answers on previous items after answering later questions may suggest students used information from later questions to update their thoughts. Changing answers while reviewing before submission, on the other hand, may suggest they worked the problem for a second time and identified an error.

- **Navigation through the task.** Navigation through a task can also inform us about preexisting knowledge. For example, a student may skip a question he or she is uncertain with at first, coming back to the question later. Studies on test navigation are scarce, but current evidence indicates that high-ability students are less likely to go back and forth while completing the assessment (Kim et al., 2012).

  It is worth noting that task navigation reflects not only student task processing, but also the structure of the assessment. For example, many formative assessments allow students to go back to previous items and ahead to future items via arrows from the current screen. Some apps, however, allow student to go to any previously-solved item from a review board (e.g., *Woot Math*, see Figure 6). Design of assessment flow will influence (a) whether students can skip a problem and come back later, (b) whether students can engage in self-monitoring actions such as marking an answer as unconfident or problematic, and (c) how students would navigate back to a previous item (see Pan, Cayton-Hodges, & Feng, 2014).

- **Problem solving and tool usage.** Problem solving processes can be revealed through students' use of virtual tools offered in the app. Virtual tools may include calculators, hints, virtual sketch paper, and other tools. The observable problem-solving behavior may include deciding whether they want to use a tool to help solve the problem (e.g., "I want to use sketch paper" vs. "It's easier to just do it in my head"; see, for example, Wilson, 2002, for a discussion on the off-loading of cognitive work), which tool they choose to use (e.g., "I used the calculator to do the math, but the number I got didn't look right. I think I am going to do it by hand"), and what they do with these tools (e.g., doing long division on the virtual sketch paper vs. creating a visual model for the problem).

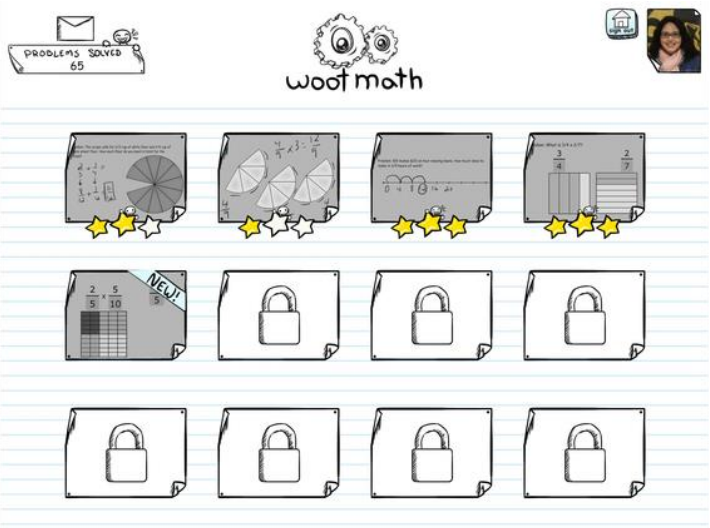


Figure 6. Review board of Woot Math

Interpretation of students' interactions with tools should happen on the item level, but the aggregated measures of such interactions can also provide insights into students' problem solving. For example, Kim et al., (2012) examined students' writing sequence in a tablet based mathematics assessment. They reported correlation between writing pattern across items and students' mathematics score. High-scoring students had fewer erased strokes than their peers. They also displayed more top-down writing movement in their problem solving. Researchers explained that multiple erased strokes at the same location might indicate a trial-and-error approach in students' problem solving, while the top-down writing manner may reveal a systematic planning and problem-solving process.

At the item level, observation of students' problem-solving behavior can provide details about students' specific strategy use as well as their modeling with mathematics. For instance, we conducted usability studies with a prototype tablet assessment for rational number understanding. In this task, students can activate a virtual cutting board where they cut and share banana bread among plates. To solve the following question (Figure 7), two students both activated the cutting board.

