A Mischief of Mice: Examining Children's Performance in Single Display Groupware Systems with 1 to 32 Mice

Neema Moraveji[1], **Kori Inkpen**[2], **Ed Cutrell**[2], **Ravin Balakrishnan**[3]
neema@moraveji.org, kori@microsoft.com, cutrell@microsoft.com, ravin@dgp.toronto.edu
[1]Learning Sciences and Technology Design, School of Education, Stanford University, Stanford, CA
[2]VIBE Research Group, Microsoft Research, Redmond, WA
[3]Department of Computer Science, University of Toronto, Toronto, Canada

ABSTRACT

Mischief is a system for classroom interaction that allows multiple children to use individual mice and cursors to interact with a single large display [20]. While the system can support large groups of children, it is unclear how children's performance is affected as group size increases. We explore this question via a study involving two tasks, with children working in group sizes ranging from 1 to 32. The first required reciprocal selection of two on-screen targets, resembling a "swarm" pointing scenario that might be used in educational applications. The second, a more temporally and spatially distributed pointing task, had children entering different words by selecting characters on an on-screen keyboard. Results indicate that performance is significantly affected by group size only when targets are small. Further, group size had a smaller effect when pointing was spatially and temporally distributed than when everyone was concurrently aiming at the same targets.

Author Keywords

Single display groupware, children, large displays, mice.

ACM Classification Keywords

H5.m. Information interfaces and presentation: Misc.

INTRODUCTION

Single-display groupware systems have been explored to support small group educational activities [3, 10, 21, 28]. Recently, these systems have been extended to support entire classrooms of students using a single large display [20]. For classroom use, there are several advantages to a shared-display design over more conventional systems. Costs are typically lower since each student does not require an individual computer, and new cooperative and competitive tasks can be performed on a shared screen in ways that may better engage students.

Typically, single-display groupware systems have supported small groups of 2-5 users [21, 28], but recently Moraveji *et al.* [20] demonstrated scenarios with up to 18 children simultaneously using a single large display. Mice, unlike audience response systems ('clickers'), calculators, mobile phones, PDAs, display continuous input to the whole class in real-time. While the idea of an entire class of students simultaneously accomplishing a task is interesting, and observations from Moraveji *et al.* [20] suggest that children are able to use such systems, their experience also shows that it can be somewhat chaotic. As the number of children and cursors grows, problems of visual clutter and occlusion, cursor differentiation, and visual and auditory feedback may occur. To date, however, there is no empirical data on how effectively large groups of children can interact with a single large display system using multiple mice. Our goal is to investigate how these systems scale to large groups in order to better guide future designs that might make such systems more usable.

This research is timely as computing facilities are being installed in classrooms at a fast pace. Low per-unit cost computers such as the One Laptop per Child initiative and Intel's Classmate PC are promising educational gains at relatively high overall costs to countries worldwide. In evaluating the feasibility of such efforts, the potential of low overall cost systems should also be considered, and empirical data on their usability is crucial in this regard.

Although classrooms are the most obvious environment for large numbers of simultaneous users, other domains may also benefit from such systems. Large groups frequently collaborate on shared representations in disaster recovery planning, team brainstorming, and geo-visualization.

This paper describes a study which evaluated children's performance in target acquisition tasks on a single shared large display in group sizes of 1, 4, 8, 16 and 32 (Figure 1), and discusses the implications of the results to the design of shared-display groupware for classroom-wide interaction.

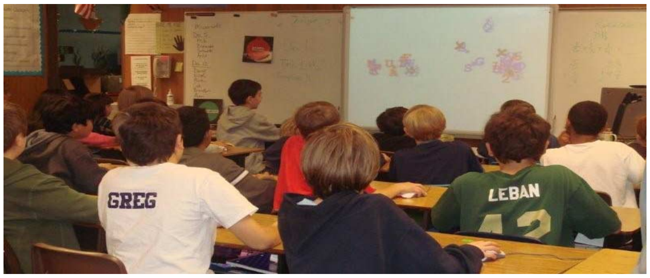

Figure 1. Children simultaneously performing a target acquisition task using multiple mice and a large shared display

CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.

RELATED WORK

Previous research has demonstrated advantages of single-display groupware [28] where one physical display is shared amongst a group of co-present users, each with their own input device(s). This model has frequently been applied to educational uses on desktop computers because of the collaborative affordances of such systems and has been shown to provide positive educational gains [3, 10, 11, 26]. Another benefit of such systems is to lower the cost of computer access for students in developing regions [21].

Traditionally, mice have been used as input devices for single-display groupware systems. Numerous studies have evaluated the performance of children using mice and other input devices [2, 8, 12, 14, 15, 16], although not necessarily within single-display groupware systems per se. Work by Hourcade *et al.* [8] showed that children's age is a significant factor in pointing task performance where younger children (4 and 5 year olds) perform significantly worse than older children (13 year olds) and adults. Target size can also significantly impact children's performance, with small targets being more difficult for them than for adults. In addition, children seem to make significantly more, and less accurate, sub-movements when acquiring targets [9]. The fine motor skills required for the homing phase of target selection are more problematic for children. Hourcade *et al.* [9] also summarized the psychology literature on reaction times in children [24, 27, 29], indicating that younger children will show greater variability. This is supported by the model human processor [1] which suggests that children's performance with pointing devices should be lower than that of adults. In summary, much of the literature in evaluating input devices with children has focused on how well children can use various input devices and understanding their movement characteristics when using those devices. In contrast, our focus is on how children's performance changes as a function of group size in large scale single-display groupware environments.

The effectiveness of an interface or interaction style is particularly important for educational activities. Inkpen's [12] work on comparing drag-and-drop versus point-and click interactions by children demonstrated that a less effective technique like drag-and-drop can significantly influence children's performance in a task. While that research was conducted in a single-user scenario, it is likely that such differences in techniques could have an even greater impact in a shared-display environment, particularly with a large number of children working concurrently.

