

# Touch Arithmetic: A process-based Computer-Aided Assessment approach for capture of problem solving steps in the context of elementary mathematics

Adewale Adesina<sup>a</sup>, Roger Stone<sup>a</sup>, Firat Batmaz<sup>a</sup> and Ian Jones<sup>b</sup>

<sup>a</sup> Department of Computer Science, Loughborough University, Loughborough, U.K.

<sup>b</sup> Mathematics Education Centre, Loughborough University, Loughborough, U.K.

{a.o.adesina, r.g.stone, f.batmaz, i.jones}@lboro.ac.uk

---

## Abstract

Technology today offers many new opportunities for innovation in educational assessment and feedback through rich assessment tasks, efficient scoring and reporting. However many Computer-Aided Assessment (CAA) environments focus on grading and providing feedback on the final product of assessment tasks rather than the process of problem solving. Focusing on steps and problem-solving processes can help teachers to diagnose strengths and weaknesses, discover strategies, and to provide appropriate feedback. This study explores a method that uses trace links on an interactive touch-based computing tool for the capture and analysis of solution steps in elementary mathematics. The tool was evaluated in an observational study among 8 and 9 year old primary school children (N=39). The approach yielded similar performance scores as compared to paper-and-pencil tests while providing more explicit information on the problem-solving process. The output data was useful for scoring intermediate and final answers as well as feedback information on types and time efficiencies of strategies used. An implication of this study for teachers and researchers is that they can more accurately assess students' understanding of important concepts, and be in a better position to provide rich and detailed feedback while motivating students with interactive tasks.

---

## Keywords:

Elementary education, Human-computer interface, Interactive learning environments, Computer-Aided Assessment

## 1. Introduction

Assessment and feedback is one of the most powerful influences on learning and achievement (Hattie, 2008; Black & Wiliam, 1998). Through assessment teachers gather data about their teaching and their students' learning (Hanna & Dettmer, 2004). Computers and electronic technology today offer numerous ways to enrich educational assessment and feedback both in the classroom and large-scale assessment situations. The practical benefits of Computer-Aided Assessment (CAA) include automatic scoring, rapid feedback, and increased accessibility (Conole & Warburton, 2005). Interactive computer based tasks are also engaging through the immediate appeal of their graphics and the sustained appeal of their interactivity (Richardson et al., 2002).

Although extensive academic research has explored CAA systems and their benefits to teaching and learning (Conole & Warburton, 2005), the focus of many CAA systems has been the final answer or product of assessment tasks. Multiple Choice Questions (MCQ's) and selected responses item-type tasks appear to be dominant. This is mainly because such tasks are readily scored by a variety of electronic means. This approach sometimes limits creative problem solving as students often have to choose an answer in a limited range of options. Some researchers have argued that such implementations only test surface learning (Ward & Bennett, 2012; Hommel, Colzato, Fischer, & Christoffels, 2011). Considering the development process leading to the answer enables better understanding of the rationale behind the product (Baker & Mayer, 1999). A process has been defined as a systematic series of actions directed to achieve a result (Pinheiro & Goguen, 1996). Process oriented systems aim at telling the story of a problem-solving effort. Many studies on process-based CAA systems have focused on problem-solving behaviour (Chung & Baker, 2003; Williamson et al., 2004; Zoanetti, 2010). Investigations have centred on how, and the sequence problem-solvers completed task. However, little attention has been paid to making the problem-solving process explicit by showing how different parts of a solution effort relate with each other. A major purpose of using computer for assessment that is relevant to the research reported here is to increase efficiency without diminishing the validity or credibility of results (Baker & Mayer, 1999). The introduction of new technology often leads to concerns about the extent to which students' scores differ between computer-based and paper-based tests (McDonald, 2002). Against this

background, the purpose of this paper is to contribute to the present understanding of how assessment environments may be designed to capture problem solving process data to better inform assessment and feedback practices. The present study explores a trace-based approach for capture of the essential details of a problem-solving process and the relationships between the various parts of the solution. A trace is a path or history showing how some particular state came to be (Pineiro & Goguen, 1996). The study is divided into two parts; the first part describes a method for capturing solution steps from traces of interactions produced by students on a touch-based prototype tool while solving arithmetic word problems. The second part of the study focuses on the evaluation of the tool and considers concerns of usability and also the applicability of providing feedback on arithmetic strategies. Specifically the evaluation answers the following questions:

- Does the tool impede students' accuracy and efficiency when solving mathematical word-problems?
- Can the use of process data output lead to the detection of arithmetic strategies?

The rest of this paper is structured as follows: First, a review of literature on related studies is provided. This is followed by discussions on the research approach, the design and implementation of this approach on the prototype tool and an observational study evaluating the tool. Following this, the results from the study are presented and discussed. Finally, implications and directions for future research are offered.

## 2. Review of Literature

### 2.1 Process and product assessments

Assessment has been defined as activities that provide teachers and student with feedback information and interpretations about the difference between the current status and learning goals (Hattie & Timperley, 2007). Black and William (1998) showed that improvement in classroom assessment has a strong effect on students' achievements. In their article Hattie & Timperley (2007) argued that feedback on assessment activities can be in four categories: tasks, processes and strategies, self-regulation and the person. The study argued that process-level feedback was most effective in assessment situations. For instance while task-level feedback is concerned about whether the task or product is correct or not, process-level feedback looks at the procedure used to create the product or complete the tasks. It addresses the questions like what is wrong and why? What strategies did the learner use? What is the explanation for the correct answer? What relationship exists with other parts of the task? What is the learner's understanding of the concepts related to the task? To answer these questions the paths to the solution need to be clearly visible (Baker & Mayer, 1999).

### 2.2 An Example

Word or story problems are commonly used in schools to train and test understanding of underlying concepts within a descriptive problem as well as to test student's capability to perform arithmetic manipulations (Hegarty, Mayer, & Monk, 1995). Figure 1 shows a two-step arithmetic word problem that involves summing double digits and single digit numbers i.e. 34, 18, and 6 on paper. The question is a two-step arithmetic problem which requires an intermediate result and a final answer. Four student solutions are shown in the diagram.

