Enabling and Understanding Instructor-designed Multiuser Classroom Activities

Neema Moraveji[1], Idara Otu[1], Jillian Harmon[1], Fred Campbell[1], David Hutchful[2], Nasha Fitter[3]
neema, idaraotu, jcharmon, fredcam@stanford.edu, davihu, nashaf@microsoft.com
[1]Learning Sciences and Technology Design, School of Education, Stanford University, Stanford, CA
[2]Technologies for Emerging Markets, Microsoft Research India, Bangalore, India
[3]Unlimited Potential Group, Microsoft Corporation, Redmond, WA

ABSTRACT

We present the results of a study assessing how teachers, as designers of multiuser educational activities, used an authoring tool to create single-display groupware activities. The study, run over 4 months with 18 schools, 50 teachers, and 3233 students, shows how teachers used different user features to achieve different participation structures and pedagogical goals. When compared to individualistic activities, collaborative activities targeted higher cognitive levels of assessment. Analyses of the time spent on different activities and the number of student mice activated showed that teachers used voice instructions in place of programming logic in activities they designed. These results motivate code-free authoring environments for novices to create multiuser activities of collaborative, rather than individual, assessments to target higher cognitive levels.

Author Keywords

End-user authoring, multiuser interaction, pedagogy, shared display, single-display groupware, classroom.

ACM Classification Keywords

H5.m. Information interfaces and presentation: Misc.

INTRODUCTION AND MOTIVATION

Teachers traditionally appropriate technologies to complement instruction and to design learning experiences. In parallel, research has increased on multiuser, virtual classroom learning activities [23]. The trend towards digital learning content may mean a break in the tradition of teacher-designed learning content, threatening the resulting customization and sense of ownership, motivation, and expertise the teacher experiences. To address this, we envision a 'long tail' of interactive assessments emerging from teachers and third parties as learning trends towards interactive mediums. This paradigm conceptualizes of teachers as interface and interaction designers of multiuser assessments.

Our work in tools to enable teachers to design such content deals specifically with formative assessment. *Formative assessment* refers to the bidirectional interaction between teacher and students used by teachers to adapt teaching to meet the learner's needs [3, 6]. Based on such interaction, teachers adapt their teaching and learners can acknowledge gaps in their knowledge.

Individual studies and meta-reviews such as [3] have shown that formative assessment produces significant learning gains especially for low-achieving and learning-disabled students. Frequent, short assessments (as would be possible were more teachers able to create them) are better than infrequent long ones. Examples of effective techniques include: ascertaining the status and extent of existing student understanding, triggering peer discussion and instruction, having all students (rather than individual volunteers) generate ideas or answers, and having students contribute possible answers before and after instruction [3].

A platform for interactive formative assessment

Our approach to enabling this long tail of interactive formative assessment has been to create a platform for teachers to design their own activities that leverage rich input from students *en masse*. The name of the tool is Mischief (the collective noun for 'mice') and it works by giving each student or small group of students a wired or wireless mouse and connecting them all to a large display at the front of the classroom [19]. The teacher guides the activities that take place on the screen and controls which mouse is activated or not. Beneficial learning and behavioral outcomes have been documented using shared display technologies with individual input in both developed [28] and developing [12, 24] contexts.

About 10-30 mice are used in a single classroom. When there are not enough mice for each student to have one, they can share in sub-groups or take turns. Mischief opens Powerpoint files and, according to metadata appended by the instructor inside those slides, makes them interactive. The system maintains point values and cursor identities during a classroom session. The teacher, using a 'supermouse', orchestrates instructional content where each slide resembles a video game combined with a chalkboard.

Motivation for the present study

The design of Mischief means that, with relatively little overhead and cost, teachers can deploy a system capable of quickly gathering rich assessment data from a classroom full of students. In prior studies with teachers designing multiuser formative assessment activities using Mischief [19**Error! Reference source not found.**], they gave reasons why they find the system beneficial: increasing student engagement, detecting social loafing, empowering quieter students, pressuring select students in front of their peers, increasing/decreasing competition, and triggering frequent collaborative opportunities.

In the same studies, many teachers consistently invented new activity ideas, many of which were impossible with the system's features and others which we folded into the system feature set. This experience demonstrated that designing interfaces that support creation of content for multiuser users is susceptible to usability problems [19] and requires pedagogical theory to be useful.

To redesign the authoring environment such that it meets teacher's behavioral and learning goals, we must a) understand what pedagogical goals teachers aim to fulfill using such activities, b) understand how they try to reach those goals, c) and how their assessments, once designed, are used in class. These goals motivated the design of the present study that spanned four months, 18 schools, 50 teachers, and 3233 students.

RELATED WORK

This section covers technological formative assessment techniques, relevant work on participation structures in classrooms, design principles for multiuser interaction, and end-user programming as it relates to the present study.

Formative assessment and participation techniques

A number of technologies aim to provide instructors with rich and scalable means of formative assessment. The most popular of these are Audience Response Systems (ARS), also known as clickers (such as [8]). Each clicker is a purpose-made device that sends answers to multiple-choice questions wirelessly to a central server in the classroom. These are aggregated and displayed on a projected screen. Clickers have been shown to be pedagogically useful, being the core tool in the Peer Instruction [17] paradigm and improving teacher adaptation to student knowledge. The primary drawbacks of such systems are the lack of collaborative features and prohibitively high cost. Mischief offers ARS functionality with enhanced interaction and feedback between students and teacher at an affordable cost (possibly with resources that already exist in schools).

Participation structures "are variable contexts that provide social conditions for children's communicative functioning" [21]. They "are differentiated by the nature of the rules governing speaking, listening, and turn-taking at different times" [30]. They include templates and examples for peer interaction that lead to desired outcomes (e.g. reflection, idea generation, prediction). They are well known in the learning literature but rarely leveraged by interface designers to create specific social interactions.