Russell *et al.* [22] showed how a larger display can improve collaboration in a single-display groupware environment. Moraveji *et al.* [20] described a groupware system with a large projected display meant for classrooms that accommodates relatively large numbers of simultaneous users. Although this work introduced some methods of supporting input from scores of simultaneous users, it did not go on to study the performance impact that might be incurred when large numbers of simultaneous users are operating cursors on screen. In particular, their cursors were large glyphs which enabled the children to easily identify their cursors but these might impact performance due to extra visual clutter as the number of cursors increase.

The study of simultaneous use of an application by multiple parties has also been studied in distributed groupware [5, 6] and tabletop groupware [23, 25]. In distributed groupware, Gutwin and Greenberg have made significant contributions in terms of usability and interaction design, and provide a nice summary of the literature in groupware usability [6]. Of particular relevance to our work is Greenberg *et al.*'s [5] exploration of how multiple pointers in a shared workspace (with some participants geographically remote) can convey more information by mapping them to underlying objects in the scene rather than to Cartesian screen coordinates. They also showed that adding semantic information to the pointers could improve user's awareness of what was happening in the shared space with minimal impact on screen real estate.

The effect of group size in single-display groupware systems has been explored by Ryall *et al.* [23] who examined the effect of table size and group size on task performance in tabletop displays. They found that group size, but not table size, affected task performance. Although they only examined groups of 2 and 4 people and did not compare to single user performance, this work does highlight some of the interface issues that designers should be aware of when supporting larger groups of users in groupware systems. In particular, they indicate that additional displays might be required for larger groups in order to mitigate issues with clutter and collisions due to overlapping human input that is particularly acute in a direct input tabletop environment.

One challenge in building such systems for large numbers of simultaneous users is the lack of software infrastructure in standard interface toolkits for supporting multiple concurrent input devices. Several research efforts have attempted to address this, including the Dynamo system by Izadi *et al.* [13] and Pawar *et al.*'s Multimouse infrastructure [21]. Our current research builds upon the infrastructure in recent work by Moraveji *et al.* [20].

In summary, our literature review indicates that the effect of group size on user performance in single-display groupware situations has not been adequately explored. This is likely because most of these systems support a relatively small number of users and cursors; as a result, they show little performance degradation due to group size. With the recent interest in deploying such systems with many more users [20], it is important that we understand how performance is impacted by increasing numbers of users and on-screen cursors. Our present work aims to provide some empirical data to inform the design of input mechanisms for large-display groupware systems with large numbers of users, with a particular focus on children's use of such systems.

STUDY

Goals

Our aim was to examine children's performance in fundamental pointing and selection tasks using multiple mice concurrently on a single shared large-screen display. More specifically, we were interested in how performance (in terms of task completion times and accuracy) changes as the number of children and mice increase. From this data, we hoped to derive some design guidance for such systems. For example, we wanted to see if there was a "sweet spot" in terms of the number of children who can concurrently use large-screen single-display groupware systems, and we hoped to determine a minimum size for objects that children may have to acquire with their cursors. We were also interested in performance differences associated with tasks that result in temporally concurrent selection of a single target by all children versus temporally semi-concurrent selection of different spatially distributed targets by each child. To achieve these goals, we needed to design tasks that would have broad implications for designers of a variety of educational activities.

Participants

Because our domain of interest is educational, we ran our study with school-age children. 40 children, with normal color vision, aged 10-12, 15 female and 25 male, from a public elementary school in the Northwest US volunteered for the study. When asked to rate how often they used computers on a discrete scale (1 being "rarely", 4 being "every day"), the average response was 3.5. 37 children operated mice with their right hand and 3 used either hand. Parental consent was obtained and a gratuity was provided to the school. The children were not directly compensated.

Setting

The study was administered in two school classrooms. Each was equipped with individual student desks, 26" wide x 20" deep, and a USB 2.0 Microsoft IntelliMouse on each desk. The desks were arranged in four rows of eight desks. Each room had a 1024 x 768 pixel resolution projector mounted in the ceiling and projected a 64" x 48" display. Figure 2 illustrates the setup. The first row of desks was positioned approximately 58" away from the screen and each subsequent row was approximately 40" behind the previous. Each room had a Pentium 4 Windows XP computer running the study software. The mice were connected to this computer via USB hubs. Custom C# software, with low-level WinAPI calls to enumerate USB mice, administered the study stimuli and logged all mouse events while ensuring there was no noticeable latency.

Two classrooms enabled two sessions to run concurrently. Both classrooms had the same configuration of desks and projection equipment. A third classroom was used as a holding room where students would go while they waited for their turn. In addition to the researchers administering the study, teachers from the school were present to supervise the children, both in the classrooms being used for the study as well as in the holding room.

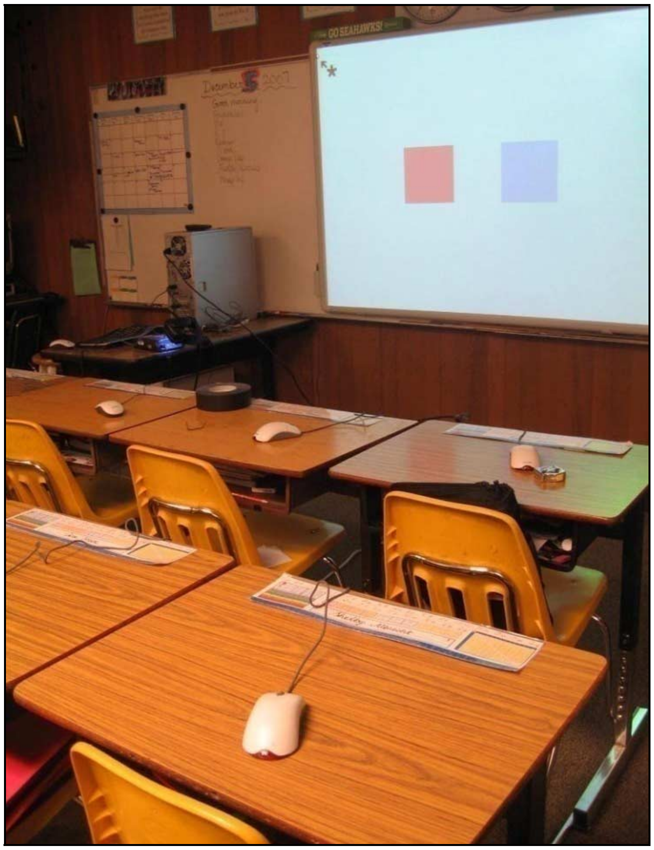

Figure 2. Study setting. Each classroom had a 64"x48" projected display in front of 4 rows of 8 desks, with a mouse on each desk.

