Multiple Mice for Retention Tasks in Disadvantaged Schools

Udai Singh Pawar[1] **Joyojeet Pal**[2] **Rahul Gupta**[3] **Kentaro Toyama**[4]

[1,3,4] Microsoft Research India
Scientia, 196/36 2nd Main
Sadashivanagar, Bangalore, India 560080
{udaip, t-ragup, kentoy}@microsoft.com

[2] TIER Research Group
University of California at Berkeley
634 Soda Hall, Berkeley CA, USA 94720
joyojeet@berkeley.edu

ABSTRACT

This study evaluates single-mouse and multiple-mice configurations for computer-aided learning in schools where access to computers is limited due to resource constraints. *Multimouse*, a single display groupware solution, developed to allow multiple mice to be used simultaneously on a single PC, is compared with single-user-single-mouse and multiple-user-single-mouse scenarios. Multimouse itself is trialed with two unique interaction designs – one where competitive interaction among students is encouraged, and another where more collaborative interaction is expected.

Experiments were conducted with 238 schoolchildren from underprivileged households in rural India on an English vocabulary retention task. On the whole, Multimouse configurations (five users each) were found to be at par with single-user scenarios in terms of actual words learned by students. This suggests that the value of a PC can be inexpensively multiplied by employing a multi-input shared-use design. Gender effects were found, where boys show significant differences in learning depending on interaction modality, whereas girls learned at similar rates across configurations. In addition, a comparison of the two Multimouse modes – collaborative and competitive – showed the striking difference in learning outcomes and user behavior that is possible due to even slight variations in interaction designs for multiple-mice.

Author Keywords

Single Display Groupware, Shared Computers, Education, Multiple Mice, Developing Nations

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces — Input devices and strategies, Interaction styles, Evaluation/methodology.

INTRODUCTION

Amidst much debate on the role of computers in children's education [17], computing technology is slowly but increasingly permeating classrooms in developing countries, even in very remote rural areas. This is often the result of government initiatives and work by the non-profit sector [14, 25].

Given that PCs are appearing in poorer schools, how can we maximize their educational value in these resource-strapped settings? To address this question, we began with an ethnographic study of schools with PCs in four states of India and three countries in Africa. The challenges were many, but one consistent finding was that PCs in these schools were used in a one-to-many fashion with multiple children – in some cases as many as ten, crowded around a single PC [18, 19] (Figure 1). Such shared usage was not uncommon even in relatively wealthy schools in urban areas, nor for that matter even in developed regions, especially among younger children. Not surprisingly, financial constraints generally prohibit the one-child-to-one-PC set-up seen in many schools in the developed world. Since standard hardware and software interfaces are designed for single-user input, the inevitable result is that the majority of children in such shared-use scenarios are not able to fully interact with the computer at any given time. Typically, a dominant child monopolizes the input devices, preventing other students from interacting with the PC, which in turn creates learning inequities [18].

Figure 1. Eight boys sharing a PC in a school in rural India.

In earlier work [19], we proposed the use of an existing paradigm known as "single display groupware" (SDG) [28] to tackle this problem. Our instantiation of this, called *Multimouse*, allowed many USB mice to be plugged into a single PC; the software presented a separate, uniquely colored cursor on-screen for each mouse. In preliminary studies [19], we investigated (1) whether children could understand the Multimouse paradigm, (2) how children interact with and share multiple mice, and (3) how engagement with the PC changes with multiple mice. We found that not only were children able to absorb the Multimouse paradigm with five mice easily and rapidly, but that excitement and individual engagement with the PC dramatically increased. These results suggested positive value in using Multimouse, but since the study was preliminary and qualitative in nature, it did not make a conclusive case for Multimouse's educational value.

RELATED WORK

Previous research on single display groupware (SDG), includes use of shared whiteboards such as [20], multi user editors (MMM [3]), handheld devices [15], tabletop interfaces like [23], and more [24].

In education, a few experiments have been conducted, but only with up to three mice and three students. These include puzzle solving games, and a mathematical game for pairs of children [9, 11, 21, 22]. In one case [7], pairs of students engaged in a collaborative search. In another case [5, 6], three children were asked to match a particular color, with each child restricted to control only one of the red, green, and blue components of color. In other work, two children were allowed to draw together using a collaborative drawing application called KidPad [1, 2, 26, 27, 28, 29].

In all of the cases above, multiple mice were observed to affect certain behavioral attributes – leading to enhanced motivation, increased student engagement, and greater task performance. SDG also seemed to provide an opportunity for more collaboration. Yet, rigorous quantitative testing of the educational value of multiple mice, compared with single-mouse, and single-user scenarios had not been conducted. One study observed, "Certain styles of collaborative behavior are desirable from an educational perspective, such as co-construction of ideas and resolution of conflict. These types of behaviors, whilst evident within some of the pairs sharing a mouse, were *not* observed within any of the two mice pairs [26]". Finally, previous work seemed to struggle for a strong reason to use SDG in education, whereas in resource-constrained schools, the case for multiple mice is clear.

CURRENT WORK

In this paper, we present novel work that takes a step toward answering some deeper questions about multiple mice: *What is the relative educational value of Multimouse compared with single-mouse configurations? And, what kinds of software designs encourage the most learning with Multimouse?*

Educational value is complex to evaluate, and conceptually open to debate, so we restricted our scope at this stage to a simple memory retention task, for which we believe the educational value is clear, if limited. Our aim was for a simple, objectively quantifiable metric that establishes some basic merit, before continuing with longitudinal studies over time to look at broader claims. In particular, we tested a simple vocabulary memorization task, supported by multiple-choice educational software under several configurations that include single-user-single-mouse, multiple-user-single-mouse, and multiple-user-multiple-mice. For the multiple-user-multiple-mice (*i.e.*, Multimouse) modes, we additionally designed and compared two kinds of interactions: one intended to foster competition, and another meant to encourage collaborative learning.