Their answers to the multiple-choice question were identical. Their written explanations were also similar, as were the final screenshots when they left the cutting board. Yet they adopted different strategies. While asking for their plan about how to use the cutting board, the first student said,

> "I thought about cut a loaf into 5 pieces. Then I think I might have more pieces but each piece is smaller. So I just cut each into 4 (as I did for the library) and see what happens. Each person get 3, and there's one piece left to share between them. So each one get 3 and a tiny bit more" [sic]

The second student was also prompted about what he planned to do when he activated the cutting board:

> "Four doesn't go into five. So I want to cut each into two and try. It doesn't work. I'll try three. Still not working. I'll try [to cut each into four]. Still not working... Wait now this is the same as the library! But there is a leftover. Someone can take the leftover, they will get more." [sic]

These two strategies were representative of many of the students interviewed. In this case, we were able to map the number of exploratory actions (i.e. how many times the students reset the cutting board, recovered from interviewers' notes and video recordings) with the students' strategy use. In large-scale field testing, however, it is unrealistic to record all the interactions student have with the tablet. It therefore becomes crucial to pre-define events of interest among multiple possibilities in the assessment design and to iteratively adjust these captured events throughout the development and testing process to ensure that they are gathering the evidence that was intended.

To summarize the lesson on interactive items: When designing an interactive item, start with what evidence is needed to make inferences about student performance, and design the interactions to collect the necessary data.

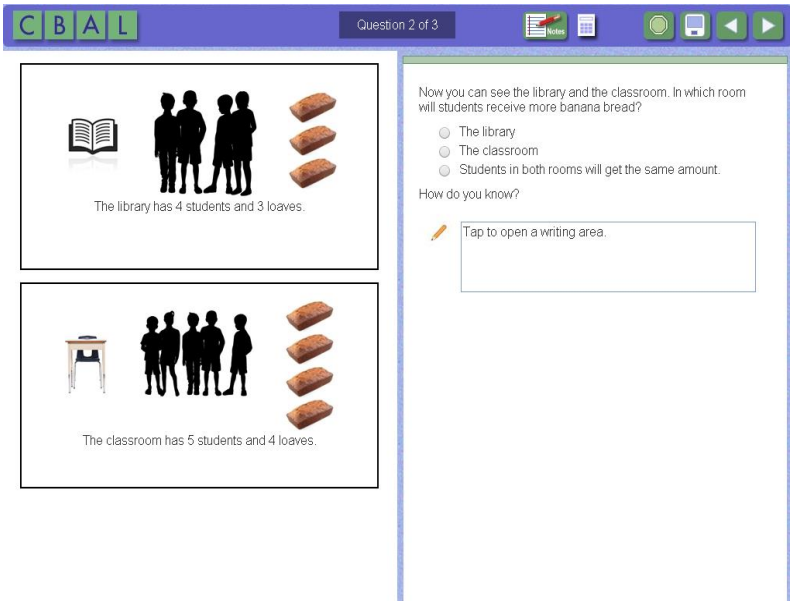


Figure 7. Example item (Cayton-Hodges, Feng, Pan, & Vezzu, 2013)

Developing innovative item types

Our review also suggests areas of weaknesses in the design of popular mathematics apps that will require innovations in designing tablet-based assessments. We focus here on two issues, namely (a) student reflections and explanations, and (b) scaffolding.

- **Self-explanation and reflection.** Literature in cognitive psychology (Chi, De Leeuw, Chiu, & Lavancher, 1994; Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001), education, and assessment (Orsmond, Merry & Reiling, 2002; Orsmond, Merry, & Callaghan, 2004) point to the importance of self-explanation and reflection. Very few of the apps reviewed offer the opportunity for users to reflect and self-explain, however. This may be a byproduct of the game design nature of the apps (i.e., having the user's goals be to gain points or to complete the game as quickly as possible). In education and assessment, where so much can be learned by listening to a student's explanation (such as the example above from bread-cutting), it would be amiss if we did not provide the opportunity for reflection when feasible.

  Prompts for verbal or written explanation can engage students in a process different from their initial reaction. These prompts sometimes act as a re-examination to avoid careless errors, while they sometimes engage a new strategy or modeling. In either case, explanation engages a deeper processing of the information, provides validity evidence for scoring, and can provide teachers enriched interpretation of students' scores. Moreover, the self-explanation process is also a recommended practice for students in promoting conceptual understanding and learning transfer (Rittle-Johnson, 2006; Berthold, Eysink & Renkl, 2009).