Cursors

One challenge in supporting a large number of simultaneous users on a single display is the design of cursors such that they are easily identifiable by users while minimizing visual clutter. The detailed investigation likely required to develop an optimal cursor representation for such environments is beyond the scope of this paper; however, in an attempt to address this issue, we designed cursors with two differentiating visual features: character and arrow direction (Figure 3).

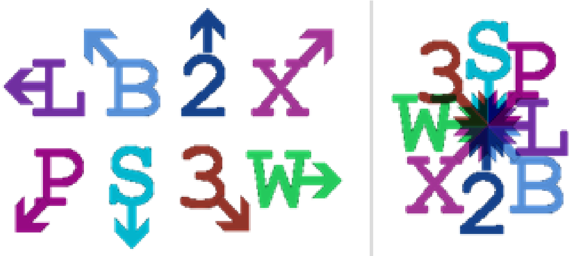

Figure 3. Cursor design. (left) Eight example cursors uniquely identifiable by different characters, with arrows pointing in different cardinal directions. (right) The eight cursors pointing to the same target result in overlapping arrows but not cursor characters, maintaining cursor identification. With more than eight cursors, more visual overlap of the characters will occur.

The characters ranged from A-Z and 2-9, omitting 0-1 because '0' (zero) could be confused with the letter 'O' and '1' with a lower case 'L'. Using different characters per cursor is intended to allow each user to quickly identify their cursor. To avoid the visual clumping of cursors that would result if all cursors had pointers aimed towards the ubiquitous upper left corner, we had the cursor arrows point in one of eight cardinal directions (Figure 3, left). The hotspot (i.e., cursor pixel which needed to be atop the target to count as a click) was at the tip of the arrow on each cursor. In our cursor-to-arrow direction assignment scheme, if there were 8 cursors, each would be assigned a different arrow direction (Figure 3, right). For more than 8 cursors, we assigned additional cursors to the 8 directions in turn, resulting in some visual overlap in the cursors' characters when pointing at the same target. For example, if there were 16 cursors, there would be visual overlap between pairs of cursors; with 32 cursors, overlap between sets of 4 cursors. When cursors overlapped, we used alpha blending with pre-randomized z-ordering to display them on screen. The size of each cursor glyph (character + arrow) was 48x91 pixels.

Tasks

Two tasks were used in this study. In the first task we modeled a "swarm" pointing scenario, in which children attempt to acquire the same targets at about the same time. Such scenarios could occur in the educational usage environments we are interested in and would seem to provide significant challenges for the children. Toward this end, we developed a reciprocal pointing task that borrows from the Fitts' paradigm [4, 17] to track how performance changes as the number of users and cursors increase for a variety of target distances and sizes. However, our task departed from the standard Fitts task for several reasons. First, the Fitts' analysis does not account for multiple on-screen cursors that may distract users. Second, visual feedback on targets is problematic because users may confuse feedback with each other's input. Finally, we expect that children will click on a target at about the same time, and then move on to the next target *en masse*, creating visual interference around the targets. Of course, it is possible that one or more children might lag behind the others and get somewhat "out of phase" in their pointing.

The second task explores a different scenario, where children are pointing semi-concurrently at different targets that are spatially distributed across the screen. While we could have designed an abstract task to test this scenario, we decided to use a more ecologically valid text entry task that was based on the applications explored by Moraveji *et al*. [20] in their Mischief system used in schools in developing regions. We designed a text entry task in which each child had to spell out a variety of words using an A-Z on-screen keyboard. In this task, each target (keyboard letter) was the same size, and successful performance required a chain of several accurate button acquisitions (e.g., to correctly spell the target word).

Each task lasted as long as the children needed; however, the children were instructed to complete the task as quickly as possible while making as few mistakes as possible. Only the left mouse button was used for clicking; middle and right mouse button input was ignored as it has been shown that children often accidentally click these other buttons [7].

Task 1: Reciprocal Pointing Task

In this task, each child was required to alternately click on two square targets displayed on-screen (Figure 4). The square on the right was blue and the square on the left was red. As the children progressed through the task, the color of their cursors changed to indicate which target they should click next. The cursor colors were similar to the targets but not identical to avoid a camouflage effect. In most reciprocal pointing studies reported in the literature, feedback on successful/unsuccessful acquisition is often provided visually on the target or via auditory cues. We instead relied on cursor color change for feedback because each cursor is unique to a particular child, whereas the two targets are common to all children in this environment. Auditory feedback via headphones would have been impractical for the large number of children in our study. It is important to note that each child had to successfully click on the indicated target before they could proceed to the next target. This effectively prevented the children from "racing" through the experiment by clicking anywhere.

Independent variables were the Movement Amplitude (or distance) between the target centers and Target Width. Movement Amplitude was either 300 or 700 pixels (actual onscreen amplitude 18.75" or 43.75"). Target Width was 30, 40, 80, or 160 pixels (actual onscreen size 1.9", 2.5", 5" or 10"). For each Movement Amplitude and Target Width combination, each child completed 10 target acquisitions in a row, after which their cursor turned gray and was disabled such that clicks were no longer registered. After all the children completed the task, the screen was cleared and they were given a short break in preparation for the next Movement Amplitude and Target Width combination.

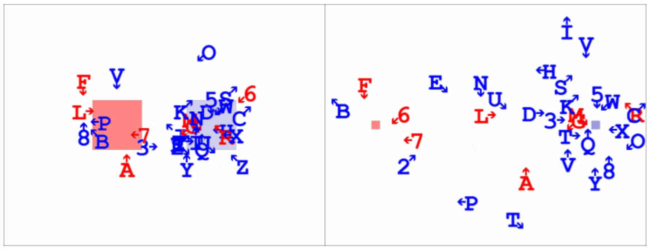

Figure 4. Stimuli for the reciprocal pointing task. (left) Stimuli with Movement Amplitude = 300 pixels and Target Width = 160 pixels. (right) Stimuli with Movement Amplitude = 700 pixels and Target Width = 30 pixels. 32 cursors are shown to illustrate the maximum density of cursors in the study.