We expected that increased engagement would lead to more learning, and that either or both of the competitive and collaborative environments would be more conducive to learning. To test these hypotheses, experiments were conducted with 238 rural schoolchildren. We found firstly, that in certain instantiations, Multimouse performed as well, *per user,* as a single-user-single-mouse model, with respect to retention tasks. This establishes Multimouse's potential to inexpensively multiply the value of a single PC for a subset of educational tasks. Secondly, we also found gender effects, adding to results from prior work [1, 10, 12]. For girls, no significant differences were seen across interaction configurations, but for boys significant differences were found. Boys fared very poorly in the shared multiple-user-single-mouse mode and the competitive Multimouse mode, and those in the collaborative Multimouse mode outperformed their peers in the other multiple-user modes. Educational value appeared to depend significantly on the way the software was written to accommodate multiple students. Specifically, we found that is important to take a nuanced view of factors such as engagement, collaboration, and identities when talking about design of multiple-mice-based interactions.

TECHNICAL DETAILS

Multimouse is an SDG design, where a single display is shared by multiple users. This entails a separate mouse and separate on-screen cursor for each user.

Multiple mice are connected via USB ports, and additional USB hubs can theoretically allow any number of mice to be added. The practical limit is in the lag between events as more mice are added and frequency of events increases. We have tested up to 15 mice simultaneously with minimal impact on the responsiveness of each mouse.

Multimouse capability is currently applied to a single application at a time and is not part of the GUI shell itself. We built our software development kit in C# in Windows, using the RawInput API in Win32. The API provides a unique identifier for each device, and allows us to access device-specific events separately. Event streams from each mouse are used to control multiple on-screen cursors, which are explicitly rendered by the Multimouse software. The normal system cursor handled by the OS is disabled.

A key point to note is that the technology is completely software based, and hence cheap and easy to distribute. This is relevant while considering scalability in the developing-nation scenarios we are discussing.

DEVISING THE EXPERIMENT

In order to test the value of Multimouse for educational purposes, much of our effort went into the design of the experiment, which we felt needed to have the following characteristics for the results to be meaningful:

- The tests should have **quantifiable, objectively verifiable metrics.**
- The results should be measurable in the **short term** to establish initial results before going on to potentially more complex results.
- The tests ought to represent a **broad class of educational scenarios**, for greater generalizability.
- Subjects should be chosen carefully, so that **use of the computer itself is not a barrier** to testing learning of the intended material.
- The tests should be designed so that they cast **single-user-single-mouse and multiple-user-single-mouse cases as one-mouse (and one-cursor) instances of the *n*-mouse Multimouse** tests. This would allow the results of the different configurations to be compared fairly with one another.
- The content of the test should be of **practical educational value**, so that the results could apply to real-world educational content.
- The test should not be held in a laboratory scenario, but designed to be **consistent with how students actually use PCs in their normal classrooms**, since field settings are unique.

We concluded that a multi-question multiple-choice retention quiz for English word-image associations, tested with middle-school children in their schools' computer rooms would fit all of the desiderata. In the following subsections, we outline our rationale for these choices.

English Vocabulary Retention Tasks

The simplest kind of learning outcome is in the 'knowledge level' at the base of Bloom's Taxonomy of Learning Outcomes [4]. This is a reasonable starting point to establish proof of concept at basic levels of cognition, rather than at higher levels such as "comprehension" or "synthesis of knowledge," which are dependent on a variety of other factors and usually require long-term studies to verify. As Bloom notes, factual knowledge is the foundation of the other learning outcomes, thus if results can be shown at this level of the taxonomy, they can form the basis for testing broader claims later.

Retention tasks also satisfied the constraint of practical value and breadth, as our study of existing computer-aided learning for children in India showed that much of the curriculum for younger children was in the knowledge level of the taxonomy, where rote retention is critical. Straightforward retention tasks are easily tested by multiple choice questions and so we settled on a multiple-choice paradigm for its simplicity and objective measurability.

As to choice of content, we studied content currently being used by primary-school students, in consultation with curriculum-development agencies. We identified ESL (English as a Second Language) as an appropriate starting point, based partly on the practical need for ESL instruction in developing regions [13], and partly on the enthusiasm among parents to have their children learn English. We decided on word-image associations to both teach and test English vocabulary, as we needed something generic within ESL, but also applicable across other learning scenarios which require mental retention.

We wanted a vocabulary of terms that we anticipated would be mostly new to our subjects, to remove confounding factors of any previous knowledge. For the actual tests, we picked animal names, as animals are readily depicted in static pictorial form, and because we could easily scale the difficulty of the vocabulary to take note of previous knowledge. Our word list, starting with 37 words, was iteratively developed, with input from teachers and curriculum developers. It comprised words known to our subjects (dog, cat), words highly unlikely to be known (koala, llama), and some that might be known (rabbit, deer).

Test Subjects

For our experiment, we sought a student population that was mature enough to be able to undergo the experiment, able to read and write basic English, and reasonably comfortable with computer use.

Our subjects were children in grades 6 and 7 from three rural public schools near Bangalore, India who were in their second or third year of English-language education. They were thus able to competently read and write elementary words in English, but had a fairly limited vocabulary. The typical student profile was of a child from a family in agricultural labor, with no cases where English was spoken at home. Overall, access to any English language media was practically non-existent. In all three schools, computer aided learning programs were in effect for two years. All the children were proficient in use of the mouse.

Software Design

The most critical component was the interaction design of the educational software that delivered the multiple-choice questions. We went through a number of iterations, running preliminary trials with some children, to finalize the design.

We knew from previous studies of computer usage in rural middle schools that students focus on interactivity, and are bored by explicitly instructional material [18]. If there is a tutorial video to watch (for example), followed by a quiz, students generally prefer to jump straight to the quiz, rather than sit through the video. So, we combined content delivery and testing in a single, interactive user interface.