A specific participation structure noted particularly in the developing world is that of "chorus response" [32] where the teacher asks a relatively simple question and the class responds in unison, mainly to ensure the pace of the class is consistent rather than as a means to identify individual knowledge status or to trigger discussion. We envision scalable systems like Mischief helping teachers administer assessments that are more granular. For example, interactive systems could elicit differentiated feedback efficiently while structuring turn-taking in sub-groups to ensure each student justifies their answers to peers.

Multiuser interaction design

Although a number of researchers have previously experimented with multiple mice for encouraging collaborative learning [7, 10, 25, 31], Microsoft's MultiPoint is the first example where this concept was applied to resource-constrained schools in developing countries as a way to expand access to computing at a lower per-child cost [23]. Mischief extends what MultiPoint did for small groups (e.g., 5 per PC [23]) in computer classrooms to whole classrooms where as many as 40 students each use mice connected to a single PC. We note that the quantitative difference in number of students leads to qualitative differences in the kinds of activities that are optimal for learning.

Researchers have studied specifically the performance of children using common mouse actions in single-user [7, 8, 11, 15, 16] and multiuser systems [19]. Guidelines for multiuser activities have also been devised for tabletop [10] displays but address on-screen interaction, not authoring.

Related end-user authoring tools

There has been a good deal of work on interfaces to support the creation of interactive applications without requiring programming. The purpose of these interfaces is often to serve application creators as prototyping tools but also for end-user experiences. HyperCard [2] used a card-stack metaphor and supported scripting as well as the placement of interactive components like buttons and text fields. Director used a 'stage' metaphor and a "carefully coordinated and timed sequence of activities by a cast of characters" [25] and, hence, was meant for heavyweight prototyping for end-user applications. Visual Basic [18] was meant for programmers who wanted a shortcut at creating the graphical user interface components. Hence, it makes drawing interfaces easy but relies on programming for any functionality. More recently, the Flash platform allows designers to create interactive animations while minimizing scripting required. Powerpoint has also made its way into prototyping environments as it supports some basic animation and logic functions. All of the aforementioned solutions, targeting different levels of expertise, are meant for designers to create single-user interactive applications to varying degrees of fidelity.

What we find missing from the literature is research on interfaces for novice users who only have basic expertise in productivity software to create multiuser activities.

NOVICE AUTHORING OF MULTIUSER ASSESSMENTS

In its first incarnation, Mischief offered a template-based content system that had teachers fill in content that would assess students in straightforward ways. The system quickly proved too constraining for the enthusiastic teachers who wanted to challenge students with novel assessments.

Beginning with a relatively simple authoring system, making only minor assumptions about the intended use, the system evolved through field deployment. To create a richer authoring tool, we used Powerpoint to leverage existing expertise observed in the East and West. The slide-based paradigm was simple to navigate and understand, fitting the classroom environment like a chalkboard. Beyond creating lectures, it was used by teachers and students alike as a medium for creative expression, communication, and note-taking, [19]. Though Powerpoint supports some amount of interaction, it is very limited. We needed to design a means of enabling the user to define activities that exploit real-time, continuous mice input.

We saw teachers use 'low-tech' methods of filling in gaps in Mischief. For example, Mischief does not currently support a 'bounding box' feature where teachers can constrain student cursors to a region. Hence, at design-time, teachers would draw a region using a standard square shape and verbally instruct students to stay within those bounds. Social norms dictated whether students followed instructions but the teacher's job was relatively simple.

Available widgets and the authoring user interface

Our three constraints for our authoring environment were that 1) it must work in Powerpoint, 2) it must maintain interoperability with standard Powerpoint files, and 3) the solution must work with both versions 2003 and 2007. To address (1) we built an add-in using VBA. To address (2), we exploited a simple means of attaching metadata to any shape on a slide: the 'alternative text' field, used when exporting a slide to HTML. In an early prototype, users were required to manually modify the alt-text field to add special keywords which would then be parsed by Mischief when launched. The format of the keywords is:

<keyword><parameter>

For example, adding "*clock*35" in the alt-text field of a shape instructed Mischief to render a timer with 35 seconds in that location and at that size. Adding "*choice" to an image makes it an answer choice in a multiple-choice slide.

The add-in helps users make modifications to the alt-text field automatically, with no requirement to remember keywords. Figure F shows a screenshot of the add-in in 2007, the 2003 version feature set is identical.

The interactive widgets a user can add to a slide are: draggable shapes [single, dual, triple-user (number of people required to simultaneously drag), and assigned], a palette for drawing, multiple choice answer choices, dialog balloons for text entry, soft keyboards, soft keypads, and timers. These are all described below.

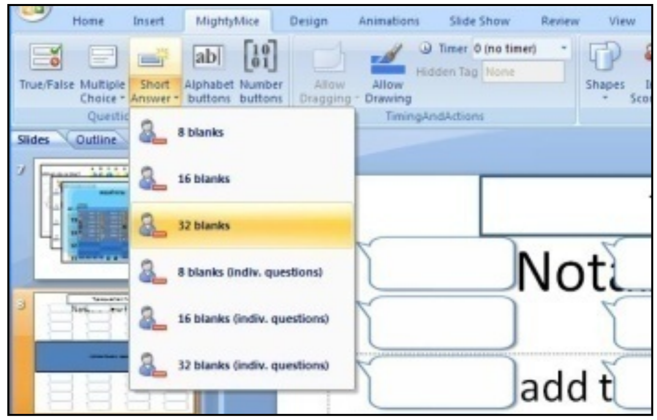

Figure 1: The Powerpoint 2007 user interface where a user is inserting 32 blanks for a short answer activity. You can see that some 'dialog balloon' objects are inserted into the slide with the alt-text automatically populated.