Task 2: Text Entry Task

In this task, each child was required to enter five-letter words by clicking on the relevant characters on an A-Z on-screen keyboard. Each letter on the keyboard was represented as a 95 x 55 pixel target. Each child was given eight words to enter in turn. All children entered the same set of words, but in different fully counterbalanced orders depending on their assigned group as discussed in the procedure section later. The words were selected from a standard linguistic database[1], with a Kucera & Francis frequency of 10 [17], and filtered to ensure they would be understood by the 10-12 year old children in this study.

Each child was assigned to a quadrant of the screen which displayed the word they were required to enter. Within a quadrant, each child was assigned a specific text display box within which the characters they selected from the on-screen keyboard would appear. Their assigned cursor's character was displayed next to their text entry box so that they could easily identify their box from the rest. No additional feedback was provided. As an example, Figure 5 shows the stimuli at the start of a trial with 32 cursors: the children with cursors K-R had to enter the word SWEAR by selecting the relevant characters from the onscreen keyboard in the middle of the display, and the other children similarly entered the words STOLE, STEAK, and ROAST respectively. Figure 6 illustrates how the display looked as the trial was partially complete. When different numbers of cursors were tested, the number of text display boxes shown per quadrant was adjusted in a balanced manner across quadrants.

The children could only input 5 characters (i.e., the length of the given word). Any additional characters were ignored. Incorrect characters could be entered, and more than one could be entered without immediate correction (up to the maximum of 5 characters), but ultimately these had to be corrected by selecting the "Del" key on the onscreen keyboard and reentering the correct characters. After each child correctly entered their given word, a large red "check mark" was displayed on their output panel, barring them from further character input. Their mouse cursor also turned gray to reinforce the fact that they had completed that trial.

Procedure

The study was conducted after school hours on two separate days in December 2007, one week apart. Eight children participated on day 1, and a different set of 32 children on day 2. At the beginning of each day, the children were asked to complete a short background questionnaire that gathered background demographic information, asked to complete a short color blindness evaluation, and asked how often they use computers. Following this, all of the students went into the holding room and were called into the study rooms when it was their time to participate.

All children did the reciprocal pointing task first, followed by the text entry task. For each task, they first did some practice trials to familiarize themselves with the task. Then, they repeated the trials in each task in different group sizes as shown in Table 1.

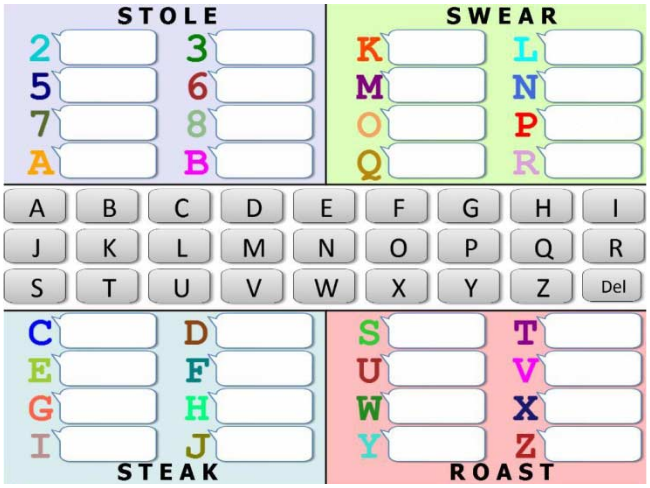

Figure 5. Stimuli for text entry task at start of trial. In this example, 32 children are divided into 4 groups of 8. Each group is assigned one quadrant of the screen, with the word to be entered displayed on top of the quadrant. Each child's cursor character is displayed next to their text display box so they could easily identify their own box.

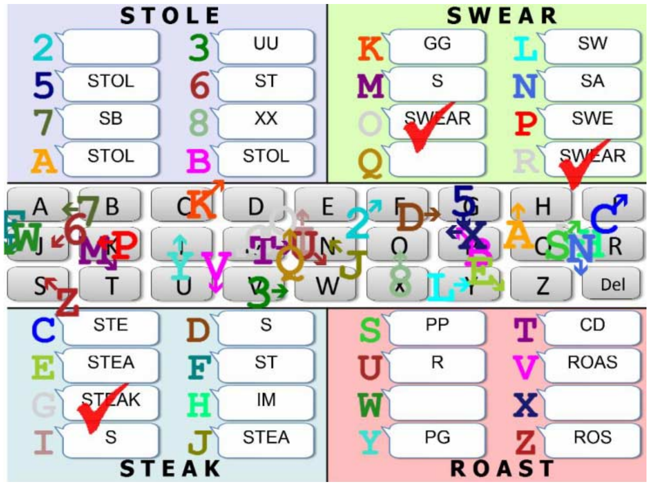

Figure 6. Stimuli for text entry task part-way through a trial. Some children have partially entered their words; three have completed entry (cursor turns gray) and a red check-mark is displayed next to the relevant text display boxes to confirm completion.

Table 1. Task and group size schedule across days.

Day 1 Sessions	Day 2 Sessions
Practice Task 1	Practice Task 1
Task 1, Group Size 1	Task 1, Group Size 8
Task 1, Group Size 4	Task 1, Group Size 16
Task 1, Group Size 8	Task 1, Group Size 32
Practice Task 2	Practice Task 2
Task 2, Group Size 4	Task 2, Group Size 16
Task 2, Group Size 8	Task 2, Group Size 32

---

[1] http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm

On Day 1, for Group Size 1, each of the 8 children did the reciprocal task individually, with a single mouse on an unshared display. Then for Group Size 4, the 8 children were divided into two groups of 4, and did the tasks with 4 cursors on one shared display. We ran both groups concurrently in the two classrooms. Finally, for Group Size 8, all 8 children did the tasks with 8 cursors on one shared display. The same assignment of children to groups was used subsequently for the text entry task. Note that we did not test the text entry task with Group Size 1 as we felt it would have been effectively similar to the reciprocal task since there would have been no issue of multiple cursors trying to concurrently acquire the same target.