Single-Mouse Modes (Single User and Multiple User)

In the single-user version of the software, the application shows a prompting image and multiple choice buttons, with each button labeled by a word (Figure 2). The screen displays "What do you see in this picture? Click on the correct answer." One of the words is the correct name for the prompting image, and if the correct option is clicked on, there is positive reinforcement – visually, the image and the correct word-button are highlighted, and the voice playback says "That's correct!" If the wrong option is chosen, there is generic negative feedback. The user is allowed to keep trying until the correct option is chosen, at which point, a new multiple-choice question is posed. Students learn by trial and error in the same multiple-choice questioning environment that also allows them to score points.

A score bar is displayed at the bottom of the screen (note the arrays of stars, depicting points, at the bottom of Figure 2). A correct answer merits a score of 10 points, and an incorrect answer results in a deduction of 4 points, meant to deter random clicks. A congratulatory pop-up comes up whenever the score crosses multiples of 100 points (these and numerous other design elements were refined through iterative testing and prototyping with small groups of children at a local after-school computer center.)

Through repetitions of this simple interaction, learning is reinforced; a single image-word pair is repeated multiple times to encourage retention. Word choice is stochastic, but not completely random. Memory studies indicate that information is most efficiently absorbed when it is repeatedly presented, but at exponentially decreasing frequency [8]. In the test application, words are thus repeated, but with decreasing frequency.

The software, as described above, is exactly what we used to test the **s**ingle-user-**s**ingle-mouse configuration (**SS**), as well as the **m**ultiple-user-**s**ingle-mouse configuration (**MS**) where the same software is used, but with many children sharing one mouse.

To extend the test application for multiple mice, we considered two different modes of interaction, one meant to encourage competition, and the other, collaboration. Both variants needed to be functional generalizations of the single-mouse case.

Multimouse: Multiple Mice in Racing Mode

In the first case, which we call ***m**ultiple-user-**m**ultiple-mice (or Multimouse) **r**acing* mode (**MM-R**), we set up a competitive environment, with multiple cursors, in which multiple students race to see who can click on the correct button for each multiple-choice question posed. The set-up is identical to that for the single-mouse case, with the generalization to multiple mice happening such that the application proceeds to a new question as soon as one of the students clicks on the correct button. We point out in particular, that once one student answers correctly, the other students do not have a chance to click on the correct button.

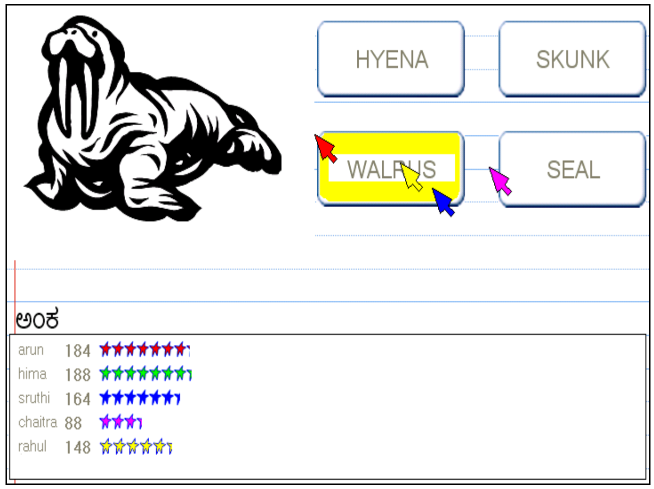

Figure 2. Screenshot (in Multimouse Racing Mode). Note the multiple cursors and the score box below. For the single-mouse modes, the difference is that there is a single cursor and single score bar.

To highlight the competitive aspect, score bars for each student (colored the same as their cursors) are stacked on top of each other. Scores are incremented as before, with +10 for correct answer and –4 for incorrect. The feedback on a correct answer includes mention of the student's name and cursor color, as well. "Red is correct!" We emphasize that the one-mouse case in MM-R mode is identical to the single-mouse application.

Preliminary trials with MM-R mode revealed that a share of the clicking was random, based on a speed-based competitive strategy. In this strategy, children hoped to score in the game through lucky clicks, so there was no real need for image-word association. Moreover, the faster turnaround of answers and consequent shift to the next screen denied those who may have learnt from visual recall from scoring because of the short screen time of each item. So, we also designed another mode of interaction that would inhibit outright competition.

Multimouse: Multiple Mice in Voting Mode

The second generalization of the single-mouse case that we developed is ***m**ultiple-user-**m**ultiple-mice **v**oting* mode (**MM-V**), detailed below, in Figure 3. All *n* children at a computer have their own mice and cursors. When a child clicks on a button, it becomes the color of that child's cursor (Step A). If another child clicks on that, then half of the button becomes one color, while the other half takes up the color of the other child's cursor (Step B). If a third child clicks, the button changes to show all three colors, split into thirds (Step C), and so on. This process can happen concurrently on all buttons on the screen (note Step C). When all the children click on a single button, that answer is chosen (Step D, for n=5). If correct, positive feedback is delivered in the same way as for the single-mouse case; otherwise, negative feedback is delivered, and points withdrawn. Once again, the interaction for the one mouse case (n = 1) of MM-V reduces to the single-mouse scenario.