  However, prompting explanations for all items requires additional time for assessments, and providing written explanations can be especially tiring for younger students with limited keyboarding skills. These difficulties raise two questions. One is when should we prompt for explanation; the other is how should we collect these explanations. Timing of the explanation prompt can be item based or response based. Item-based prompts should be used in items where multiple steps are needed to solve the problem, as well as when multiple strategies are present in solving a problem. Response-based prompts may be used when students provide an incorrect answer, to differentiate a careless error from more serious misunderstanding. Such prompts may also be used when the student has spent too much time on a question to differentiate between a reading comprehension problem, confusion and lack of knowledge, and poor planning.

  With regards to ways to collect student explanations, tablet devices offer a number of possibilities, such as using the soft keyboard or an external keyboard, handwriting, drawing, and speech input. Our pilot study (Cayton-

Hodges, Feng, Pan, & Vezzu, 2013) showed that fourth-grade students liked writing numerical responses on the tablet by hand, but preferred typing their short-response answers over writing on the tablet (although they also appreciated the ability to supplement their written explanation with drawings, which is typically difficult on a pc with a keyboard-mouse interface). Besides collecting responses through a multi-touch screen, verbal input may also be utilized. For example, Siri is available on iPad as a real-time verbal input processor. It can be called through the assessment interface and return recognized strings in real time.

Reasoning and argumentation are practices called for by the Common Core State Standards (see the Standards for Mathematical Practice, Common Core Standards Initiative, 2010), but these self-explanation skills are often downplayed in practice. There is much room for innovation in this area, and tablets provide a very versatile platform for capturing explanations and reflections, in speech, writing, or drawing. Scoring such evidence is out of the scope of this paper, though a number of natural language processing tools may be used to automatically categorize student explanations. We suggest the following: As part of the design of engaging, interactive tasks, also create opportunities for students to self-reflect or explain their problem-solving process. Consider the use of appropriate modality to capture such evidence.

- **Providing aids and scaffolds.** Tutorials, hints, scaffolds, and other tools are widely used in the mathematics apps we reviewed. They are done in such a way to quickly orient the user to the task and to guide the user along the path through challenges. This type of aid is particularly important for game-like apps, which risk losing users if they cannot jump into the game quickly or get stuck at a level and cannot find a way out. These tools are typically presented on an as-needed basis. Part of the reason is the limited screen real estate – apps need to maximize the area of useful interactions and avoid clustering. Hence, when and how to provide assistance to the user becomes a key consideration in app design.

  In comparison, in traditional assessments the pressure is on students to follow directions and to figure things out. Copying the practice of paper-and-pencil tests, computer-based assessments today typically print instructions statically on the screen, regardless of whether the user needs them. If an assessment task is long and complicated, students often have to read one or more pages of directions. Few apps can afford to do this; they must minimize the effort and working memory load of the user, or else they lose out.

  Extensive reading does take a toll on students. Through our tablet piloting, we found that students are often not very careful while reading the items, and they sometimes get confused about the demands of the item (see Cayton-Hodges, Feng, Pan, & Vezzu, 2013). Because the directions are static, students get no feedback or opportunities to confirm or challenge their understanding. In such cases, students often turn to the interviewer for help. Some students commented that they would be helpless in a summative assessment because they would not have anyone to ask.

  The problem here is how we deliver "directions." In games and apps, as well as in daily communication, helpful directions are given on an as-needed basis. Directions in testing are an anomaly, in that being standardized is a necessity. While this makes sense when item types are simple and familiar, the individualized helpfulness will become increasingly important as tasks become more complex and creative, particularly in formative assessment settings. Letting students struggle in frustration not only goes against the "iPad experience" but also threatens the validity of the assessment; a score is not a proper reflection of mathematical ability if the student is confused about task requirements. If we decide to engage students in complex assessment tasks, then it is our responsibility to ensure that all students understand and can navigate the tasks. The mathematics apps we reviewed provide many positive examples to guide students through challenges, through animated tutorials, adaptive instructions, and options to request help.