Multiple choice

Often, the first activity teachers create is a multiple-choice question. This can come in the form of true/false or answer choices marked A-F. There are two ways to create a multiple-choice activity: 1) click the 'True/False' or 'Multiple Choice' template button (Figure 1) to insert a question and a set number of answer choices with text labels to overwrite or 2) add these elements manually and select the answer choice on the canvas and choose 'Convert selection to answer choice' from the 'Multiple Choice' drop-down menu. As such, the teacher can use images and clip-art as answer choices. The question can be answer on the slide or can be absent, given via voice in class.

When run, students can click answer choices and then their mouse disappears after 5 seconds (configurable), giving them some time to change their choice. It can also be run 'anonymously' (see Figure 2b) whereby the cursor is not shown and users must click a mouse button combination correlating with an answer choice (left button is 'A', right is 'B', right+left is 'C', etc.).

Short Answer

Short answer activities present a means of eliciting free-form text input from students and can be used in multiple ways. The most straightforward way of creating one is to use a template from the Powerpoint add-in. The 'Short Answer' drop-down menu provides options to insert 2, 4, 8, 16, and 32 dialog balloons (places where an individual's input is displayed) and automatically arranges them along with a soft keyboard (see Figure 2a) and/or number keypad. Teachers can instead manually insert blanks, keyboard, and keypad as well and adjust their sizes according to their requirements. Each blank is assigned a single student, whose cursor is displayed adjacent to it. Though done rarely, teachers can type inside the blank to create a prompt for that individual – when run in Mischief that prompt is displayed immediately above the blank.

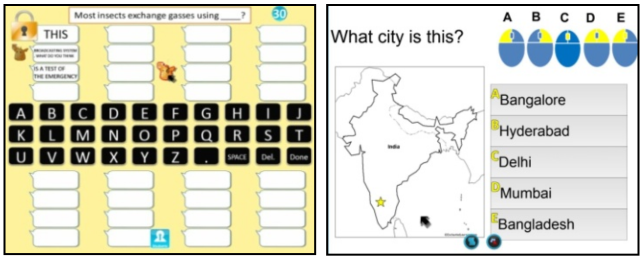

Figure 2: a) A short answer activity with 32 blanks and a soft keyboard. b) A multiple choice activity being conducted anonymously where answer choices are done via combinations of mouse clicks of mouse buttons instead of target acquisition. The 'legend' in the top-right would appear automatically if the teacher decides to administer the activity anomymously.

Drag and drop

The most flexible feature of the authoring environment is the ability to select an object and make it 'draggable'. It is flexible because it creates many reasons to drag items on-screen. Images, shapes, and text labels could all be made draggable. Teachers can set the number of students required to drag a piece from 1-3 or they can set a piece to support 'Assigned Drag' which means that, in run-time, the teacher can assign a student who 'owns' that piece. Drag-and-drop was used for categorization (Figure 3a), recall (Figure 3b), collaborative, and creative tasks. To illustrate the versatility, one teacher created a dissection activity for science class where part of a crawfish image was cut-out and draggable and a student had to perform a virtual dissection of it in front of their peers before performing the actual dissection in class.

Drawing

By toggling the 'Allow Drawing' button on a slide, a small palette is added to the top-left of the screen. Students could then individually choose a color to draw in and could delete their own ink via the 'X' button. The teacher could also draw and clicking 'X' would clear everybody's ink. Like drag-and-drop, drawing was used for a variety of purposes. For example, a student could volunteer to circle a region and explain its meaning to the class (Figure Fb). It could be used in a traditional 'matching' activity as well.

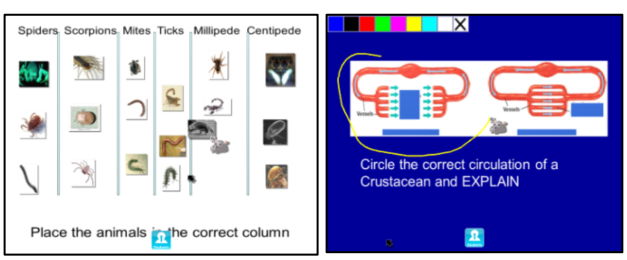

Figure 3: a) Drag-and-drop used to categorize insects into their phylum. b) A drawing activity where a student is asked to circle the correct illustration of a crustacean circulatory system and explain why it is correct.

Teachers, as they are traditionally viewed, are already designers – they design learning environments, curriculum, and activities. Our goal is to create a system that allows them to leverage that knowledge to create lightweight interactive learning activities without scripting or programming. We found that teachers would use verbal instructions to provide or replace logic, error-checking, and intelligent system behavior that usually requires code. We had received some feedback about our system but wanted to better understand how true teacher needs would arise in a longitudinal deployment so we created a study.

FIELD STUDY

We had three research questions:

a) Which pedagogical goals did teachers aim to fulfill using the system?
b) Which participation structures did they use to address these goals?
c) How were their assessments, once designed, actually used in class?

Building upon these, the goal of the present field study was to understand how teachers designed formative assessment activities for shared display environments and the nature of use of those activities by students. Our aim was not to ascertain learning effects of its use but rather to uncover common and unexpected uses to help us understand how teachers want to design multiuser formative assessments.

Logistical difficulties in conducting design-based research in learning environments plague the research community. Given tight schedules and high potential opportunity costs, it is difficult for researchers to gain insight into how a classroom technology is appropriated *in situ*.