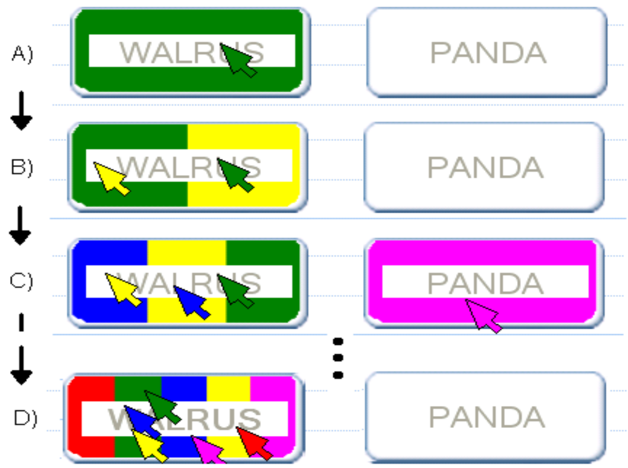

Figure 3. The voting mechanism for *n*=5. As a child clicks on an option (Steps A-C), the button takes up the color of that child's cursor, until all *n* children have chosen a single option.

Here, a collective, single decision is required to proceed, and so we felt that there may be benefits to educational value, due to the collaboration required. At the same time, each child has a chance to "catch up" with the leader, because the application does not proceed without clicks from all mice. Also, the leading child has some incentive to help the others, in order to proceed to the next question.

To summarize, we designed four different modes for testing, based on multiple-choice image-word matching questions. The four modes are as follows:

1. ***SS** or **S**ingle-User-**S**ingle-Mouse*: one child at a single PC with one mouse
2. ***MS** or **M**ultiple-User-**S**ingle-Mouse*: n children at a single PC with only one mouse
3. ***MM-R** or **M**ulti**m**ouse **R**acing*: n children at a single PC with n mice; interaction is such that a new question is posed as soon as any child clicks on the correct word
4. ***MM-V** or **M**ulti**m**ouse **V**oting*: n children at a single PC with n mice; interaction is such that a new question is posed only if all children click on the correct button

Both the MM-R and MM-V modes reduce to the SS and MS interaction model when n=1. This allows a fair comparison of the configurations.

EXPERIMENT

Description of Test Session

The computer-based testing sessions were designed to be about 30 minutes long per child. To measure number of words retained per session, a paper-based pre-test was administered immediately to the student before the computing session, and a post-test, immediately after it. The students were picked and allotted groups and configurations randomly from attendance registers.

The computer sessions were in the regular computer room of the schools, whereas the pre- and post- paper-tests were conducted in a separate area. The time for the paper tests was set at 7 minutes based on previous iterations. To allay fears among children that they were being administered an exam, the exercise was called an 'Animal Game'. For the final test, there were 16 questions. The score was graded on 12 questions out of these, as the others were words which every child knew, to get them started on choosing answers and a repeat of one question, to see validity and consistency of answers.

Figure 4. Children were administered a pre- and post-test for English vocabulary so that we could measure the words learnt in the different experimental modalities. (In rural schools, sitting on the floor is common, due to lack of classrooms.)

The groups with multiple users (MM-R, MM-V and MS) had 5 children each, randomly assigned, stratified by gender, into single-gender or mixed-gender groups. All computing sessions were overseen by a trained research assistant (RA). We picked five as the size of the group since this was typical in rural schools and because we wanted to differentiate strongly between single- and multiple-user cases. Also, beyond five, physical resources such as table space and screen viewability become non-uniform.

Description of Field Tests

We conducted a total of four field tests. The first, Pre-Trial#1 (P1) and third, Pre-Trial #2, (P2) were small-scale trials of just 2-3 sessions with N~30 each, to check modalities of the testing paradigm and software – the data and settings were not fully controlled. Experiment#1 (E1) was a controlled experiment in real-life conditions at a rural school, and we gained qualitative observations for N=140 children. Experiment#2 (E2, N=98) was the final controlled experiment. All trials were with 11- and 12-year olds (from the 6th and 7th grade) in rural government schools near Bangalore. Schools were picked to be as similar as possible in infrastructure, curriculum, and computing experience.

Tests P1, P2, and E1 were used to refine the testing scheme for E2 for quantitative data. Particularly, we arrived at the following set-up: a session had 12 words with ~7 repeats each. All words in E2 were generally new to the children.

Experimental Measures

The main measure of learning was the difference in pre- and post-test scores. Additionally, detailed transcripts of behaviors were made by RAs stationed at each session. They specifically noted conversations between the children, positions and mouse-sharing behaviors. Also, the application logged all mouse events with mouse IDs.

(NOTE**:** It is possible that the Hawthorne effect may have skewed all discussed results slightly, especially due to the presence of the RAs. Children could have paid more attention than usual. There may have been less extreme dominance and more sharing than normal, and in MM-R, possible impacts on competition.)

RESULTS

Test Data Summary: Learning Outcomes

The quantitative results showcased in this paper are from the controlled experiment, E2, of 98 children (48 boys, 50 girls). For analyzing statistical significance, we used the standard t-test, with alpha of 0.05, unless otherwise specified. The main metric used was the difference in pre- and post-test scores, measuring words learnt. The graphs in Figure 5 and 6 show the mean values for these for various modalities, classified by gender and grouping conditions. We ensured that children were from the same grade in each group, but found no significant differences in performance across grades, perhaps as there were significant age overlaps.

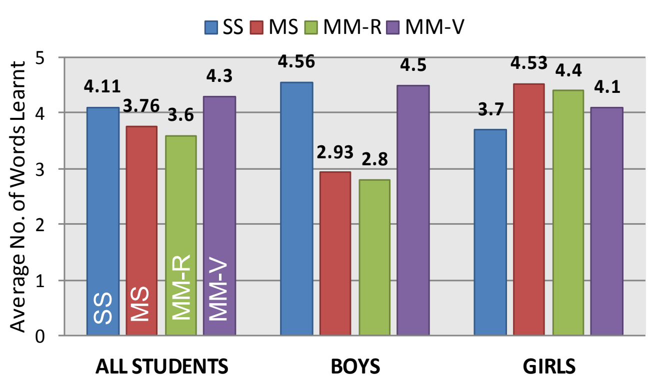

Figure 5. Aggregated learning outcomes. Results consistent for girls, while boys have more variations. Overall, Multimouse (MM) results match the single-user SS mode. (Trial E2, N=98.)