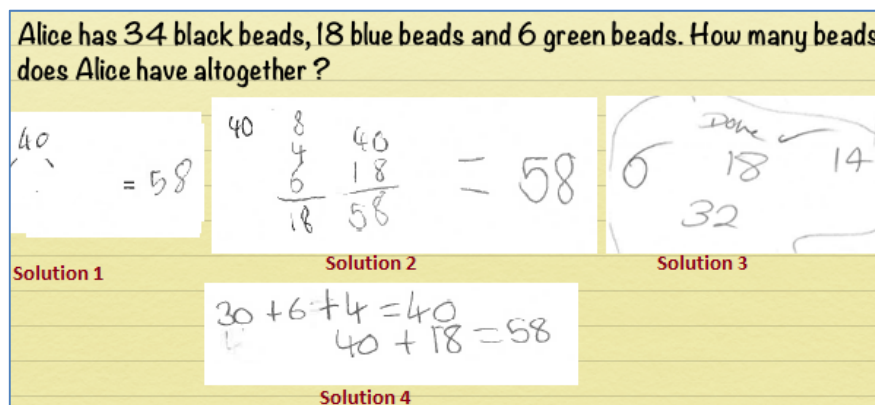


Fig 1: Record of solutions

Solution 1 shows a final answer of 58 without showing explicitly how the answer was arrived at. Solutions 3 and 4 show the workings using vertical and horizontal arrangement for the addition. Solution 3 is less clear; the number 14 is not in the problem text but appears in the solution thereby making the student's thinking less obvious. Too often, the opportunities to provide meaningful feedback are missed when student's responses to these types of assessment tasks are not detailed or clear.

### 2.3 Conceptual understanding and arithmetic strategies

In many fields of study students are taught important concepts and correct procedures. Conceptual knowledge has been defined as explicit or implicit understanding of the principles that govern a domain and the interrelations between them while procedural knowledge is seen as the action sequences for solving problems (Rittle-Johnson & Alibali, 1999). Rittle-Johnson & Alibali (1999) argued that mathematical ability lies in students developing and connecting their knowledge of concepts with procedures. The association between conceptual knowledge and procedural knowledge has been identified in many studies in the mathematical domain for instance counting (Cowan & Renton, 1996), single-digit arithmetic (Baroody & Gannon, 1984; Cowan & Renton, 1996), fraction arithmetic (Byrnes & Wasik, 1991), and proportional reasoning (Dixon & Moore, 1996). Overall, the literature suggests that conceptual understanding plays an important role in strategy adoption and generation. Imbo (2008) described three classes of strategies that can be used to solve mental arithmetic problems. These are; direct memory retrieval, procedural strategies such as counting (e.g.,  $8+5=9\dots10\dots11\dots12\dots13$ ;  $9 \times 6 = 9\dots18\dots27\dots36\dots45\dots54$ ) and transformational strategies (e.g.,  $8 + 5 = 8 + 2 + 3$ ;  $9 \times 6 = 10 \times 6 - 6$ ).

When students are given tasks to solve it is often valuable to know what strategy was applied to solve the problem. For example, a concept usually taught in maths classes is number bonds. Number bonds refer to useful pairs or numbers that bond to 10 or multiples of 10, they help students see that numbers can be transformed or "broken" into pieces to make computation easier and to recognize relationships. Teachers may want to know if students understand and use this concept in the context of a word or story problem. To accurately assess the strategy so as to give meaningful feedback, they need to have access to the detailed steps and final answer. Performance on such problem-solving tasks is dependent on both strategy selection and strategy efficiency (Mayer & Wittrock, 1996). According to the definitions provided by Imbo (2008), strategy selection is the choice of a strategy among several alternatives while strategy efficiency refers to how fast and how accurately a strategy leads to the solution.

Throughout the remainder of this paper the term *place-value* strategy will be used to refer to a student solution that reflects understanding of number bonds (i.e. the student recognises and uses the number bonds concept). *As-presented* strategy will refer to solutions that pair numbers based on the order they occur in the problem only. *Other* will simply refer to any strategy outside the aforementioned. Getting information about these types of strategy requires techniques and tools that reveal the problem-solving process.

### 2.4 Computers and process assessment

Computer environments have long been recognized as being valuable for process-based, problem-solving tasks. Video-recorded observations used with think-aloud procedures have been used for gathering information about student's thinking and the way they solve tasks (Bannert & Mengelkamp, 2008; Young, 1995). Also, audio and screen recordings have been used to follow precisely students' movements on the computer (Barmby, Harries, Higgins, & Suggate, 2009). While the studies showed that students' strategies can be captured in a more in-depth way, these methods are often expensive and difficult to implement in classroom situations. The use of interactions in a computing environment for elementary mathematics has been found to be beneficial. For example, Silke (2011) showed that children externalize their concept of numbers through touching a multi-touch screen with their fingers and thus producing tokens. Jones and Pratt (2012) used a specially-designed microworld to investigate primary students' understanding of arithmetic symbol structure. They found that students interactions with the software more closely resembled typical algebraic than arithmetic strategies. In particular, students tended to look for equivalence relationships between arithmetic equations rather than calculating arithmetic results.

The use of software that produces a list of events a student has carried out has also been studied (Baker & Mayer, 1999). The date, duration and actions such as clicks on the screen are registered in a log file and examined to detect what the student is trying to do. This method has the advantage of being non-intrusive and can be useful in assessing a solution process. However, fewer studies have explored dependencies and relationships in the problem solving process in ways they can be unambiguously represented. This research argues that solution process examination can be extended to include connections and relationships. Showing how problem specification, solution steps and final answers are related is likely to increase precision and clarity and may lead to better interpretation of the results and strategies.

### 3. Methodology

The aim of the study is to capture students' detailed solution steps in elementary mathematics for process-level feedback information. Effective solution steps capture will require a computer user interface that provides problem-solvers opportunities to explicitly enter their solutions. The approach of capture adopted for the present study uses interactive problem-solving tasks which aim at reducing inconsistencies and ambiguities in the representation of solution steps. The first part of this section discusses this method for capture. The second part presents the details of an observational study that evaluates the tool and its applicability in investigating arithmetic strategies.

#### 3.1 The method of capture and tool implementation

We propose the use of structured trace information for the capture of problem-solving steps to reduce the problem of unclear steps and ambiguous relationships between the different parts of a solution effort as described in Section 2.2. This method is similar to the concept of *on-line requirement traceability* described in software engineering and design rationale studies (Dick, 2002; Rochimah, Wan Kadir, & Abdullah, 2011; Tang, Jin, & Han, 2007). Traceability allows for tracing the chain of reasoning why particular choices were made by establishing links between paragraphs of documents (Von Knethen, Antje, & Barbara Paech, 2002).