Given these constraints, and to reduce the impact of the novelty and Hawthorne [9] effects as much as possible, we designed a field study whereby teachers and schools would have access to the technology in their schools and use it over a various amount of time independently. We instrumented the technology to log data that our distributed team members collected periodically.

Method

Administering a field study across a number of different countries required the involvement of a number of local researchers trained to administer a common procedure. We created a document with instructions for collecting data and content to train teachers. All teachers were trained using these source materials but there was certainly variability between trainers in teaching style.

Participants

Our team contacted teaching institutions in several Southeast Asian countries to make the study available to willing institutions. This call made the intent of the study known and included a video of the system in action. The study ran from Jan. 10, 2009 – Apr. 30, 2009. Participating schools and teachers were not compensated. The teachers volunteered and were proficient with Powerpoint before the study began. This included 50 teachers in 18 schools (7 primary, 15 secondary) with children and teachers varying in socioeconomic status. 3233 students used the system for approximately 310 classroom hours in total.

User Training

The teachers that responded were granted permission by their school principals to attend an optional training workshop conducted by a local stakeholder who was familiar with Mischief and was from the local country. The training session consisted of going through using the system as an author and as a student and creating sample slides.

Installation

Local technologists helped install the Mischief software on a school's computer(s) and decided how best to install the mice according to the characteristics of the target room. We did not purchase mice or USB cables for schools. This shows that teachers were interested in the system for pedagogical, not financial, reasons.

Data collection

The participating teachers were told about the data we would collect over the course of a few months. The system was instrumented to record individual user activity while in use. We manually collected these log files along with the Powerpoint files created and responses to paper surveys collected from the instructors and students.

Each row in each log files contained the following columns of data: time since session started (in milliseconds), unique cursor identifier, position of cursor on screen, event type, object type under the cursor, coordinates of said object, and unique object identifier. The surveys were administered at three points during the study: before the study began, once during the study, and after it was over. The surveys asked about their opinion of technology and impressions of its applicability and utility in the classroom.

RESULTS

We analyzed our data from two angles: 1) analysis of the teaching content created by teachers and 2) the nature of student in-class usage of those activities.

Content Analysis

Because the interactive content created by teachers was created in Powerpoint, we use the word 'slide' to refer to a single, independent activity and the word 'file' to refer to the set of content (both interactive and non-interactive) for a class. A teacher may use a single file for all her classes in a single day or only once. It is also possible she use one file over the course of two or more sessions but we did not report these occurrences. Further, anecdotal evidence showed us sometimes teachers would have students do an activity using paper and pencil and use the slides to augment the activities.

Creating a sample

We selected a sample set of slides from the Hanoi, Vietnam area to study because that data was returned to us early enough to do the analysis in a timely manner and it represented schools at multiple socioeconomic levels from the same culture. This sample contained 32 files from 19 teachers at 12 schools. There were 9 science, 9 math, 10 English, 2 social studies, and 2 computer literacy classes. These files contained a highly varied number of non-interactive slides that we ignored for the purposes of this study. Table 1 describes the sample dataset.

Statistic	**Population**			**Sample**	
	n	**%**		**n**	**%**
Content files	242	100%		32	N/A
Total slides	3773	100%		440	N/A
Interactive slides	1877	49.7		201	45.5
Drag-drop	338	18.0		31	15.4
Multiple-choice	867	46.1		110	54.7
Short-answer	539	28.7		47	23.3
Paint	133	7.0		13	6.5

Table 1: Descriptive statistics of the slides analyzed.

To confirm that our sample was representative of the population of content created by teachers, we ran a *chi square* test of association between the distributions in the two groups. Our null hypothesis was that there is no significant difference between the distribution of interactive activities between the population and the sample such that any differences can be attributed to random factors. We found the distributions not to be significantly different, $c^2(3, N=8)=5.43$, $p > .05$. Our sample serves as an appropriate representative of the population.

Mischief is meant as an authoring tool, not a collection of activity templates. As such, teachers can mold widgets into multiple purposes. For example, a short answer activity can be collaborative or individualistic, high level or low level. For this reason, we are interested less in which widgets teachers used than with what types of activities they created and what structures they used to do so.

Dimensions: pedagogical goal and participation structure

Our team inspected a number of slides as a group and used two dimensions along which to classify the content: Pedagogical Goal and Participation Structure (see Table 2).

Dimension	**Possible values**
Participation Structure	Individual, Collaborative
Pedagogical Goal	Knowledge/Comprehension (Level 1), Application/Analysis (Level 2), Synthesis/Evaluation (Level 3)

Table 2: Each slide was rated using the possible values for each of two dimensions.

For the participation structure dimension, 'Individual' activities were activities where a group of students does individual tasks in parallel. Most multiple-choice questions fall into this category because the each student answers and is assessed individually even though they do the task as a group. 'Collaborative' structures are those where the students interact with the system and with one another either virtually or physically. Technically speaking, 'cooperative' activities are those where a larger task is split up and students work independently towards a shared goal and 'collaborative' is when students work actively together on aspects of a shared task [25]. We did not distinguish between the two in this paper.

The pedagogical goal dimension is based on Bloom's Taxonomy of Cognitive Levels [4]. Table 3 summarizes this taxonomy in a succinct manner using applicable keywords. For the purposes of this study, we consolidated each pair of proximal levels into a meta-level, indicated on the right-hand side of the table.

Bloom's level	Keywords	Meta-level
Knowledge	define, duplicate, label, name, order, recognize, relate, recall.	*Level 1*
Comprehension	classify, describe, discuss, explain, express.	
Application	apply, choose, demonstrate, dramatize, employ, illustrate.	*Level 2*
Analyze	analyze, appraise, calculate, categorize, compare, contrast.	
Synthesis	arrange, assemble, collect, compose, construct, create.	*Level 3*
Evaluation	appraise, argue, assess, attach, choose compare, defend.	