In aggregated results (Figure 5), the number of words learned did not differ significantly between the different modes tested. MM-V showed, on average, the highest number of words learnt (4.30), and MM-R the least (3.60), but these differences were not statistically significant. Thus, all results, even for the multiple-user-single-mouse case, showed learning outcomes comparable to the situation of a single child to a computer, SS (4.11). This suggests that it is possible with multiple mice to have outcomes, at least for learning rote material, that are fairly comparable to a single child per PC, even with as many as five children. This is of great interest, as it points out the benefits of using multiple mice in resource-strapped settings, where there are multiple children per PC, not by choice, but by constraint.

The results are more interesting when examined by gender. Girls have more consistent learning outcomes across modes, and generally outperform boys in multiple-user modes. Overall, for girls, shared-use scenarios of MS, MM-V, MM-R (4.38 average) were a slight improvement over the single user SS scenario (3.70). Boys showed greater variation in learning outcomes.

Boys did not do well when they had to share a mouse, and also did poorly in the competitive mode. The best learning results are for MM-V (4.50) and SS (4.56), while MS (2.93) and MM-R (2.80) were laggards. Multimouse voting MM-V was significantly better than the multiple-user-single-mouse (MS) mode (with $t(20) = 2.13$, $p<0.05$) and MM-V was also significantly better than the Multimouse racing mode, MM-R (to $p<0.06$, with $t(23) = 1.99$). It seems the competitive nature of MM-R hampered thoughtful decision-making, skewing instead towards impulsive clicking, consequently affecting word learning and retention. Similarly, boys did not cooperate well in the shared mouse MS mode. The mouse-controlling child tended to control the pacing – with the others often losing interest, which could explain the low MS mode scores.

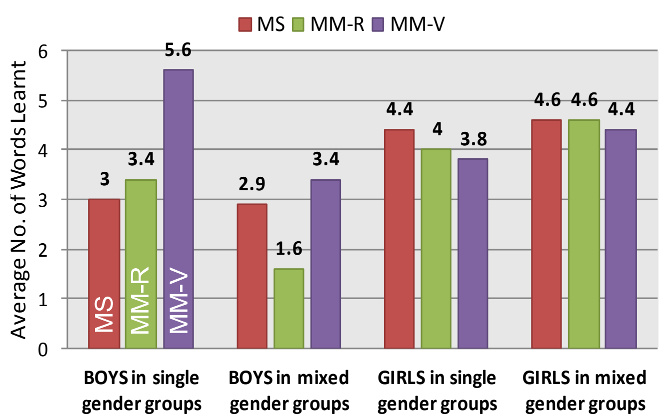

Figure 6. Learning outcome results classified by gender grouping scenarios, for all multiple-user modes. There are significant relative gains in using the MM-V mode for boys.

In all, the data supports the claims that there can be strong gender-based differences in shared-use scenarios [1, 10, 12], with added nuances and stronger empirical backing in educational terms. We also noted a marked contrast between the two groups in their performance in single-gender versus mixed-gender groups (Figure 6). For boys in single-gender groups, MM-V (5.6 words) was significantly better than MM-R with 3.4 words, (with $t(13) = 2.29$, $p<0.05$). MM-V was significantly better than MS, too (with $t(7) = 3.58$, $p<0.05$), whereas in mixed-gender groups, despite the consistent performance of MM-V (3.4), over MM-R (1.6) and MS (2.9), overall outcomes are lower. The

Figure 7. Children pointing and gesturing at the screen in multiple-user-single-mouse (MS) mode, hoping to influence the decision of the mouse-controlling child.

Figure 8. A group of boys in full concentration in the competitive Multimouse Racing (MM-R) mode.

outcomes for girls on the other hand were improved by mixed-group collaborations. Girls in all the mixed-gender groups (4.56 words on average) did significantly better than boys ($t(42) = 2.39$, $p < 0.05$) in all the mixed-gender groups (2.7), but marginally better than girls in single-gender groups (4.06). This diverges slightly from past work [1] where girls using KidPad in mixed-gender groups fared worse than girls in single-gender groups, perhaps as the nature of the task is vastly different.

In all, the data showed very encouraging results for the idea of shared input. The data for boys demonstrates statistically significant improvements for MM-V over MM-R and MS modes, comparable to the SS mode. Since MS mode is the dominant mode of use of computers in developing regions, and SS is infeasible financially, Multimouse offers tangible benefits. For instance, boys register significant gains (2.4 words) in MM-V in comparison to MS modes (in all-boy groups). The uniform outcomes for girls suggest their greater ability to work with sharing and collaboration.

Qualitative Observations: Engagement

Qualitative data gathered by the RAs (Table 1) supported some of the quantitative findings and also provided interesting contrasts on engagement across various modes. There was little or no conversation in the SS or MM-R modes – in the latter primarily due to the competitive nature of the game. The few conversations that took place in MM-R were competitive in nature, such as children rejoicing on correct clicks and mocking others over scores. Conversely, the MS mode had the most dialogue: the main pattern was conversations by non-mouse controlling children, verbalizing their choices for answers (Figure 7). Some cases of fighting, as well as argumentative banter over mice were recorded. In MM-V mode, the talk consisted mostly of children instructing each other on correct answers. Here, it was frequently recorded that distracted children were brought back into the game by others, and directed to click on appropriate answers, as the other children could not move on without that. In some cases, the mice were snatched and clicked by other team members.

Several cases of distraction and off-task behavior were recorded – the most were in SS, with children enquiring about game length or drifting off into boredom, as the flow of words and repetitions continued. In fact, in the SS mode some children lost attention due to loss of interest in the game, and this did not occur in any other setting. The non-mouse-controlling children in MS also lost interest easily.