#### 3.2 Trace capture: The basic idea

On-line trace capture is based on the idea that information can be recorded during process execution by providing a set of user interactions by which the product is created, deleted, and/or modified. The actions are responsible for recording the trace information in a repository. As the actions are performed by the problem-solver dependency or relationship links are created between the objects involved in the action. This enables precise recording of solution items and dependencies that are caused by the execution of nested actions. Recent studies have shown that trace capture can be enhanced by integrating the problem text/scenario with the solution space such that interactive actions are executed between them (Batmaz, Stone, & Hinde, 2009; Stone, Batmaz, & Rickards, 2010).

Taking the word problem example used in section 2.2 it is possible to capture the solution process through interactive actions by using an integrated problem text and solution workspace. This is illustrated in Figure 2.

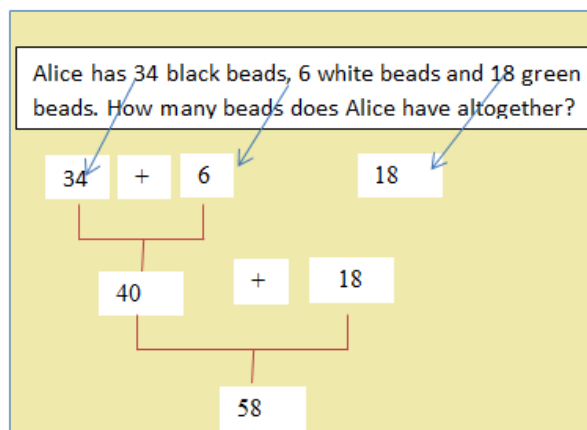


Fig 2. Problem space (top) and solution space (bottom) integration with trace links

Here a student first determines the numbers to use for the solution and takes them to the solution area. Once this is done the student decides on two of the numbers to first work with (e.g. 34 and 6), and chooses an operator to use on them. The student then computes and enters the result concluding the first stage with an intermediate result (40). Following this, the intermediate result is paired with the remaining number from the question (18) after another decision on the correct operator. A final answer (58) is recorded, concluding the second and final stage. All this information when captured can be analysed to provide a full picture of how the final solution was arrived at. This simple example illustrates important concepts of the research approach. Solving the question by interacting with the numbers from the problem text allows for direct linking of solution items with their origins. This provides the opportunity to describe and follow the life of an answer or response to a question in both forward and backward directions (i.e. from its origins, through any subsequent use and through all periods of refinement or modification). This tracing approach may help to verify solution items against problem specifications, and more easily identify error sources. It also makes it possible to identify relationships among the components of a solution.

### 3.3 Tool Design and Implementation

A prototype tool based on the research approach outlined above was developed. The tool will be referred to as the *Multi-Touch Arithmetic Tool* (MuTAT). Although it is possible to implement the approach on several technologies, the multi-touch technology was chosen for two main reasons. First, multi-touch can be an intuitive and a more expressive way of using touch interactions with computers (Moscovich, 2007), reflecting how children explore physical objects with their hands. Studies have shown the multi-touch environment removes abstraction from the interaction process (Moscovich, 2007), is engaging (Heinrich, 2011), permits performance of complex tasks in a reduced time (Jiao, Deng, & Wang, 2010), supports bi-manual interactions (Bailey, Müller, & Lecolinet, 2012) and aids collaboration (Mercier & Higgins, 2013).

The second reason for this choice is the popularity of multi-touch devices. In recent years, the technology has become widespread with tablet PCs and other multi-touch surfaces increasingly used in different teaching and learning scenarios (comScore, 2011). Analysts have predicted that the market for tablet computing will soon eclipse the market for desktop computing including laptops (UPI.com, 2012). As the digital landscape advances, it would seem appropriate that assessment and feedback tools should leverage on these innovations.

### 3.4 MuTAT Layout for problem-solving

The layout and design of the MuTAT application is presented in Figure 3. It is divided into two; the problem pane and the solution pane as discussed in Section 3.1.1. The problem pane displays specific and relevant word problem tasks suitable for the level of the student while the solution pane displays responses constructed by the problem solver. Both panes are placed together on the workspace to allow for easy movement between panes.