Table 3: Categories for the 'Pedagogical Goal' dimension, based on Bloom's Taxonomy. Pairs of levels were grouped.

To ensure a common understanding of the codes before commencing individual coding, we used a sample of 25 slides from 7 files from the Hanoi area to conduct a test of inter-coder agreement. These slides were selected by the researcher most familiar with the content to cover the range of activities possible with the system. Three of the team members rated 10 slides then met to resolve discrepancies. The team rated 15 additional slides individually. Table 4 reports agreement scores for raters R1, R2, and R3.

Variable	R1+R2	R1+R3	R2+R3	Average
Participation Structure	PA=0.68 PE=0.50	PA=0.72 PE=0.50	PA=0.64 PE=0.50	PA=0.68 PE=0.50
Pedagogical Goal	PA=0.68 PE=0.33	PA=0.88 PE=0.34	PA=0.72 PE=0.33	PA=0.76 PE=0.33

Table 4: Agreement scores among raters R1, R2, and R3. PA=Proportion of Agreement. PE=Proportion of Agreement due to Chance.

The inter-coder agreement scores show substantial agreement for participation structure and pedagogical goal with the latter being stronger. We attribute the lower score for the participation structure to the fact that some activities that look as if they are individual could indeed be interactive and vice versa.

Inspecting slides alone cannot account for exactly how they were being used in the actual classroom. We did not have video recordings of all classrooms sessions to confirm ratings. This is particularly poignant when verbal instructions make a seemingly pedagogically low-level activity higher or an individual activity collaborative.

Assessment types

By rating a slide on its pedagogical goal and its participation structure, we ascribe an 'assessment type' to slide. There are 6 possible assessment types (see Table 5).

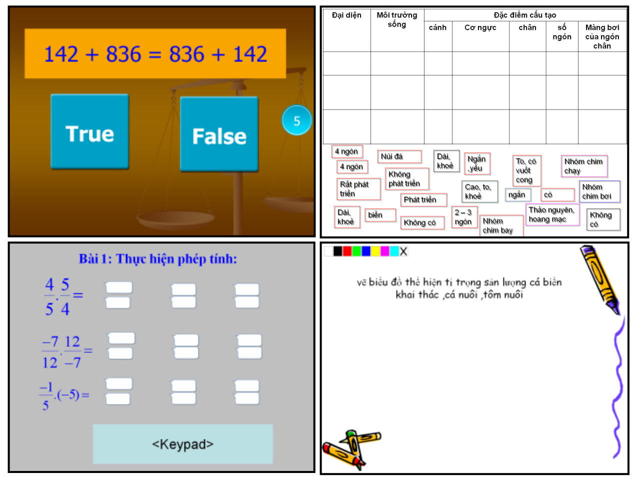

Figure 4abcd (clockwise from top-left): a) Multiple-choice; individual participation structure, level 2. b) Short answer, individual, level 2. c) Short answer, collaborative (18 students, 3 problems), level 2. d) Draw, individual (a student must diagram the ratio of production between sea fish harvesting, fish feed, and shrimp feed), level 3.

We explain this table with examples that use different interaction widgets to accomplish varying goals. First, Figure 4a is a multiple-choice slide that is Level 2 pedagogical goal (it assesses a student's ability to calculate an expression) and treats each student as an individual. Compare this to Figure 4b uses drag-and-drop to create a collaborative structure where the group must categorize (Level 2) a set of environmental characteristics at the bottom into the boxes. Figure 4c uses short answer to create a collaborative structure whereby the group must show the work to solve 3 problems. Each person is responsible for one part (numerator, denominator) and the class ensures the proof is correct (Level 1). Figure 4d uses the draw feature to create an individual structure where the student must draw a diagram comparing the proportions of sea fish harvested with fish feed and shrimp feed (Level 3).

Relating pedagogical goal with participation structure

How is pedagogical goal related to participation structure? This question is interesting because if teachers are using collaborative assessments for higher pedagogical goals, this motivates improvement of the way such assessments are designed. Thus, if authoring tools ease the design of collaborative assessments, classrooms would have higher-level assessments in them, even in resource-poor schools.

To illustrate this point, consider a teacher in a classroom of 60 attempting to ensure each student understands the current material. She does not have time to interact with all of them individually so she resorts to asking rapid-fire questions and getting volunteers to answer them or eliciting chorus response answers. Using a technology system to engage her students and to get granular input, she can also ask rapid-fire questions. However, if the technology effectively supports the design of collaborative activities, she can present higher-level assessments (i.e. ones that spawn formal argument, analysis, and application of concepts as opposed to simply recall) that can create constructive conflicts between students and lead to learning.

Participation structure	Pedagogical goal		
	Level 1	**Level 2**	**Level 3**
Individual	87 (82.1%)	49 (62.9)	10 (58.9)
Collaborative	19 (17.9%)	29 (37.1)	7 (41.1)
Total	106 (100%)	78 (100)	17 (100)

Table 5: Two-way table of associations between Pedagogical Level and Participation Structure from our sample set.

To determine the significance of the differences in the proportions of participation structures between the different pedagogical levels, we ran a *chi square* test of association between pedagogical goal and type of participation structure (see Table 5). Our expectations are that:

a. At Level 1, there are proportionally more individual activities.
b. At Level 2, individual and interactive activities would occur at an equal rate.
c. At Level 3, there are proportionally more collaborative activities.

Our null hypothesis is that the proportional makeup of participation structure does not change as pedagogical goal changes and our alternative hypothesis is that is does. We found that the proportions are different, c^2(2, *N*=201)=10.16, *p*=.006, but we must now understand how they differ, using a *post-hoc* analysis.