On Day 2, for Group Size 8, the 32 children were divided into four groups of 8, and did the reciprocal pointing task with 8 cursors on one shared display. We ran two groups concurrently in the two classrooms. For Group Size 16, two of the earlier groups of 8 were merged to form groups of 16, and did the task with 16 cursors on one shared display. We again ran the two groups concurrently in the two classrooms. Finally, for Group Size 32, all 32 children did the task with 32 cursors on one shared display. This same assignment of children to groups was used subsequently for the text entry task. Note that the Group Size 8 condition was common to both days for the reciprocal pointing task mainly because we did not want to start the Day 2 children off with the likely harder Group Size 16 condition.

On each day, all children kept the same seat position throughout all tests, so they were in the same physical place relative to the screen for each of the 3 group size conditions (albeit sometimes in a different room).

In summary, for the reciprocal pointing tasks, each child completed 10 target acquisitions for each combination of Group Size x Target Width x Target Amplitude for a total of 10 x 3 x 4 x 2 = 240 target acquisitions per child. On Day 1, this resulted in 8 x 240 = 1920 target acquisitions across all children; on Day 2, 32 x 240 = 7680 target acquisitions.

For the text entry task, each child entered 8 words for each Group Size for a total of 8 x 2 = 16 words per child. On Day 1, this resulted in 8 x 16 = 128 words entered by all children; on Day 2, 32 x 16 = 512 words.

The children were free to talk amongst themselves during the study, as they might do in a real-world scenario. Breaks were given between each condition. At the end of each task the children were asked to rate how easy the activity was on a 5 point scale. Freeform comments were also solicited.

Hypotheses

We expected to observe the following:

H1: Group size would significantly affect individual target acquisition error rates.

H2: Group size would significantly affect individual target acquisition speed.

H3: Group size will have less of an effect on performance in the text entry task compared to the reciprocal pointing task, since the children would not all be trying to select the same target concurrently.

RESULTS

We analyzed the data from Day 1 and Day 2 separately. As such, all analyses were conducted within-participant. Because of the complexity and number of analyses performed, all results are reported with a conservative significance threshold of α=0.01. Bonferroni corrections were used on all post-hoc t-tests.

Reciprocal Pointing Task

We calculated two dependent measures for this task: *movement time* and *error rate*. Movement time was computed as the interval between when a cursor's color changed to indicate the next target to select and the button press that successfully selected that target. This measure thus includes the time to correct for errors. Error rate was computed as the number of targets that were not selected at the first attempt.

When performed by a single person on a non-shared display, our reciprocal pointing task becomes essentially a standard Fitts' Law [4, 17] task. However, as we anticipated when designing the study, the multi-user multi-cursor nature of this task results in fundamental differences from the single user case, in terms of difficulty in providing visual feedback on the target, visual clutter due to overlapping cursors, and visual distraction due to multiple cursors. Not surprisingly, these differences seem to manifest themselves in terms of high error rates for the conditions with a large number of cursors, as well as large differences in performance across Target Widths as compared to Movement Amplitude (whereas Fitts' Index of Difficulty would indicate that Width and Amplitude would have similar impact on performance). Also, our experimental design had a relatively low number of observations per condition. Taken as a whole, these issues confirm our expectation that a standard Fitts' analysis would not be appropriate, Therefore we report analyses of movement time and error rate as a function of Group Size, Target Width, and Movement Amplitude. To minimize the skewing commonly seen in response time data, we looked at the median response time for each child in the last 8 movements of each condition (the first two were discarded). For each day, we performed a 3 (Group Size) X 4 (Target Width) x 2 (Movement Amplitude) within subjects ANOVA. On day 1, Group Size = 1, 4, and 8; on day 2, Group Size = 8, 16, and 32 children.

Day 1 Movement Time and Error Rate

Day 1 had group sizes of 1, 4 & 8. For movement time, we found no significant main effect for group size. We did find a significant effect for Target Width ($F_{3,21}$=184.8, p<0.001) and Movement Amplitude ($F_{1,7}$=370.2, p<0.001). As expected, movement time increased with increasing Movement Amplitude and decreasing Target Width (Figure 7, top row). We also found a significant interaction between

Group Size and Target Width ($F_{6,42}=6.52$, $p<0.001$). Further analyses revealed no significant pair-wise differences between Group Sizes for each Target Width.

We found no significant effects for any factor for error rates on Day 1 (Figure 7, bottom row).

Day 2 Movement Time and Error Rate

On Day 2, we looked at group sizes of 8, 16 and 32 children. As in Day 1, we found significant main effects on movement time for Target Width ($F_{3,81}=261.54$, $p<0.001$), and Movement Amplitude ($F_{1,27}=90.29$, $p<0.001$). We also found a significant main effect for Group Size, ($F_{2,54}=36.41$, $p<0.001$) on movement time. *Post hoc* t-tests indicated that the Group Size of 32 was significantly slower than either 16 or 8, $p<0.001$ (Figure 7, top row). There was also a significant interaction between Group Size and Target Width ($F_{6,162}=17.98$, $p<0.001$). We performed a series of 12 pair-wise t-tests between each Group Size at each Target Width to further investigate specific differences. We found three significant differences: the Group Size of 32 was significantly slower than either 16 or 8 for Target Width = 30, and Group Size of 32 was significantly slower than 16 for Target Width = 40 (Figure 7, top right). These findings suggest that movement times are only seriously impacted by Group Size when targets are smaller.

We found significant main effects on error rate for Target Width ($F_{3,81}=115.16$, $p<0.001$) and Group Size ($F_{2,54}=35.23$, $p<0.001$). *Post hoc* t-tests for Target Width showed no significant differences between the two largest Target Widths, but error rates increased significantly for each successively smaller width, $p<0.001$ (Figure 7, bottom right). *Post hoc* t-tests for Group Size revealed significantly larger error rates for the group of 32 children than either groups of 16 or 8, $p<0.001$ (Figure 7, bottom row).