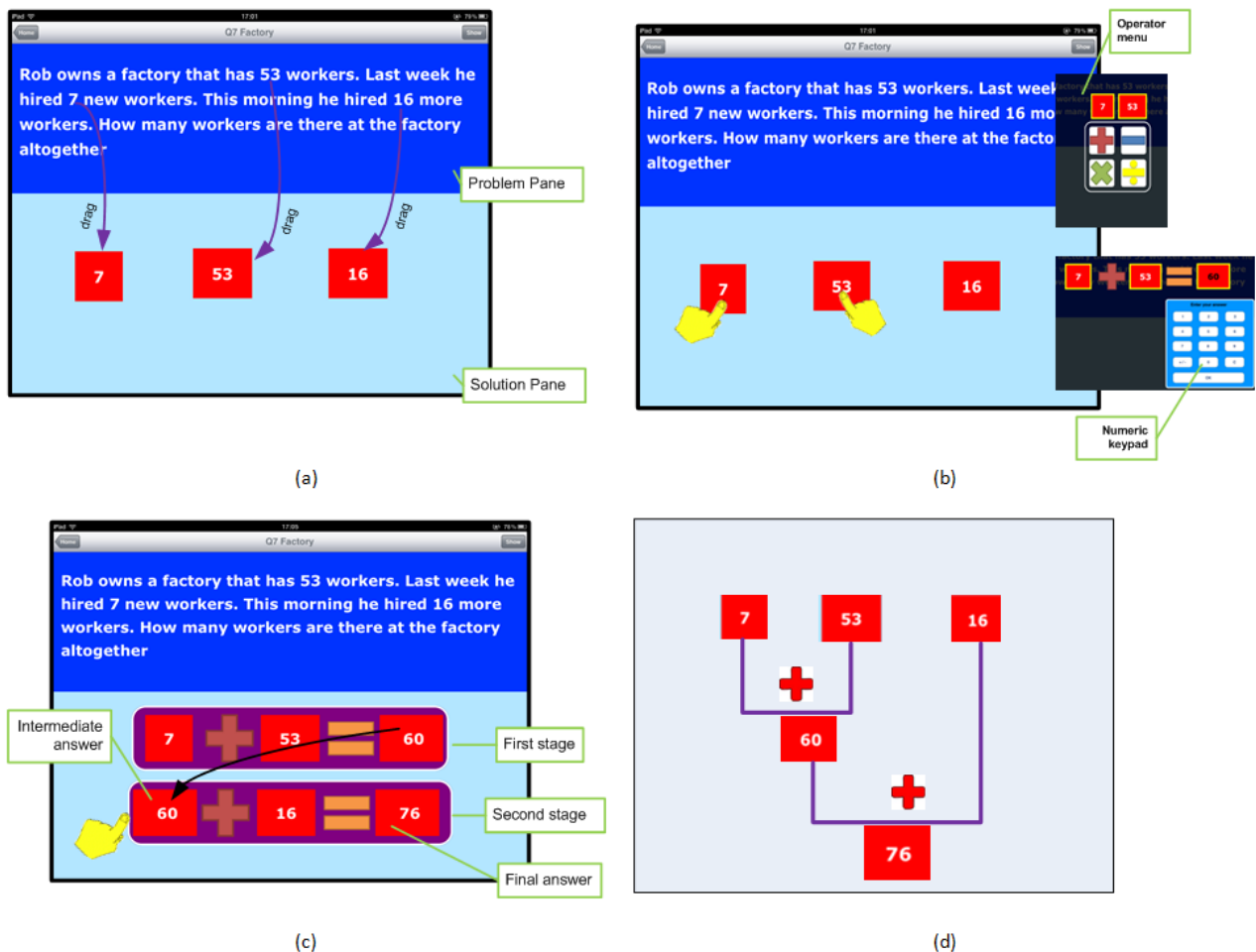


Fig 3. MuTAT user interface

The problem pane contains word problems with draggable numeric values. One or more numbers can be dragged from the problem pane down to the solution pane with one or both hands (Figure 3a). The pan (drag) gesture can be

used for this step. To enter an arithmetic computation a multi-touch gesture is used. A user touches and holds two chosen numbers simultaneously (Figure 3b). The action of touching and holding an interface element for a few seconds is referred to in literature as touch-and-hold gesture or long press gesture (Bailly et al., 2012; Wigdor & Wixon, 2011). This action in the environment activates a menu containing four arithmetic symbols (+ - × ÷) for selection. The selection of one of the symbols activates a numeric key pad and a text box for receiving and displaying entered results. In multi-stage arithmetic problems the product of the first step is used in subsequent steps (see Figure 3c). The process is repeated until a final answer is arrived at. The record of interactions and solution are captured and stored in a repository from which visual representations of the process data is constructed. Figure 4d shows a pictorial representation of the solution generated from traces of the problem-solving actions.

## 4. The Study

The design discussed above was implemented on a tablet computer for the evaluation study. The study examines usability of the tool and evaluates students' performances with reference to equivalent paper-and-pencil tests. Furthermore, it examines the applicability of using the tool to investigate arithmetic strategies.

### 4.1 Participants

The participants were 10-11 year-old children (24 girls and 15 boys) attending Year 5 classes in two rural schools in England. The schools were attended by children of mixed socio-economic background. The Year 5 children are expected to have knowledge of elementary multi-digit addition and word problems (Borthwick & Harcourt-Heath M, 2012). The typical school year runs from September to July. The study was carried out during the last month of the school year. The experiment was divided into two parts: the first involved the participants solving word-problems on the MuTAT application while the second part involved solving similar problems on paper.

### 4.2 Experimental design and procedure

In the first part of the study, the participants were required to solve word problems using the MuTAT application. Well-researched word-problems specially designed to allow detection of strategies were chosen. The problems contained numbers that were chosen to support the use of the place-value strategy by students as discussed in Section 2.3. In each problem there are two large two-digit numbers, and one single-digit number. Two of the numbers bond to multiples of 10 as shown in questions 3, 6 and 7 in Table 1. In each problem the two numbers that bond to multiples of 10 are presented in different positions: 2<sup>nd</sup> and 3<sup>rd</sup> in question 1; 3<sup>rd</sup> and 1<sup>st</sup> in question 6; 1<sup>st</sup> and 2<sup>nd</sup> in question 7. The particular values were selected so that adding the two-digit numbers requires a carry over. The large numbers were selected such that the computational burden is minimised by starting with the large number, and then one of the smaller numbers or vice versa. Question 7 presents the numbers in strategic order, with the large and the small numbers placed together. This was a control question which served to help us work out if any participants consistently either (i) just chose numbers from left to right or (ii) just choose numbers arbitrarily.

**Table 1**

A selection of study questions

| Questions |                                                                                                                                                                                                      |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3         | Joe had <b>17</b> toy cars. His father gave him <b>26</b> toy cars. His mother gave him <b>4</b> more toy cars. How many toy cars did Joe have altogether? <sup>a</sup>                              |
| 6         | Sara has <b>8</b> sugar donuts. She also has <b>15</b> plain donuts and <b>32</b> jam donuts. How many donuts does Sara have altogether? <sup>a</sup>                                                |
| 7         | Jason owned a factory that employs <b>53</b> workers. He hired another <b>7</b> workers. He then hired another <b>16</b> workers. How many workers are there at the factory altogether? <sup>a</sup> |
| 10        | There are <b>22</b> sheep in the field, <b>13</b> more are put there and some moved away. At the end there are <b>28</b> sheep in the field. How many were moved away? <sup>b</sup>                  |

Adapted from (Carpenter & Moser, 1984) <sup>a</sup> and (Gilmore & Bryant, 2006) <sup>b</sup>

Drawing on Gray and Tall (1995) and Gilmore and Bryant (2006), we hypothesized that some students who understand arithmetic conceptually would consistently use the place-value strategy in the questions to ease the

computational burden. Students with a less secure conceptual understanding would only make partial use of the place-value strategy. Students who do not view arithmetic conceptually would consistently go from left to right, or select numbers arbitrarily in no particular order.

The first two questions the students attempted were practice questions not included in the analysis. The next five questions were similar to Questions 3 and 6 and 7. In addition, there were 5 exploratory questions similar to question 10 above. The exploratory questions required the students to make use of addition and subtraction operations. They are distractors, and served to help us check if the students will understand the requirement in the problem text to use a mix of plus and minus operators.

In the second part of the study the students worked on paper-and-pencil test. The paper tests were a research instrument to check the performance scores of the students in similar tasks. In the test word problems were not used, instead the same question structure was presented symbolically (e.g.  $36 + 4 + 18$ ).