Overall, engagement (as defined and quoted in past work as lack of off-task behavior [11]) was greatest in MM-R (Figure 8), though observations showed the focus seemed to be more on rapid, competitive clicking, rather than on the content, especially for boys. This maps to the poor quantitative results for boys in MM-R. In MM-V, children were quite engaged, and less distracted than in MS – which is reflected in the generally higher word retention for boys in the former. Also in MM-V, children seemed more engaged than SS overall, especially if we note that in SS, interest often tailed off over time. This suggests that given their own input device, children are highly engaged, but more so in a shared scenario than in a single-user scenario.

In all the scenarios, engagement was not impacted by game competency - children took a few minutes clicking all over the place to understand the game, until they accidentally clicked on a right answer, and after a few tries they usually understood the game. This was slower in SS than in MM-R and MS, as in those modes, often one child picked up the game quickly, and either taught the others, or they observed and learnt. MM-V took the longest time for the children to pick up, at about 5 minutes. Both MM-V and SS modes went by slower than others in actual game mode.

The key finding related to engagement was that both MM modes created a highly screen-attentive environment. Even though discussion was quite high in MS too, the exclusion of peripheral members was high as well, decreasing the aggregate group engagement. We conclude that engagement is best seen through the frame of ‘group goals’ rather than attention to content. Group goals [16], like moving on through the game or avoiding negative marks, had a stabilizing effect through the team, with a sense of collective well-being being tied to good performance. We

	Words Learnt	Engagement	Decision-making	Response error	Conflict (Boys)	Conflict (Girls)	Intra-group Competitiveness	Dominance by a child
SS	4.11	High, tails off	Individual	Low	n/a	n/a	n/a	n/a
MS	3.77	Low	Collaborative	Very Low	High	Low	Medium	Varied
MM-R	3.6	Very High	Individual	Med-High	Low	Low	Very High	None
MM-V	4.3	High	Collaborative	Very Low	Medium	Low	Low	Varied

Table 1: Findings Matrix for qualitative observations from experiments E1 and E2, N=238 ('Words Learnt' from E2).

find strong evidence of this by contrasting MS and MM-V against MM-R engagement. In MM-R, there were very few off-task activities, quick response to stimuli, high eye contact with the screen, but because of the lack of group goals, there was less cognitive benefit. So while both MM-R and MM-V relied on 'good performance' to do well with points, the realization of the need for careful decision-making emerged faster in a collaborative setting of MM-V than in the individualistic setting of the competitive MM-R.

Qualitative Observations: Identity and Dominance

In group-use scenarios, two important and related patterns emerged – sense of identity and pockets of power within groups in the interactive mode. In both the Multimouse modes, color played an important role in defining identity. In MM-R, the winners were defined by color bars corresponding to the cursors, while in MM-V, the changing color of the buttons, depending on who clicked, made color an important factor. A few minutes into the game, children began to refer to each other by color: "Click here, Red". The colors (or mice) also became associated with success – in MM-R, children in some cases said that a specific color was lucky, and that it had a better chance of correct responses, and fought over that mouse. In MM-V, colors impacted group and dominance, as leader-follower patterns emerged, when some students would click, following 'trusted' others, with trust often dependent on past success with correct answers.

In MS and MM-V, a sense of group developed. In both cases, if a child picked the wrong answer, there would be a reprimand from the rest of the group. There was sense of urgency in some cases ("Quick, we have to click quickly!") and joint celebrations on the "100 points" popup screens. In MS, some of the non-mouse-controlling children started noting answers from other groups, and began to drive inter-group competitions.

These findings also afford a deeper look at the single-child dominance in MS-type situations that has been noted in the past [18]. We found that children would get impatient in the MS mode if one child had the mouse too long. Failed attempts to get the mouse caused snatching incidents, complaints to the RAs, and in some cases slapping and pinching. However, observations showed no single power-sharing pattern. At one extreme, we observed an MS case where the group collaboratively decided that the privilege of being the mouse-clicker belonged to one 'representative'. At the other end, there were cases where the mouse controlling child ran a dictatorship.

Dominance is tied to knowledge legitimacy, of getting answers correct (like in MS, a mouse-controlling child getting answers wrong would be changed) but also to sense of initiative (early clickers in MM-V would often be followed). Accordingly, dominance impacted learning – either directly through the unequal use of resources, though more frequently in the shaping of interactions – as was the case of children following the leading student in going after a specific response in MM-V. However, the adaptive nature of the computer games seems to level dominance over the course of a game – for example, the child picking the wrong answers rarely went unpunished.

The observations on dominance revealed both technical and human interventions that spread control of resources. MS clearly had a great risk of single-user dominance, though this was also checked by group intervention during failures. This helped slow down MS, and gradually raised the amount of discussion. In contrast, MM modes had a lot less dominance by any single user, but lent themselves to an 'invisible free-rider' risk whereby one or a few users could consistently get wrong answers or not contribute, but never stand out for not learning.

Qualitative Observations: Collaboration

The term "collaboration", as opposed to mere "engagement", ought to apply more to the constructive discussion, rather than to imposing of dominance relationships. Like in the observations on engagement, we saw that goals were critical in defining the level of collaboration. In MM-R, where goals were more individualistic, collaboration was the least. The fast clicking observed in MM-R, discussed earlier, became faster and more competitive in advanced stages of the game as users

Figure 9. In Multimouse Voting, a slow child would be pulled up by the rest of the group. Sometimes though, an impatient child would try to snatch a slower child's mouse and click.