As with movement time, we found a significant interaction between Group Size and Target Width on error rate, ($F_{6,162}=5.10$, $p<0.001$). We performed a series of 12 pair-wise t-tests between each Group Size at each Target Width to further investigate specific differences. We found five significant differences. At Width=30 and 40 the Group Size of 32 had significantly larger error rates than either groups of 16 or 8. At Width=80, the Group Size of 32 had significantly larger error rates than groups of 8 (Figure 7, bottom right). As with the analyses of movement time, these findings suggest that error rates for smaller targets are disproportionately affected by larger groups.

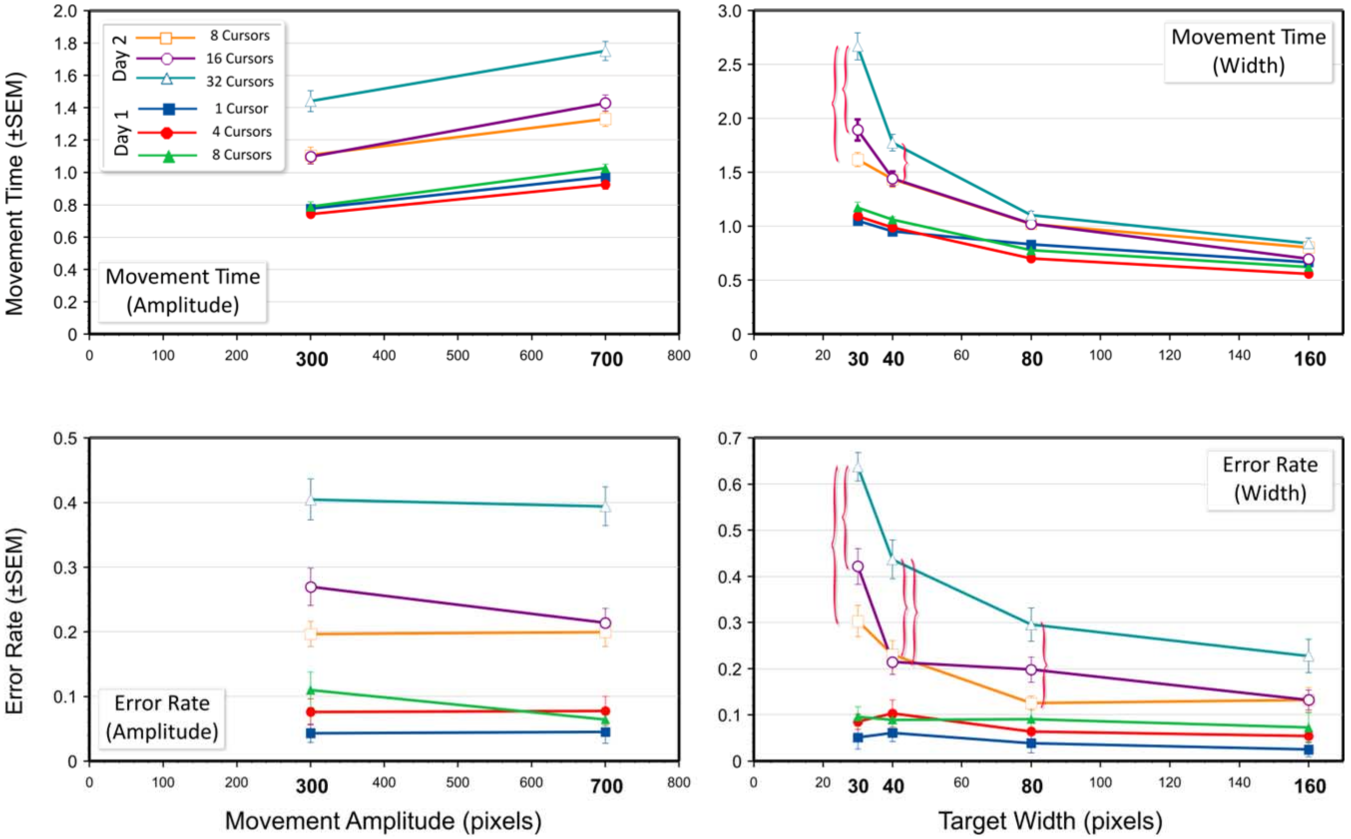

Figure 7. Results for reciprocal pointing task. (top row) Movement time in seconds, by Movement Amplitude and Target Width. (bottom row) Error rate as percentage, by Movement Amplitude and Target Width. The curly braces on the right column graphs highlight the primary significant differences found in our analysis.

Text Entry Task

We calculated two dependent measures for this task: *movement time* and *error rate*. Similar to the reciprocal pointing task, we focus on individual movement times for acquiring keys on the virtual keyboard. Unlike the reciprocal pointing task, where errors simply meant that the child kept trying until they selected the target successfully, errors in the text typing task often resulted in the incorrect selection of an unintended letter which had to be subsequently corrected. To account for these incorrect letters, our response time measure was a normalized measure of the average movement time per correct character (accounting for correction characters) for each word. For example, if the child entered "ROAD<Del>ST" for the word "ROAST", the normalized time per correct character was calculated as the total time to enter "ROAD<Del>ST" divided by 5 (i.e., the number of correct characters). Error rates were calculated as the total number of errors committed for each word. Since each child completed 8 words in each condition, we performed a 2 (Group Size) X 8 (Word Order) within subjects repeated measures ANOVA for normalized movement time per word and for number of errors per word. As in the reciprocal typing task, the data was analyzed separately for each day.

In all four analyses (normalized movement time and error rate for each day), only one significant effect was found: there was a significant effect for Word Order for Day 2 for normalized movement time ($F_{7,189}=4.52$, $p<0.001$). Unsurprisingly, the normalized movement time for the first word was slower than subsequent words, although no post-hoc pair-wise comparisons were significant. Although there were no significant differences found for Group Size, the data does show a trend of increasing movement times and error rates as Group Size increased, particularly moving from 16 to 32 children (Figure 8). However, note that neither the differences in movement time nor error rate come anywhere close to the differences seen in the reciprocal pointing task. In addition, even at the largest Group Size, the error rate is very low, at about one error for every two words completed (Figure 9). Also note that the magnitude of the movement times were significantly higher in this task compared to the reciprocal pointing task, likely due to the higher cognitive effort required.