We found that the proportions between Levels 1 and 2 are significantly different, *c2*(1, *N*=184)=8.64, *p*=.003, and that Levels 1 and 3 are as well, c^2(1, *N*=123)=4.75, *p*=.029. Levels 2 and 3, however, are not significantly different, c^2(1, *N*=95)=0.95, *p*=.758. Expectation (a) and (c) were met but (b) was not. This is an unexpected result: teachers employed more collaborative activities to access higher levels of cognitive learning. Further *chi square* analysis showed there is an interaction between pedagogical goal and number of slides, c^2(4, *N*=201)=402.019, *p*=.000, each level being significantly different from the others (p<.05).

<table>
<tr><th>Ped. level</th><th>Part. structure</th><th colspan="2">Drag-drop</th><th colspan="2">Multiple choice</th><th colspan="2">Short answer</th><th colspan="2">Draw</th><th>All</th></tr>
<tr><td rowspan="2">1</td><td>I</td><td>5</td><td rowspan="2">14</td><td>64</td><td rowspan="2">65</td><td>17</td><td rowspan="2">21</td><td>1</td><td rowspan="2">6</td><td>87</td></tr>
<tr><td>C</td><td>9</td><td>1</td><td>4</td><td>5</td><td>19</td></tr>
<tr><td rowspan="2">2</td><td>I</td><td>0</td><td rowspan="2">13</td><td>33</td><td rowspan="2">38</td><td>11</td><td rowspan="2">19</td><td>5</td><td rowspan="2">8</td><td>49</td></tr>
<tr><td>C</td><td>13</td><td>5</td><td>8</td><td>3</td><td>29</td></tr>
<tr><td rowspan="2">3</td><td>I</td><td>0</td><td rowspan="2">6</td><td>2</td><td rowspan="2">2</td><td>8</td><td rowspan="2">8</td><td>0</td><td rowspan="2">1</td><td>10</td></tr>
<tr><td>C</td><td>6</td><td>0</td><td>0</td><td>1</td><td>7</td></tr>
<tr><td colspan="2">Sum</td><td colspan="2">33</td><td colspan="2">105</td><td colspan="2">48</td><td colspan="2">15</td><td>201</td></tr>
</table>

Table 6. Interface widgets teachers used to create different assessment types (2I means 'Level 2, Collaborative').

Table 6 shows how teachers used different user interface widgets to different ends. While multiple-choice is used more for Level 1 assessments, the others are more evenly distributed, c^2(6, *N*=201)=17.449, *p*=.008. Drag-and-drop is used for proportionally more collaborative activities than multiple-choice, c^2(2, *N*=138)=20.153, *p*=.000, and short answer are, c^2(2, *N*=81)=5.714, *p*=.057.

In-class Usage

We also collected log data of mouse movements and actions in the classrooms. Some of these files were inadvertently deleted, lost, or corrupted before we could look at them so we present results from a set of 123 slides from 12 classroom sessions across 9 teachers (these were not the same slides as the ones we rated). The goal of this was to understand how the designs were used in classes.

The mean number of mice connected to the system in these classrooms was higher than expected (*M*=17.13, *SD*=3.09) as was the mean number of mice activated in a slide (*M*=10.43, *SD*=7.53). We attribute the high standard deviation of the latter to the fact that some assessments required only one mouse while others required all the students in the class (e.g. multiple-choice).

The average times (in seconds) spent using different interaction (shown in Figure 5) are as follows: non-interactive (*M*=68.2, *SD*=76.24), multiple-choice (*M*=39.34, *SD*=25.64), short answer (*M*=158.6, *SD*=182.95), drag and drop (*M*=233.3, *SD*=259.09), draw (*M*=332.6, *SD*=408.02). Non-interactive slides are slides that had no clear widgets added to the slide but the teacher activated mice on the slide for some purpose. There is a multitude of reasons a teacher might activate students on a non-interactive slide. One example of a reason for this is where the teacher puts images of different insects on the slide and asked a student(s) to simply move their cursor to the picture of an arachnid. There is no need for widget-level interaction.

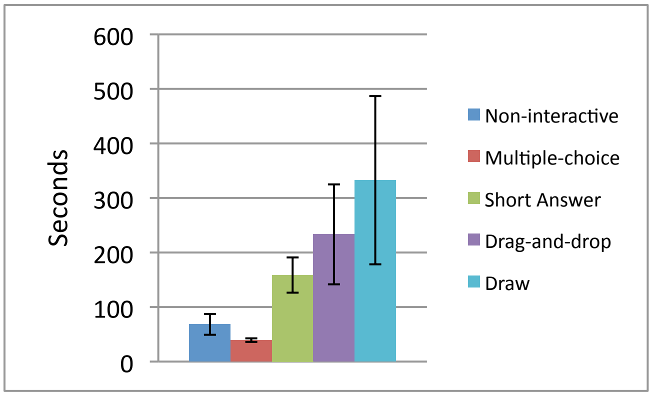

Figure 5: Average time spent using different interaction activities. Error bars represent standard error.

A one-way ANOVA analysis shows a significant interaction between activity type and duration, $F(4, 118)=9.71$, $p<.01$. A *Tukey HSD post-hoc* analysis showed the following activities are significantly different from one another: non-interactive/draw ($p=.001$), multiple-choice/short answer ($p=.004$), multiple-choice/drag-and-drop ($p=.007$), multiple-choice/draw ($p=.000$), and short answer/draw ($p=.047$).