### 4.3 Measurements

The variables recorded by the tool included question number, type of gesture interaction (i.e. drag, long-press, and touch), X and Y coordinates of each element on the screen to obtain data about the physical and logical locations of the solution items. The time-stamp for each operation was also recorded to establish the sequence of events. For the paper-and-pencil tasks, the workings and answers were collected for comparative analysis with data collected with the MuTAT application.

A total of 546 questions were answered by the 39 participants (390 on the MuTAT application and 156 on paper). All the responses were recorded in a log file. The file contained a total of 8,221 records.

## 5. Results

The results of the experiment are presented in both qualitative and quantitative forms in this section. The analysis of the responses on the MuTAT and paper are first examined. This is followed by an analysis of the strategies used by the students.

### 5.1 Performance scores

On the MuTAT application all the participants were able to complete all word-problems. Also, all the students made the correct choice of arithmetic operation (addition) in answering the study questions. Most of the students used a mix of the plus and minus operators for the distractor questions. This suggests that the students were able to use the tool to enter their desired solutions and understood the semantic relations in the problem text. This claim was also confirmed by observations made on the participants while working on the experiments. The participants also successfully completed the pencil and paper questions.

Figure 4 compares the results obtained from the MuTAT application and paper-and-pencil task in a bar plot with standard error as error bars

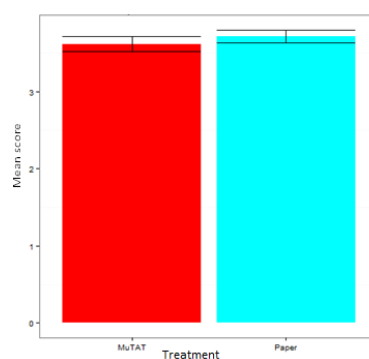


Fig 4. Bar plot with standard error as error bars of mean scores

This figure showed that the mean performances on the MuTAT software and paper are comparable (MuTAT-> M = 3.61, SD=0.59; Paper ->M = 3.72, SD = 0.51). Both treatments had a mode score of 4. However the performance on the MuTAT application appeared to have a greater variance. To test for significant differences between the two methods, a paired t test suggested that there were no significant differences between the means of the two groups ( $t(38) = 0.94, p = 0.35$ ).

### 5.2 Time on task

An analysis of task time can indicate problems in using a tool or method. Several scholars have argued that task times are very useful in diagnosing usability problems such poor interactions and inefficient procedures (Bailey, 1993; Lewis, 2006). Table 2 provides univariate descriptive statistics for the times the participants spent on completing the tasks on the MuTAT platform. It can be seen that the average time spent on all the questions was 38 seconds.

**Table 2**

Univariate descriptive statistics for the task times (n = 39)

|                           | Minimum | Maximum | M     | SD   |
|---------------------------|---------|---------|-------|------|
| Completion time (seconds) | 19      | 92      | 38.46 | 0.25 |

The mean score of 38 seconds suggests that the participants did not spend excessive amount of time in using this approach. Because the current research was not specifically designed to measure the task time for the paper-and-pencil test a comparative task time analysis could not be made between the two methods. However observations made during the experiments suggested no notable difference in the task times.

### 5.3 Detailed feedback

As mentioned in Section 3, traditional paper-and-pencil tests can capture detailed solution steps but the absence of consistent notation for representing solution steps and the difficulty in enforcement may present some challenges. Figure 5 shows the responses from three students Lucas, Julia and Mark.

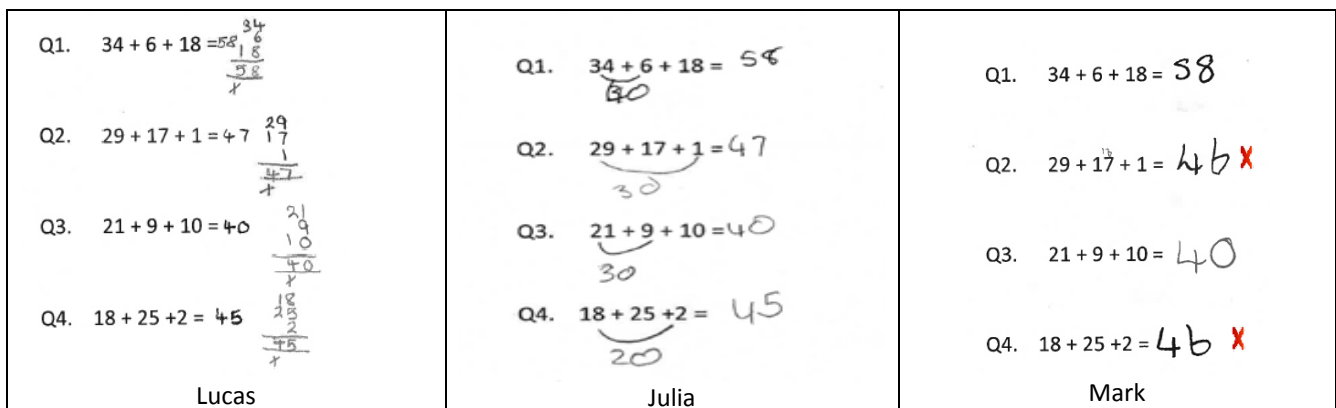


Fig 5. Paper-and-pencil responses from three students

In Figure 6, Lucas appeared to have used as-presented strategy, arranging the solution items. Julia appeared to have used the place-value strategy in all the questions, and intermediate answers are shown. In contrast to Lucas and Julia, Mark did not show his workings hence it may not be possible to provide accurate feedback on where and why he missed the correct answers for questions 2 and 4.

The MuTAT application provides rich and consistent data that can be used to assess in detail the solution of each student. Figure 6 shows the diagram of the solution steps obtained from the tool for one of the study questions. Here Alan paired 6 and 14 as a first step, then entered an intermediate result of 30. He used this result in the second calculation stage pairing it with 18 before entering a final result of 48.