  Help-seeking behaviors can also provide valuable assessment information. In a formative assessment, the teacher may want to know which students clicked on the "help" button and under what circumstances. Having context-dependent assistance in this case creates assessment opportunities. Imagine an interactive instruction (or a computer agent) that can read, rephrase, or explain parts of the directions, depending on users' needs. This would be a helpful tool for English language learners and students with special needs, and would enhance the validity of the assessment as a result. Finally, more sophisticated aids may provide students with learning resources, from "cheat sheets" to instructional videos (see *Math* by YourTeacher.com). This creates opportunities to assess not only what students know (and do not know) but also how much and how quickly they can learn.

While not every method of aid is appropriate for all assessments, we recommend the following: Assessment developers need to adopt the mindset of app developers: it is the designers' responsibility to keep the user engaged, on task, and moving forward. A test is only valid to the extent that students are "in the game" However, take caution that the construct of measure is never sacrificed in the name of engagement.

Summary

The aim of this report was to summarize a survey of mathematics education apps in the Apple App Store, conducted as part of a research project to develop a tablet-based assessment prototype for elementary mathematics. This survey was performed with the goal of understanding the design principles and techniques used in mathematics apps designed for tablets. We focused our reviews on four areas, (1) the quality of mathematical content, (2) feedback and scaffolding, (3) richness of interactions, and (4) adaptability of the applications. These four areas were cultivated from prior research on digital tools in mathematics (e.g., Digital Tools for Algebra Education criteria; Bokhove & Drijvers, 2011), designing principles of learning objects (e.g., Learning Object Evaluation Metric; Kay & Knaack, 2008), as well as quality of mathematics instruction (e.g., Hill et al., 2008). This review culminates in the formulation of four recommendations for researchers and assessment developers on designing tablet-based mathematics assessments: (1) Thoroughly review the mapping between concepts/operations and objects/actions early in the task design stage; (2) Start with what evidence is needed to make inferences about student performance, and design the interactions to collect the necessary data; (3) Create opportunities for students to self-reflect or explain their problem-solving process; and (4) Adopt the mindset of app developers to keep the user engaged, on task, and moving forward to ensure that students are "in the game" enough to accurately assess content knowledge.

References

Abrams, R. A., Davoli, C. C., Du, F., Knapp III, W. H., & Paull, D. (2008). Altered vision near the hands. *Cognition, 107*, 1035–1047.

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement*, *8*(2–3), 70–91.

Berthold, K., Eysink, T. H., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*(4), 345–363.

Blume, H. (2013). L.A. school board OKs $30 million for Apple iPads. *Los Angeles Times*. Retrieved from http://articles.latimes.com/2013/jun/18/local/la-me-ln-lausd-chooses-ipads-for-pilot-20130618

Bokhove, C., & Drijvers, P. (2010). Digital tools for algebra education: Criteria and evaluation. *International Journal of Computers for Mathematical Learning*, *15*(1), 45–62.

Bokhove, C., & Drijvers, P. (2011). *Effects of feedback conditions for an online algebra tool*. In M. Joubert, A. Clark-Wilson, & M. McCabe (Eds.), Proceedings from the Tenth International Conference for Technology in Mathematics Teaching (ICTM T10) (pp. 81-86). Retrieved from http://mccabeme.myweb.port.ac.uk/ictmt10proceedings2.pdf

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4–5), 514–527.

Cayton-Hodges, G.A., Feng, G., Pan, X., & Vezzu, M. (2013). *iPad task design for elementary math: End of year report to the CBAL initiative.* Princeton, NJ: Educational Testing Service.

Chan, M. S., & Black, J. B. (2006, April). *Learning Newtonian mechanics with an animation game: The role of presentation format on mental model acquisition*. Paper presented at the American Education Research Association Annual Conference, San Francisco, CA.

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477.

Chi, M. T. ., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, *25*(4), 471–533.

Common Core Standards Initiative. (2010). *Common core standards initiative*. Retrieved from http://www.corestandards.org/Math/Practice

Davoli, C. C., & Brockmole, J. R. (2012). The hands shield attention from visual interference. *Attention, Perception, and Psychophysics*, 74(7), 1386–1390.