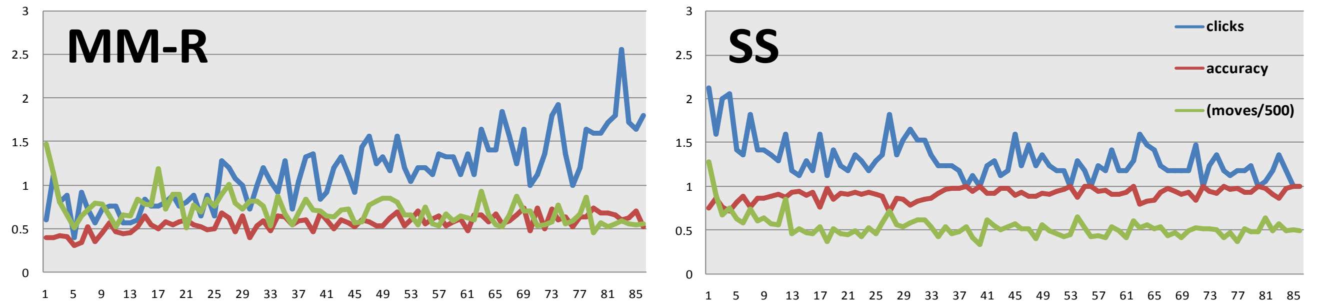

Figure 10. A sample of the log data for two modes, MM-R and SS, averaged over all children, and plotted against the number of words displayed. BLUE denotes total no. of clicks, ORANGE is accuracy and GREEN is normalized rate of mouse moves. Regular peaks appear in the curves, corresponding to the first displays of new words. The curves are smoothed here by averaging.

started coming closer to 'winning', with almost no discussion. MS saw more discussion, but was low on collaboration. Exchanges in MS had a tendency to be confrontational, without resolution, especially for boys, while girls generally showed evidence of some collaborative discussion. We can see the effect in the quantitative data, too (boys did poorly in MS, compared to girls, and to boys in MM-V). MM-V encouraged more children to collaborate and discuss, where children even tried to speak out the words on the screen, which hardly happened in other case. At the same time, there was often pressure on the slower children such as "Quick, click!" or the more extreme "I will kick you if you don't click". There were some occurrences of 'ganging up' – such as in one case, even after a 'slow' child in a group had sped up to match the rest, she was still attributed to slowing the team down when it performed poorly. There were also frequent cases of 'clicking for others' (Figure 9). In MS, the modes of collaboration were recurring – but these were largely related to mouse dominance or to deliberations on what to click. In contrast, nuanced patterns emerged in MM-V – the three that stood out are:

- *Leader/Followers:* In this, one child instructed and the others followed. This strategy was most common, and was usually led by a charismatic first clicker – though the same demotions applied here as in the case of dominant MS users who got answers wrong.
- *Joint Decisions:* This was less common, and data suggested that consciousness of negative marking increased the likelihood of this mode of interaction.
- *Majority following:* Here, the 'leader' was the option that got a couple of clicks by any members, often random – causing it to become a 'favorite' answer.

These observations on collaboration underline some important issues on educational value especially in MM-V mode. Quantitative data suggests that the collaboration inherent in MM-V played an important role, in contrast to MM-R or MS, in making it a more effective learning tool for boys. MS, which had a lot of discussion of a varied nature, and a high rate of 'correct' responses, did not lead to the best learning outcomes. In contrast, in MM-V, each child needed to read and physically click on a correct answer arrived at through discussion or through some leader-follower system. This may have supported visual recall, reinforced by discussion and tangibility of the click. The observations suggest a mix of factors that could have led to the generally higher word retention in MM-V.

Miscellaneous Observations: Activity Log Summary

The software collected detailed logs, allowing measures like *Activity* (rate of mouse events), *Clicks* (rate of clicks, including random clicks), *Accuracy* (Ratio of correct to in-correct clicks). We performed only a preliminary analysis, but such data is worth investigating, as we could see indications of some interesting trends, such as (Figure 10):

- At the start of a session, the values for activity (green line) and click rate (blue values) fall as users learn the game. From the graph, we can see the fall is slower for SS than MM-R, matching behavioral observations.
- In MM-R mode, there is a clear increase (blue line) of the rate of clicking, as a session goes on. This could tie in with the heightening competition and engagement.

CONCLUSION

Although the importance of engagement and collaboration in computer-aided learning has been pointed out in the past, here we note the significance of these in designs of multiple mouse modes of interaction. We also see that the effect of competition in such settings can promote engagement, yet still affect learning negatively. The mixed-methods approach of using qualitative and quantitative research brought out nuanced views of both engagement and collaboration in perspectives of shared use. Of particular note is the idea of engagement, which in past literature was defined in some terms to be synonymous with 'attention' or 'lack of off-task behavior'. We see here, that simple attention to on-screen content does not necessarily mean better learning. This study also provides statistically significant findings to support the hypotheses of collaboration leading to better learning, in terms of quantifiable and tangible educational value attributable to the use of multiple mice. This work is distinct in another important way - our experiments were directed not towards scenarios where collaboration was encouraged for the inherent value of learning to collaborate as such, but where it was a necessity out of resource shortage.

We saw that the two modes of multiple mice explored, MM-V-collaborative and MM-R-competitive, show the striking difference in learning outcomes (for boys) and user behavior that is possible due to variations in interaction designs for multiple-mice, even though both were designed for the same end task. We also saw how saliently gender impacts learning, especially as in shared-use settings, social aspects of the experience gain importance.

Overall, Multimouse configurations were found to be on par with single-user configurations in terms of actual words learned, suggesting that PC value for certain kinds of education can be inexpensively multiplied through shared use. The outcomes of this research suggest that multiple-mice as a teaching technology needs to be brought into greater attention, especially in primary education. In future work, we anticipate further designs of software developed for Multimouse, and tests of higher-level cognitive tasks.