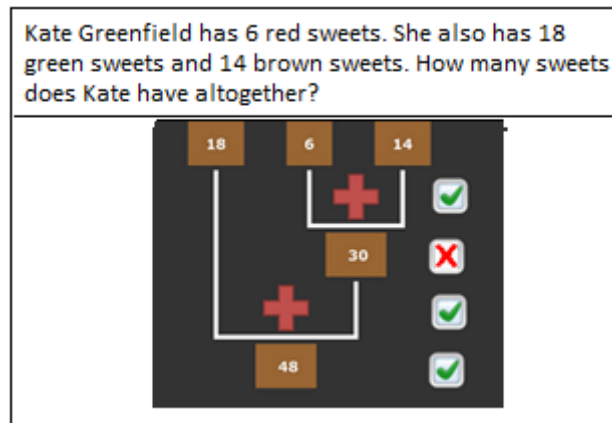


Fig 6. MuTAT Diagrammatic representation of solution steps from Alan

From this figure it is easier to see why and where Alan had difficulties in answering the question correctly. While he understood the problem and selected the correct operators, he got the first addition step wrong. A more accurate feedback can be given to him and he may be assigned partial credit for the parts of the solution he got right.

#### 5.4 Strategies

As discussed in the literature review, performance on a problem-solving task is dependent on both strategy selection and strategy efficiency (Mayer & Wittrock, 1996). It is possible to use the data from the MuTAT application to examine these two characteristics. The following sub-sections discuss the strategies used by the participants.

##### 5.4.1 Strategy selection

As discussed in Section 2.3, strategy types can be inferred from the way the students paired the numbers. The students all completed 5 study questions. The strategies obtained from the MuTAT platform, ignoring pairwise commutations is summarised in Figure 7. In this Figure the five study questions are identified as 3, 4, 5, 6 and 7 respectively. Question 7 is the control question described in Section 5.2. In Figure 7, it can be seen that the use of strategies varied across questions for the 39 participants. More students used the place-value strategies on questions 4, 6 and 7, while the as-presented strategy was used more frequently on questions 3 and 5. Most of the students used as-presented on question 3 possibly because it was the first study question and they were starting to get used to the application. The reason why no participants were coded as using the as-presented strategy in the control problem (question 7 in Figure 7) is that it was designed such that the as-presented and place-value strategies could be counted together (both were coded as place-value).

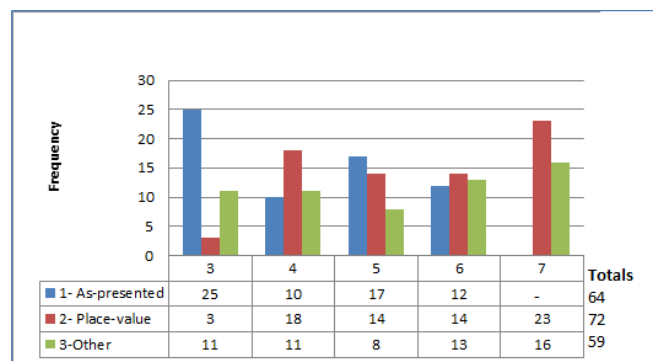


Fig 7. Strategies output across questions

In order to investigate whether the participants were working strategically the place-value strategy was subjected to a goodness-of-fit test. Had the students been selecting and pairing the numbers randomly the expected frequency for each of the three strategies on a given question would be 13 (i.e.  $39 \div 3$ ). The observed place-value strategy frequencies across the five questions were 3, 18, 14, 14 and 23 respectively. A one-sample chi-square test was significant but marginal,  $\chi^2(4, N=39) = 9.58, p = .048$ , suggesting that some but not all of the students were working strategically. This outcome is consistent with the hypothesis that some participants would use the place-value strategy, and others would add the numbers in the order presented or arbitrarily in each question.

A closer examination of how individual students consistently used the place-value strategy is presented in Table 3.

**Table 3**  
Usage of the place-value strategy

| No of times place-value was used | No of students |
|----------------------------------|----------------|
| 0                                | 6              |
| 1                                | 14             |
| 2                                | 8              |
| 3                                | 2              |
| 4                                | 9              |

The Table showed that about half of the students (20) used place-value once or not at all; Of the 14 who used place-value once, 8 used it on the control question; and of the 20 who used place-value once or not at all, 9 used mostly the as-presented strategy. This suggests the MuTAT application may have detected strategic groupings amongst the students between those who selected the numbers as presented and those who were more strategic.

#### 5.4.2 Strategy efficiency

Strategy efficiency examines how quickly and accurately a strategy leads to the solution. Figure 8 presents data from the MuTAT on the strategies and the average times it took children using the strategies. The data shows that the students spent less time using the place-value strategies in three out of the five questions.

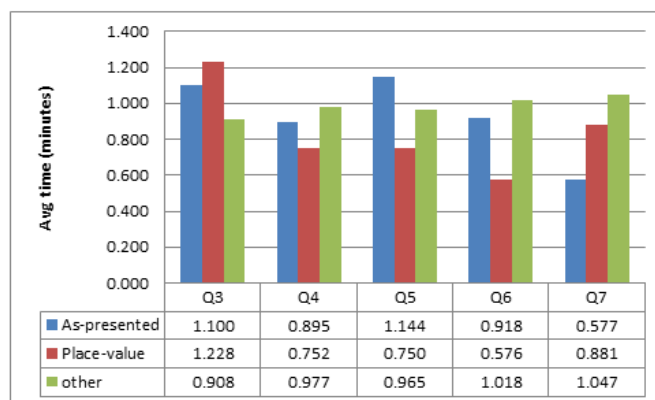


Fig 8. Strategies and the average times

To compare the completion times of the three strategies, a one way ANOVA F-test was used. The result was significant ( $F = 4.9, p = 0.01$ ) indicating that there are differences between at least two of the average time values. To check for where the differences lie, a TukeyHSD pairwise comparison was carried out at 95% confidence interval. Figure 9 shows a plot of the differences in the mean times.