Davoli, C. C., Du, F., Montana, J., Garverick, J., & Abrams, R. A. (2010). When meaning matters, look but don't touch: The effect of posture on reading. *Memory & Cognition*, 38(5), 555–562.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3–31.

Gorin, J. S. (2006). Test design with cognition in mind. *Educational Measurement*: *Issues and Practice*, *25*(4), 21–35.

Haughey, M., & Muirhead, B. (2005). Evaluating learning objects for schools. *E-Journal of Instructional Sciences and Technology*, *8*(1), 229–254.

Harrison, T. R. (2013). The evaluation of iPad applications for the learning of mathematics. Retrieved from http://www.lib.ncsu.edu/resolver/1840.16/8707

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Hieb, J. L., & Ralston, P. A. S. (2010). Tablet PCs in engineering mathematics courses at the J.B. Speed School of Engineering. *International Journal of Mathematical Education in Science and Technology*, *41*(4), 487–500.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430–511.

IDC (2013). Tablet shipments forecast to top total PC shipments in the fourth quarter of 2013 and annually by 2015, according to IDC [Press Release]. Retrieved from http://www.idc.com/getdoc.jsp?containerId=prUS24314413

Interactive Education Systems Design, Inc. (2013). National survey on mobile technology for K-12 education. Retrieved from http://dnwssx4l7gl7s.cloudfront.net/amplifylive/default/page/-/Amplify_Mobile_Technology_Research_Report.pdf

Isabwe, G. M. N. (2012, June). Investigating the usability of iPad mobile tablet in formative assessment of a mathematics course. In *2012 International Conference on Information Society (i-Society)* (pp. 39–44). Paper presented at the 2012 International Conference on Information Society (i-Society), London, UK.

Kay, R., & Knaack, L. (2008). A multi-component model for assessing learning objects: The learning object evaluation metric (LOEM). *Australasian Journal of Educational Technology*, *24*(5), 574–591.

Kim, Y., Lim, C., Choi, H., & Hahn, M. (2012, August). *Effects on training mathematics problem-solving behaviors using a tablet computer.* Paper presented at the 2012 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), Hong Kong.

Kowalski, S. E., Kowalski, F. V., & Gardner, T. Q. (2009). Lessons learned when gathering real-time formative assessment in the university classroom using Tablet PCs. *Proceedings of the 39th IEEE Frontiers in Education Conference* (pp. 926-930). San Antonio, TX. Piscataway, NJ: IEEE Press. doi: 10.1109/FIE.2009.5350639

Laughlin Davis, L., Strain-Seymour, E., Gay, H. (2013). Testing on tablets: Part II of a series of usability studies on the use of tablets for K-12 assessment programs (Pearson white paper). Retrieved December 1, 2013, from http://researchnetwork.pearson.com/wp-content/uploads/Testing-on-Tablets-Part-II_formatted.pdf

Mainelli, T. (2013, September). U.S. consumer tablet survey results, part 1: Hardware, OSs, brands, and replacement cycles. IDC. Retrieved from http://www.idc.com/getdoc.jsp?containerId=243327

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger III, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*(2), 399–414.

McKnight, L., & Fitton, D. (2010). Touch-screen technology for children: giving the right instructions and getting the right responses. *Proceedings of the 9th International Conference on Interaction Design and Children* (p. 238-241). New York, NY: ACM Press. doi: 10.1145/1810543.1810580

McMorris, R. F., Schwarz, S. P., Richichi, R. V., Fisher, M., Buczek, N. M., Chevalier, L., Meland, K. A. (1991). *Why do young students change answers on tests?* Research Report to the State University of New York at Albany. (ERIC Document Reproduction Service, ED 342 803).

Moyer-Packenham, P.S., Salkind, G., & Bolyard, J.J. (2008). Virtual manipulatives used by K-8 teachers for mathematics instruction: Considering mathematical, cognitive, and pedagogical fidelity. *Contemporary Issues in Technology and Teacher Education*, *8*(3), 202–218.

Nicol, D. J., & Macfarlane- Dick, D. (2006). Formative assessment and self- regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218.

Orsmond, P., Merry, S., & Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, *41*(3), 273–290.

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, *27*(4), 309–323.