ACKNOWLEDGMENTS

We are deeply indebted to the Azim Premji Foundation for their support, specifically Prachi Gaur, Ramya, S.Santhosh Kumar, Sukumar Anikar and Sunita Amenchelra. Thanks too to the RAs – Ajay Anand, Arjun V., Gautam Prakash, Girish P.S., Karthik Kambatla, Naga Yasodhar, Naveen M., Rajesh T.M., Shravan Kumar and Sireesh S.

REFERENCES

1. Abnett, C., Stanton, D., Neale, H and O'Malley. The effect of multiple input devices on collaboration and gender issues. Proc. EuroCSCL (2001), 29-36.
2. Benford, S., Bederson, B. B., Åkesson, K., Bayon, V., Druin, A., Hansson, P., Hourcade, J. P., Ingram, R., Neale, H., O'Malley, C., Simsarian, K. T., Stanton, D., Sundblad, Y., and Taxén, G. Designing storytelling technologies to encouraging collaboration between young children. Proc. CHI '00, ACM Press (2000), 556-563.
3. Bier E.A, and Freeman S.. MMM: A User Interface Architecture for Shared Editors on a Single Screen. Proc. UIST (1991), 79-86.
4. Bloom B.S. and David R.K., *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I*. Longmans, Green (1956), NY.
5. Bricker, L. *Cooperatively Controlled Objects in Support of Collaboration.* Ph.D. Thesis (1998), Univ. of Washington.
6. Bricker, L.J., Tanimoto, S.L., Rothenberg, A.I., Hutama, D.C., and Wong, T.H. Multiplayer Activities that Develop Mathematical Coordination. Proc. CSCL (1995), 32-39.
7. Druin, A., Revelle, G., Bederson, B. B., Hourcade, J. P., Farber, A., Lee, J., & Campbell, D. A Collaborative Digital Library for Children: A Descriptive Study of Children's Collaborative Behaviors. J. Computer-Assisted Learning 19, 2 (2003), 239-248.
8. Ebbinghaus, H. *Memory: A Contributon to Experimental Psychology.* Dover Publications (1964), New York, NY.
9. Inkpen K., Booth K.S., Gribble S.D., and Klawe M. Give and Take: Children Collaborating on One Computer. Short papers in CHI' 95, ACM Press (1995), 258-259.
10. Inkpen, K., Booth, K. S., Klawe, M., and Upitis, R. Playing together beats playing apart, especially for girls. Proc. CSCL (1995), 177-181.
11. Inkpen, K., Ho-Ching, W., Kuederle, O., Scott, S.D., & Shoemaker, G. This is fun! We're all best friends and we're all playing.: Supporting children's synchronous collaboration. Proc. CSCL (1999), 252-259.
12. Inkpen, K., McGrenere, J., Booth, K. S., and Klawe, M. The effect of turn-taking protocols on children's learning in mouse-driven collaborative environments. Proc. GI (1997), 138-145.
13. Kam, M. Notes Towards a Framework for Designing Mobile Games for Children in the Developing World to Learn English as a Second Language in Out-of-School Settings. Proc. DREAM 2006 (2006).
14. Mitra, S., Dangwal, R., Chatterjee, S, Jha, S., Bisht, R.S., and Kapur, P. Acquisition of computing literacy on shared public computers: Children and the 'Hole in the Wall'. Australasian Journal of Educational Technology 21, 3 (2005), 407-426.
15. Myers, B.A., Stiehl, H, and Gargiulo, R . Collaboration using multiple PDAs connected to a PC. Proc. CSCW (1998), 285-294.
16. O'Leary-Kelly, A., Mortochhio, J, Frink, D. 1994. A Review of the Influence of Group Goals on Group Performance. Academy of Management Journal 37, 5(1994), 1285-1301.
17. Oppenheimer, T. *The Flickering Mind.* Random House Publishers (2003).
18. Pal, J., Pawar, U., Brewer, E., and Toyama, K. The case for multi-user design for computer aided learning in developing regions. Proc. WWW 2006, ACM Press (2006), 781-789.
19. Pawar, U. S., Pal, J., and Toyama, K. Multiple mice for computers in education in developing countries. Proc. IEEE/ACM ICTD (2006), 64-71.
20. Pederson, E.R., K. McCall, T. P. Moran, and F. G. Halasz. Tivoli: An electronic whiteboard for informal workgroup meetings. Proc. CHI '93, ACM Press (2003), 391–398.
21. Scott, S., Mandryk, R., and Inkpen, K. Understanding Children's Interactions in Synchronous Shared Environments. Proc. CSCL (2002), 333-341.
22. Scott, S.D., Mandryk, R.L., and Inkpen, K.M. Understanding Children's Collaborative Interactions in Shared Environments. J. CAL 19, 2 (2003), 220-228.
23. Shen, C. Multi-user interface and interactions on direct-touch horizontal surfaces: collaborative tabletop research at MERL. Proc. IEEE TableTop (2006), 53-56.
24. Shoemaker G.B.D. Single Display Groupware research in the year 2000. TR2001-1 (2001), Simon Fraser University.
25. Siriginidi S.R. Bridging digital divide: Efforts in India, Telematics and Informatics 22, 4 (2006), 361-375.
26. Stanton D., and Neale H.R. The Effects of Multiple Mice on Children's talk and Interaction. J. Computer Assisted Learning 19, 2 (2003), 220-228.
27. Stanton, D., Neale, H. and Bayon, V. Interfaces to support children's co-present collaboration: multiple mice and tangible technologies. Proc. CSCL (2002), 342-352.
28. Stewart J., Bederson B., and Druin A. Single Display Groupware: A model for Co-present Collaboration. Proc. CHI '99, ACM Press (1999), 286-293.
29. Stewart J., Raybourn E., Bederson B., and Druin A. When Two Hands are Better Than One: Enhancing Collaboration Using Single Display Groupware. Ext.Abst. CHI '98, ACM Press (1998), 287-288.