One unexpected finding from this analysis is that multiple-choice activities are not significantly longer than non-interactive activities. There are multiple possible explanations of this. Those non-interactive slides could be instructional material where the teacher is teaching and activating students for various reasons during instruction. Why are multiple choice questions so short? One would expect that teachers would be administering a question and giving feedback about why an answer is wrong or right. This could mean that the questions that are too simple, teachers are not expanding upon or explaining answers, or that they are not leveraging peer instruction practices.

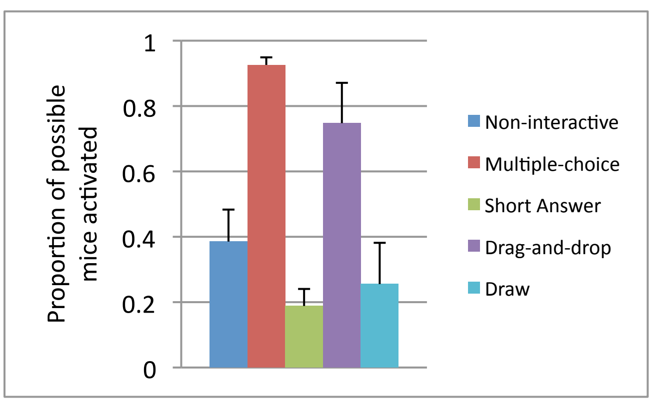

Figure 6: The proportion of mice activated out of the total number available in the class, across different activities. Error bars represent standard error.

A given class had 6-30 mice available to them, depending on hardware setup. The teacher decides how many mice to activate for each activity. She can activate all of them via the Enter keyboard key or a single soft-button press. Or, she can activate individual students by clicking their icon on the on-screen student list. Upon each slide change, all the mice are automatically deactivated. Figure 6 shows the proportion of students activated across activity types.

A one-way ANOVA analysis shows a significant interaction between activity type and duration, $F(4, 118)=47.316$, $p<.01$. A *Tukey HSD post-hoc* analysis showed the following activities are significantly different from one another: non-interactive/multiple-choice ($p<.001$), non-interactive/drag-and-drop ($p=.018$), multiple-choice/short answer ($p<.000$), multiple-choice/draw ($p<.001$), and drag-and-drop/draw ($p=.005$).

Seeing multiple-choice as having the largest proportion is not surprising because the activity scales well. On the other hand, it is surprising to see short answer as the lowest. The reason is that it is relatively simple to add a high number of dialog balloons to a slide and activate many people. Drag and drop is also surprisingly high given the potential visual clutter than can arise when many students are moving objects simultaneously. Perhaps teachers are not as bothered by visual clutter as much as we expected. Interestingly, non-interactive slides had a non-negligible amount of students activated on them. One interpretation of this is that teachers are creating collaborative activities that are not supported by the existing feature set.

Anecdotally, we observed various activation patterns among teachers and activities. For example, one teacher using a drawing activity might activate 3 students at a time to draw something on-screen together while another might let each one try it individually and watch them converge sequentially on the right answer.

Survey Results

Pre- and post-pilot survey data was collected from 12 teachers across six schools in the Philippines, Malaysia, and Thailand to elicit expectations and reactions after using the system over time. Surveys were taken from these schools because data the pre/post surveys were fully completed and returned on time. The surveys asked largely different questions but several were repeated on either survey. The two limitations of the surveys are that a) local cultural norms meant responses were subject to ceiling effects (teachers did not want to offend us) and b) our study participants were a self-selecting group: these teaching already believed technology had great potential to improve teaching. As a result, Table N presents the data as an addendum to illustrate teacher experiences. The question read, “What expectations of impact do you have of the Mischief system…” followed by the different areas. We ran a dependent sample t-test to identify significant changes between the pre- and post-study surveys.

The results show that, generally, expectations were overwhelmingly high (likely due to cultural norms) and, as a result, difficult to surpass. We see the expectations of positive impacts on classroom discussion ($t(11)=-2.28$, $p=0.043$) and student learning ($t(11)=-3.023$, $p=0.012$) was significantly higher after the pilot. This may speak to how teachers entered the pilot with conventional opinions of how technology can be useful: engaging students and teachers and assessing them. After the pilot, however, their opinions change to think of technology as useful for initiating positive classroom discussions and positively impacting student learning.

Expectations of system impact on…	**Before**	**After**
…classroom discussion.	5.1	5.8*
…student engagement.	6.3	6.7
…assessing students.	6.3	6.4
…student learning.	6.0	6.6*
…teacher engagement.	6.3	6.3

Table N: Mean teacher ratings (on Likert scales of 7 where 1 is negative) before and after the study. *significant (p<0.05).

In free-form comments from the surveys, teachers desired more animations and hyperlink support (they are accustomed to Powerpoint), which Mischief does not currently support. Both of these requests speak to how teacher desire to have more control and flexibility of the authoring environment and motivates our future work.

Study Limitations

The first limitation of the study is that we do not know how the teachers used some of these activities (i.e. what she said as she presented the slides and led the activities). That is, a slide that looks as if it is being used in one way could actually have been used in a different way in the classroom.

Another limitation of the study is that we do not know the act use structure (i.e. did students use the mice in small groups or individuals) and how turn-taking occurred. A multiple choice activity that seems as if it is meant for individuals may have been created for small groups to come to consensus on. Similarly, as noted earlier, our raters could not know exactly how an activity was used in the classroom setting due to a lack of video data accompanying the slides. Lastly, Bloom's taxonomy has frustrated many researchers for its vagueness and inaccuracy

System and interaction paradigm limitations

We encountered a number of limitations to deploying and using the system design under study (using multiple mice on a large display in a classroom setting) that certainly affected the study results. First and foremost, the system was a research prototype and there were efficiency problems most noticeable when opening Powerpoint files. The system did not support non-trivial Powerpoint animations nor did it support hyperlinks, significantly limiting what teachers were able to design. Setting up the room for use with mice was a non-trivial problem that required technicians to help extend USB cables from the computer to the desks and to install covering to ensure students did not trip over the wires. The costs of wireless mice were prohibitive and tests with wireless mice available to the region proved the connections unstable when used in concert and at long range. The system is not designed for group-use in mind and, hence, students that might have been facing one another and working together instead are seen to face the screen, speaking to one another as an aside. Technical problems, as are common in such deployments, caused further complications and disruptions to class, as did power disruptions and overheating of the projector.