Pan, X. Cayton-Hodges, G.A., and Feng, G. (2014, April). *Not all changes are created equal: Reasons, contexts, and outcomes of answer change.* Paper presented at the American Educational Research Association Annual Conference, Philadelphia, PA.

Ren, X., & Moriya, S. (2000). Improving selection performance on pen-based systems: A study of pen-based interaction for selection tasks. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *7*(3), 384–416.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*(1), 1–15.

Roediger III, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*(4), 382–395.

Strain–Seymour, E., Craft, J., Davis, L. L., & Elbom, J. (2013). *Testing on tablets: Part I of a series of usability studies on the use of tablets for K-12 assessment programs* (Pearson white paper). Retrieved December 1, 2013, from http://researchnetwork.pearson.com/wp-content/uploads/Testing-on-Tablets-PartI.pdf

Smarter Balanced Assessment Consortium. (2012, April 25). Smarter balanced and PARCC issue guidance for new instructional technology purchases. Retrieved from http://www.smarterbalanced.org/news/smarter-balanced-and-parcc-issue-guidance-for-new-instructional-technology-purchases/

Toby, M., Ma, B., Lai, G., Lin, L., & Jaciw, A. (2012). *Comparative effectiveness of Houghton Mifflin Harcourt Fuse: Algebra 1– A report of randomized experiments in four California districts*. Palo Alto, CA: Empirical Education Inc.

Tsuei, M., Chou, H.-Y., & Chen, B.-S. (2013). Measuring usability of the Mobile Mathematics curriculum-based measurement application with children. In A. Marcus (Ed.), *Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience* (pp. 304–310). doi: 10.1007/978-3-642-39241-2_34

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.

Appendix

List of Apps reviewed

Fluidity Software, Inc. (2013). FluidMath 2013. (Version 1.2). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/fluidmath–2013/id594744681?mt=8

Hands on Equations. (2013). Hands–on Equation (Version 3.0) [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/hands–on–equations–1–fun–way/id505948222?mt=8

Helttula, E. (2013). Long Division (version 2.8). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/long–division/id341041204?mt=8

Houghton Mifflin Harcourt. (2013). HMH Fuse: Algebra 1 (version 2.1). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/hmh–fuse–algebra–1/id415533582?mt=8

INKids. (2013).Math Champ Challenge (version 1.1). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/math–champ–challenge–common/id566463831?mt=8

Interactive Elementary. (2012). Middle School Math HD (version 2.6). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/middle–school–math–hd/id439682332?mt=8

Khan Academy (2013). Khan Academy (Version 1.3.2). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/khan–academy/id469863705?mt=8

Martinez, J. (2011). touchyMath (Version 1.2.7). Mobile application software]. Retrieved from https://itunes.apple.com/us/app/touchymath/id388884486?mt=8

MobileRise. (2013). Map Ruler (version 1.3) [Mobile application software]. Retrieved from https://play.google.com/store/apps/details?id=mobilerise.MapsRuler&hl=en

Motion Math. (2013). Motion Math HD: Fractions! [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/motion–math–hd–fractions!/id410521340?mt=8

Nimbee LLC. (2013). Woot Math. [Mobile application software]. Retrieved from http://wootmath.com/

Nuance Communications (2013). Dragon Dictation (version 3.0.28). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/dragon–dictation/id341446764?mt=8

Play Power Labs. (2013). Fraction Planet (beta). [Mobile application software]. Retrieved from http://www.fractionplanet.com/

Scolab (2013). BuzzMath Middle School (version 1.3.1). [Mobile application software]. Retrived from https://itunes.apple.com/app/buzzmath–middle–school/id593186620.

Teachley. (2013). Teachley: Addimal Adventure (version 1.2). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/teachley–addimal–adventure/id661286973?mt=8

Thoughtbox. (2013). Numerosity: Play with multiplication (Version 1.0). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/numerosity–play–multiplication!/id670585230?mt=8

West, B. (2013). Pizza Fractions: Beginning With Simple Fractions (Version 1.5). [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/pizza–fractions–beginning/id374084320?mt=8

WeWantToKnow AS. (2013). DragonBox Algebra (Version 1.1.6) [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/dragonbox–algebra–5+/id522069155?mt=8