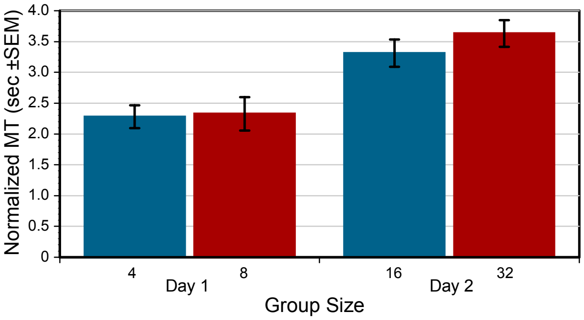

Figure 8. Normalized movement time per correct character for the text entry task.

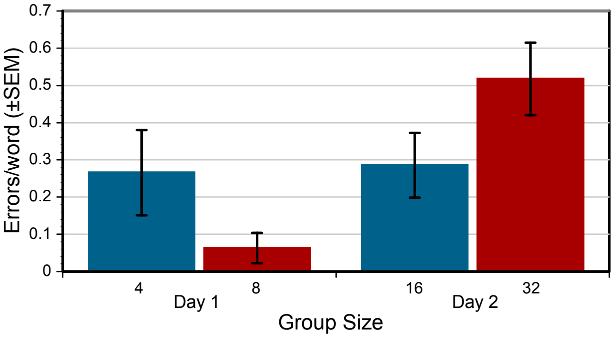

Figure 9. Error rate (number of errors per word) for the text entry task.

Survey Results

At the end of each task the children were asked to rate how easy the activity was on a 5 point scale, where 5 represented very easy and 1 represented very difficult. The results are summarized in Table 2. Not surprisingly, the children's mean ranks went down as the Group Size increased; however, a Wilcoxon paired test showed that these differences were only significant for the children who participated on day 2 ($p<0.00625$).

Table 2. Mean ranks for difficulty of task. 1 = very difficult, 5 = very easy.

	Day 1			Day 2		
	Group Size			Group Size		
	1	4	8	8	16	32
Reciprocal Pointing	4.50	3.81	3.13	3.98	3.26	2.74
Text Entry	n/a	3.75	3.81	n/a	4.50	3.87

When asked whether they would like to use a system like Mischief at school, most of the children responded favorably except for three who felt that the system was too confusing. Others recognized that the system was best when there were a small number of cursors. Some comments included:

"It was FUN!!!",

"I "heart" this system!!!"

"I think this system should be used for educating. I think it is AWESOME!"

"It is fun to use but sort of confusing."

"I like the system, but when too many people click on one thing it gets confusing."

"I think it's really fun but only when there's a limited amount, or it gets confusing, but when you do it a few times, you get used to it!"

DISCUSSION

Our results demonstrate that for large targets (e.g. 160 pixels) group size does not make a significant difference for either movement time or error rate. As targets decrease in size to a width of 80 pixels, the effect of group size begins to have an impact on error rate. However, we do not see an effect for movement time until targets decrease to 40 pixels. When targets reduce to 30 pixels the task becomes extremely difficult for groups of 32, while there were no significant performance difference for groups of up to 16 children. Thus, our hypotheses H1 and H2 are only confirmed for smaller target sizes.

The results from the text typing task are interesting because they demonstrate that the impact of group size can be mitigated by designing tasks that don't require all children to be aiming for the same targets concurrently. Our results did not reveal any significant difference based on group size for the text typing task, despite the targets being relatively small (similar in size to the 80 pixel targets in the reciprocal pointing task). Since the children were separated into four groups, and each group was given a different word to type, there was less contention for each individual target. This enabled the children to perform the task equally well, regardless of group size. Thus, hypothesis H3 is confirmed.

The Group Size=8 condition was repeated on both days for the reciprocal pointing task; however, the performance results were different. While a small difference was not unexpected since different children completed the activities on the different days and mouse performance is impacted by individual differences, we did not expect such a large performance change. There are several additional factors that may have contributed to this difference. First, there may have been a practice effect due to the order in which the children did each Group Size condition on the different days. On the first day, the children completed the Group Size = 8 condition last, after having gained some experience with the task in previous conditions. In contrast, the children on the second day completed the Group Size = 8 condition first. Additionally, the children who participated on the second day received their practice and instruction in a group (as opposed to individually). These children may have still been getting comfortable with the task when they completed the Group Size = 8 condition. Ideally, one would counterbalance the presentation of Group Size in such studies and run more trials; however, given the logistical complexities of running studies of this nature with relatively large numbers of children in a classroom, it is unclear if the insights that might be gained by such counterbalancing would make the significant additional logistical efforts worthwhile.

Our results should be considered a lower bound on performance for these sorts of tasks, since the children had relatively little experience with the system even towards the end of the study. Should such systems be deployed in real classrooms, children using them every day will likely develop improved strategies for handling the difficulties posed by even 32 cursors.

It is worth noting some qualitative observations made during the study. One that was immediately apparent was that the participants seemed to feel a strong sense of competition. While such competitive behavior is likely a natural occurrence when children perform the same task concurrently, we also believe that because of the relative simple nature of the tasks, speed became a salient means of injecting some extra "fun" into the task. Thus, even when the task was completed alone in the Group Size = 1 condition, the children felt that they had to do it quickly.

CONCLUSION AND FUTURE DIRECTIONS

Overall, our results demonstrate that children can perform tasks comprising of target selections on a shared large display in large group sizes with minimal impact on performance as long as the targets are not too small. The increased clutter and potential occlusion resulting from having many cursors active on the display did not impact children's performance for most of the conditions we examined, although this could be partially due to the steps we took to ensure minimal visual overlap in the cursors we used. This means that multiple mouse single-display groupware configurations can be scaled up to include whole-class interactions, if care is taken to ensure that targets are a reasonable size, or that the task is structured such that not all children are trying to acquire the same targets at the same time. This opens the door to new user interfaces that enable large groups of simultaneous users.