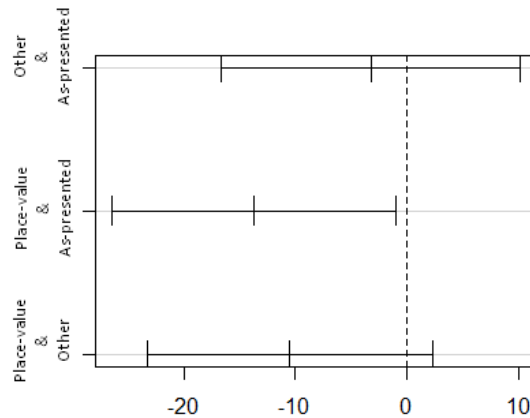


Fig 9. TukeyHSD of difference in average times of strategies

The test revealed significant differences between the mean time of the place-value strategy ( $p=0.01$ ) and the as-presented ( $p=0.13$ ) and other ( $p=0.58$ ) strategies. This agrees with Imbo and Vandierendonck (2008) who argued that computational strategies are more efficient than procedural strategies. Therefore the MuTAT application can detect the relative efficiency of strategies

#### 5.4.3 Qualitative analysis

To determine if the strategies used on paper-and-pencil corresponded with that used on the MuTAT application, the responses of Lucas, Julia, and Mark (see Figure 6) were closely analysed. Lucas used the as-presented strategy in all four questions of the paper-and-pencil test and he used the same strategy in all the five questions on the MuTAT. Julia used the place-value strategy in all the paper-and-pencil test questions and in four of the five questions on the MuTAT. Mark on the other hand used a mixture of strategies; as-presented on 2 questions, place-value on 2 questions and *other* strategy of the five questions on the MuTAT. As mentioned in Section 6.3, Mark's strategy on paper cannot be accurately determined since he did not show any workings. Analysis of the strategy output from the MuTAT suggests that he may require more support in learning to recognize relationships between numbers and working more strategically.

This result suggests that the strategies used on paper may not be different from the one used on the application. However this finding cannot be extrapolated to all students as most of them did not completely present their workings on paper.

## 6. Discussion

In this paper an approach for capturing detailed problem-solving steps in arithmetic word problems was presented. Overall the MuTAT application successfully achieved the research aims. It is a novel study that demonstrated that solution steps can be captured from traces of interactive problem-solving actions in an assessment environment by using multi-touch gestures on integrated workspaces.

### 6.1 Summary of findings

The Multi-Touch Arithmetic Tool showed how solution steps can be captured from trace information obtained from interactive actions in problem-solving. It showed that the products of an assessment task and the dependencies between the different parts of a solution effort as well as links to problem specification can be obtained and represented in a way that makes the thinking behind a solution visible. This is likely to provide valuable information that may be useful for detailed assessment of solution steps and for the provision of more accurate feedback.

The empirical study which investigated the practicality of using the tool in classroom situations suggests that students aged 8 and 9 years are able to successfully use the approach to solve problems. The performance scores and

timings from the tool were comparable with those obtained from paper-and-pencil tests. A close examination of three student's MuTAT and paper-and-pencil solutions indicated that strategies were consistent across the two media. This suggests that the tool did not impede students' accuracy and efficiency when solving mathematical word-problems. While performances and strategies output were similar in both media, the process data output from the MuTAT application provided deeper insights into students' strategies than the paper responses. This can facilitate small-scale and large-scale assessments focused not only on answers but also on strategies. Moreover, the approach provides students with increased opportunities to get relevant feedback to revise and improve their understanding and reasoning. Additionally, the approach provided opportunities for more detail scoring, intermediate steps such as the use of correct operations and intermediate answers can be scored independently of final answers. With this assessors can assign proper credit for the evidences collected.

## 6.2 Implications

The study may contribute to understanding how gesture-based user interfaces are used to capture the thinking process and reveal strategies in elementary mathematics problem solving. While well researched word-problems were used in the study to allow strategies to be inferred MuTAT can generally be used where such strategic problems do not exist.

Computer-Aided Assessment and feedback generation can be extended beyond just providing comments and scores to the final product of elementary maths assessment tasks. Because consideration of the whole steps and actions during the problem-solving process provides useful information, teachers, parents and practitioners can use this approach to improve on their feedback processes.

Some scholars have suggested that word problems may restrict children in the use of their conceptual knowledge (Gilmore & Bryant, 2006). Teachers and researchers can use the tool to examine how students understand the wording of a question.

## 6.3 Limitations

Three limitations may be related to data collection and the interpretation of results. A first limitation might be the constraint on the tool to present the stepwise information horizontally. Some students may feel better doing the sums vertically, especially when it involves a carry-over digit.

Another potential shortcoming in the study is the bias which may be introduced by the non-randomization of the MuTAT and paper-and-pencil test treatments on the participants. All the students started with the MuTAT application before crossing over to the paper-and-pencil paper test. This however should not have significant effect on the overall outcome of the study.

A third potential limitation is related to design challenges that may result from the small size of the tablet screen (9.7-inches) that was used in the experiment. Interacting with objects on touch screen requires reasonable screen size. This constraint may affect problems that require three or more stages.

## 6.4 Directions for future research

The capture of detailed solution steps results in a large amount of data to be processed and analysed. In future investigations it might be possible to automate or semi-automate the analysis and feedback activities of teachers and thus avoiding a significant increase in workload.

## 7. References

- Bailey, G. (1993). Iterative methodology and designer training in human-computer interface design. In *INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Vol. CHI '93). Amsterdam, The Netherlands.
- Bailly, G., Müller, J., & Lecolinet, E. (2012). Design and evaluation of finger-count interaction: Combining multitouch gestures and menus. *International Journal of Human-Computer Studies*, 70(10), 673–689. doi:10.1016/j.ijhcs.2012.05.006
- Baker, E. ., & Mayer, R. . (1999). Computer-based assessment of problem solving. *Computers in Human Behavior*, 15(3–4), 269–282. doi:10.1016/S0747-5632(99)00023-0
- Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning? *Metacognition and Learning*, 3(1), 39–58. doi:10.1007/s11409-007-9009-6

- Barmby, P., Harries, T., Higgins, S., & Suggate, J. (2009). The array representation and primary children's understanding and reasoning in multiplication. *Educational Studies in Mathematics*, 70(3), 217–241. doi:10.1007/s10649-008-9145-1
- Baroody, A. J., & Gannon, K. E. (1984). The Development of the Commutativity Principle and Economical Addition Strategies. *Cognition and Instruction*, 1(3), 321–339. doi:10.1207/s1532690xci0103\_3
- Batmaz, F., Stone, R., & Hinde, C. J. (2009). Personal Feedback with Semi-Automated Assessment Tool for Conceptual Database Model. In *10th Annual Conference of HEA-ICS* (pp. 115–119). University of Kent, Canterbury UK.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. doi:10.1080/0969595980050102
- Borthwick, A., & Harcourt-Heath M. (2012). calculating: what can year 5 children do now ? In *Proceedings of the British Society for Research into Learning Mathematics* (Vol. 32).
- Byrnes, J. P., & Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology*, 27(5), 777–786. doi:10.1037/0012-1649.27.5.777
- Carpenter, T. P., & Moser, J. M. (1984). The Acquisition of Addition and Subtraction Concepts in Grades One through Three. *Journal for Research in Mathematics Education*, 15(3), 179–202. doi:10.2307/748348
- Chung, G. K. W. K., & Baker, E. L. (2003). An Exploratory Study to Examine the Feasibility of Measuring Problem-Solving Processes Using a Click-Through Interface. *The Journal of Technology, Learning and Assessment*, 2(2). Retrieved from <https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1662>
- comScore. (2011). Smartphones and Tablets Drive Nearly 7 Percent of Total U.S. Digital Traffic. *comScore*. Retrieved February 2, 2014, from <http://www.comScore.com/DigitalOmnivores>
- Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *ALT-J*, 13(1), 17–31. doi:10.1080/0968776042000339772
- Cowan, R., & Renton, M. (1996). Do They Know What They Are Doing? Children's Use of Economical Addition Strategies and Knowledge of Commutativity. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 16(4), 407–20.
- Dick, J. (2002). Rich traceability. In *Proceedings of the 1st international workshop on traceability in emerging forms of software engineering*, (pp. 18 – 23). Edinburgh, Scotland.
- Dixon, J. A., & Moore, C. F. (1996). The developmental role of intuitive principles in choosing mathematical strategies. *Developmental Psychology*, 32(2), 241–253. doi:10.1037/0012-1649.32.2.241
- Gilmore, C. K., & Bryant, P. (2006). Individual differences in children's understanding of inversion and arithmetical skill. *The British Journal of Educational Psychology*, 76(Pt 2), 309–331. doi:10.1348/000709905X39125
- Hanna, G. S., & Dettmer, P. A. (2004). *Assessment for effective teaching using context-adaptive planning*. New York: Pearson.
- Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487
- Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87(1), 18–32. doi:10.1037/0022-0663.87.1.18
- Heinrich, P. (2011). The iPad as a Tool for Education: A study of the introduction of iPads at Longfield Academy, Kent. Naace. Retrieved from <http://www.naace.co.uk/publications/longfieldipadresearch>
- Hommel, B., Colzato, L. S., Fischer, R., & Christoffels, I. K. (2011). Bilingualism and Creativity: Benefits in Convergent Thinking Come with Losses in Divergent Thinking. *Frontiers in Psychology*, 2. doi:10.3389/fpsyg.2011.00273
- Imbo, I., & Vandierendonck, A. (2008). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: differences between children and adults? *Psychological Research*, 72(3), 331–346. doi:10.1007/s00426-007-0112-8
- Jiao, X., Deng, H., & Wang, F. (2010). An Investigation of Two-Handed Manipulation and Related Techniques in Multi-touch Interaction. In *2010 International Conference on Machine Vision and Human-Machine Interface (MVHI)* (pp. 565 –568). doi:10.1109/MVHI.2010.27
- Jones, I., & Pratt, D. (2012). A Substituting Meaning for the Equals Sign in Arithmetic Notating Tasks. *Journal for Research in Mathematics Education*, 43(1), 2–33. doi:10.5951/jresmetheduc.43.1.0002
- Lewis, J. . (2006). Usability testing. In *Handbook of Human Factors and Ergonomics* (pp. 1275–1316). New York: John Wiley.
- Mayer, R. ., & Wittrock, M. C. (1996). Problem-solving transfer. In *Handbook of educational psychology* (pp. 47–62).
- McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education*, 39(3), 299–312. doi:10.1016/S0360-1315(02)00032-5

- Mercier, E. M., & Higgins, S. E. (2013). Collaborative learning with multi-touch technology: Developing adaptive expertise. *Learning and Instruction, 25*, 13–23. doi:10.1016/j.learninstruc.2012.10.004
- Moscovich, T. (2007). *Principles and Applications of Multi-touch Interaction*. Brown University, Providence, RI, USA.
- Pinheiro, F. A. C., & Goguen, J. A. (1996). An object-oriented tool for tracing requirements. *IEEE Software, 13*(2), 52–64. doi:10.1109/52.506462
- Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*(1), 175–189. doi:10.1037/0022-0663.91.1.175
- Rochimah, S., Wan Kadir, W. M. N., & Abdullah, A. H. (2011). Utilizing Multifaceted Requirement Traceability Approach: A Case Study. *International Journal of Software Engineering & Knowledge Engineering, 21*(4), 571–603.
- Silke, Ladel, & Kortenkamp, Ulrich. (2011). Implementation of a multi-touch-environment supporting finger symbol sets - Google Scholar. *Proceedings of the Seventh Congress of The European Society for Research in Mathematics Education, CERME-7*, pp. 2278–2287.
- Stone, R. G., Batmaz, F., & Rickards, T. M. (2010). A multi-touch ER diagram editor to capture students' design rationale (Vol. 1). Presented at the World Congress on Engineering and Computer Science, San Francisco, USA. Retrieved from <https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/7956>
- Tang, A., Jin, Y., & Han, J. (2007). A rationale-based architecture model for design traceability and reasoning. *Journal of Systems and Software, 80*(6), 918–934. doi:10.1016/j.jss.2006.08.040
- UPI.com. (2012). Tablets giving PCs a run for the money. *UPI*. Retrieved August 18, 2013, from [http://www.upi.com/Business\\_News/2012/03/05/Tablets-giving-PCs-a-run-for-the-money/UPI-37341330981057/](http://www.upi.com/Business_News/2012/03/05/Tablets-giving-PCs-a-run-for-the-money/UPI-37341330981057/)
- Von Knethen, Antje, & Barbara Paech. (2002). A survey on tracing approaches in practice and research. *Fraunhofer Institut Experimentelles Software Engineering*, (IESE-Report No, 95).
- Ward, W. C., & Bennett, R. E. (2012). *Construction Versus Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment*. Routledge.
- Wigdor, D., & Wixon, D. (2011). *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Elsevier.
- Williamson, D. M., Bauer, M., Steinberg, L. S., Mislevy, R. J., Behrens, J. T., & DeMark, S. F. (2004). Design Rationale for a Complex Performance Assessment. *International Journal of Testing, 4*(4), 303–332. doi:10.1207/s15327574ijt0404\_2
- Young, M. (1995). Assessment of situated learning using computer environments. *Journal of Science Education and Technology, 4*(1), 89–96. doi:10.1007/BF02211586
- Zoanetti, N. (2010). Interactive computer based assessment tasks: How problem-solving process data can inform instruction. *Australasian Journal of Educational Technology, 26*(5), 585–606.