While our results are promising, large group interactions with multiple mice systems can be further improved with new interaction methods. For example, interaction with small targets was challenging when many users were all trying to acquire the same target because the target (as well as some of the cursors) would become occluded. This might be alleviated by dynamically cycling the z-order of targets and cursors so that no cursor or target remains occluded for long, by expanding targets [19] when many cursors are over it, or by "blooming" the cursors away from one another by stretching their arrows when many cursors overlap.

Another concern is how to provide appropriate feedback. Since simultaneous attempts are being made to acquire the target, it is infeasible to provide visual feedback on the target. Nor is audio feedback feasible in a busy classroom. In our study we used feedback on the cursor (the reciprocal pointing task), or in a region of the screen assigned to a child (the text entry task). While these seemed to work well, other alternatives clearly merit further investigation.

In our study, we used wired mice which required significant gaffer taping to the ground to prevent wires from being disconnected from hubs. Wireless mice would be more suitable if the technology can be scaled support large numbers of currently active mice and to work over the relatively large size of a classroom.

ACKNOWLEDGEMENTS

We thank Kathleen Mulcahy for assistance and the teachers, parents, and students who participated.

REFERENCES

1. Card, S.K., Moran, T.P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
2. Crook, C. (1992). Young children's skill in using a mouse to control a graphical computer interface. *Computers and Education, 19*(3). p. 199 - 207.
3. Druin, A., Stewart, J., Proft, D., Bederson, B., & Hollan, J. (1997). KidPad: a design collaboration between children, technologists, and educators. *ACM CHI Conference.* p. 463-470.
4. Fittts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J. of Exp. Psychology, 47*. p. 381-391.
5. Greenberg, S., Gutwin, C., & Roseman, M. (1996). Semantic telepointers for groupware. *Australian Conference on Computer-Human Interaction (OZCHI).* p. 54.
6. Gutwin, C., & Greenberg, S. (1999). The effects of workspace awareness support on the usability of real-time distributed groupware. *ACM Transactions on Computer Human Interaction, 6(3)*. p. 243-281.
7. Hourcade, J.P., Bederson, B., & Druin, A. (2004). Preschool children's use of mouse buttons. *Extended Abstracts of the ACM CHI Conference.* p. 1411-1412.
8. Hourcade, J.P., Bederson, B., Druin, A., & Guimbretiere, F. (2004). Differences in pointing task performance between preschool children and adults using mice. *ACM Transactions on Computer Human Interaction, 11(4).* p. 357-386.
9. Hourcade, J.P. (2006). Learning from preschool children's pointing sub-movements. *ACM Conference on Interaction Design and Children (IDC)*. p. 65-72.
10. Inkpen K., Booth K.S., Gribble S.D., & Klawe M. (1995). Give and take: Children collaborating on one computer. *ACM CHI Conference.* p. 258-259.
11. Inkpen, K., Ho-Ching, W., Kuederle, O., Scott, S.D., & Shoemaker, G. (1999). This is fun! We're all best friends and we're all playing: Supporting children's synchronous collaboration. *ACM Conference on Computer Support for Collaborative Learning (CSCL).* Article No. 31. 12 pages.
12. Inkpen, K. (2001). Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Trans. on Computer Human Interaction, 8(1).* p. 1-33.
13. Izadi, S., Brignull, H., Rodden, T., Rogers, Y., & Underwood, M. (2003). Dynamo: a public interactive surface supporting the cooperative sharing and exchange of media. *ACM Symposium on User Interface Software and Technology (UIST).* p. 159-168.
14. Joiner, R., Messer, D., Light, P. & Littleton, K. (1998). It is best to point for young children: A comparison of children's pointing and dragging. *Computers in Human Behavior, 14*(3). p. 513-529.
15. Jones, T. (1991). An empirical study of children's use of computer pointing devices. *Journal of Educational Computing Research, 7*(1). p. 61-76.
16. King, J. & Alloway, N. (1993). Young children's use of microcomputer input devices. *Computers in the Schools, 9*. p. 39-53.
17. Kucera, H & Francis, W.N. (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Press.
18. MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction, 7*. p. 91-139.
19. McGuffin, M., & Balakrishnan, R. (2005). Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Transactions on Computer-Human Interaction, 12(4*). p. 388-422.
20. Moraveji, N., Kim, T., Ge, J., Pawar, U.S., Mulcahy, K., & Inkpen, K. (2008). Mischief: supporting remote teaching in developing regions. *ACM CHI Conference.*.
21. Pawar, U.S., Pal, J., Gupta, R., & Toyama, K. (2007). Multiple mice for retention tasks in disadvantaged schools. *ACM CHI Conference*. p. 1581-1590.
22. Russell, D.M., Drew, C., & Sue, A. (2002). Social aspects of using large public interactive displays for collaboration. *International Conference on Ubiquitous Computing (UbiComp).* p. 229-236.
23. Ryall, K., Forlines, C., Shen, C., & Morris, M. (2004). Exploring the effects of group size and table size on interactions with tabletop shared-display groupware. *ACM Conference on Computer Supported Cooperative Work (CSCW).* p. 284-293.
24. Salmoni, A. W. & McIlwain, J. S. (1979). Fitts' reciprocal tapping task: A measure of motor capacity? *Perceptual and Motor Skills, 49*. p. 403-413.
25. Scott, S., Grant, K., & Mandryk, R. (2003). System guidelines for co-located, collaborative work on a tabletop display. *European Conference on Computer Supported Cooperative Work (ECSCW).* p. 159-178.
26. Stanton, D., & Neale H.R. (2003). The effects of multiple mice on children's talk and interaction. *Journal of Computer Assisted Learning*, *19(2).* p. 220-228.
27. Sugden, D. A. (1980). Movement speed in children. *Journal of Motor Behavior, 12*. p. 125-132.
28. Stewart, J., Bederson, B. B., & Druin, A. (1999). Single display groupware: a model for co-present collaboration. *ACM CHI Conference.* p. 286-293.
29. Wallace, S. A., Newell, K. M. & Wade, M. G. (1978). Decision and response times as a function of movement difficulty in preschool children. *Child Development, 49*.