DISCUSSION

Our research questions were:

a) Which pedagogical goals did teachers aim to fulfill using the system?
b) Which participation structures did they use to address these goals?
c) How were their assessments, once designed, actually used in class?

To address (a), we saw that teachers used significantly more activities that were of Level 1 pedagogical goal than of Level 2 or 3. We attribute this largely to the prevalence of multiple-choice activities, which are simple to create, administer, and understand. Teacher usage of Level 2 and 3 activities was not significantly different. In short, teachers used the system for both simple assessments as well as for challenging ones. This speaks to a teacher need for system support beyond simple recall assessments.

To address (b), we found that about half of the slides created by teachers overall used interactive widgets and a portion of those that were not interactive were also used in some interactive way. This significant portion of interactive slides speaks to teachers' underlying belief or desire in interaction complementing or integrating with instruction. We did find an interaction between participation structure and pedagogical goal, with the finding being that higher-level assessments used a significantly greater proportion of collaborative assessments.

To address (c), we found that teachers desire more functionality than clickers allow (as illustrated by the significant number of collaborative assessments in general) and, crucially, such new collaborative functionality may bring higher level assessments to the classroom. Indeed, this finding is corroborated by the finding that teacher expectations of the system's ability to affect classroom discussion and student learning improved significantly after the pilot. It is also supported by finding that multiple-choice activities, numerous but extremely terse, were not used for peer-instruction, as is recommended.

A paramount concern of multiuser classroom systems is that the resulting visual clutter and potential chaos of the system would distract teachers. However, the teachers in our study, evolved expert practices both in designing activities (by creating multiuser activities that went beyond the widgets available to them) and in administering them (by employing strategies for activation patterns instead of activating/de-activating students *en masse*). For example, teachers consistently activated large numbers of mice in drag-and-drop activities, assumedly because they had ideas for addressing the clutter (something we saw in prior work).

Multiuser systems may complement individual workstations in the future. Teachers who desire to do so may take on the role of designers of novel interactive assessments that challenge students. Giving teachers tools to design, study, and share such multiuser activities may lead to beneficial outcomes in the modern classroom.

CONCLUSION

We described the evolution of a tool for non-experts to author multiuser activities in a classroom context. We described a longitudinal field study in a large number of classrooms across different contexts. Data showed that teachers designed far more low-level activities that are relatively simple. For those activities, they used more individualistic designs. For higher-level forms of assessment, they tended to create multiuser activities rather than individualistic ones. In classes they taught where the system was used, they used it frequently and spent significantly more time on multiuser activities than on individualistic activities. Teachers appropriated the system to create novel multiuser activities that were not built into the system. After using the system over time, their expectations about the system's effect on classroom discussion and student learning improved significantly.

Next Steps

We will analyze how teachers create interactive activities out of slides with no interactive widgets on them for this may inform new activity design. Next, we will inspect effective instructional routines and pedagogies for collaborative assessment design to incorporate those techniques into the user interface. This will require observation of existing collaborative classroom activities as well as participatory design with instructors and students.

REFERENCES

1. Anderson, R., Anderson, R., VanDeGrift, T., Wolfman, S., Yasuhara, K. (2003). Promoting interaction in large classes with computer-mediated feedback. *In Proc. of ACM CSCW '03*
2. Apple Computer. http://apple.com/hypercard
3. Black, P. and Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 80 (2): 139-148. http://www.pdkintl.org/kappan/kbla9810.htm
4. Bloom, B., Englehart, M. Furst, E., Hill, W., Krathwohl, D. (1956). Taxonomy of educational objectives: The classification of educational goals: Cognitive domain. Longmans, Green.
5. Bricker, L.J., Tanimoto, S.L., Rothenberg, A.I., Hutama, D.C., Wong, T.H. (1995). Multiplayer Activities that Develop Mathematical Coordination. *In Proc. of CSCL '95.*
6. Cowie, B., & Bell, B. (1999), A model of formative assessment in science education, Assessment in Education, 6: 101-116.
7. 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.
8. eInstruction. www.einstruction.com. Accessed June 2009.
9. Landsberger, H. A., *Hawthorne Revisited*, Ithaca, 1958.
10. 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.
11. Hourcade, J.P. (2006). Learning from preschool children's pointing sub-movements. *In Proc. of ACM IDC'06*.
12. Hourcade, J., Bederson, B. (1999). Architecture and Implementation of a Java Package for Multiple Input Devices, *University of Maryland Technical Report*.
13. 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.
14. 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. *In Proc. of CSCL*.
15. 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.
16. 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).
17. Mazur, E. Peer Instruction: A User's Manual. Series in Educational Innovation. *Prentice Hall*, NJ, USA, 1997.
18. Microsoft Corp. Accessed June 2009.
19. Moraveji, N., Kim, T., Ge, J., Pawar, U.S., Mulcahy, K., Inkpen, K. (2008). Mischief: supporting remote teaching in developing regions. *In Proceedings of ACM CHI'08.*
20. Moraveji, N., Inkpen, K., Cutrell, E., Balakrishnan, R. (2009). A Mischief of Mice: Examining Children's Performance in Single Display Groupware Systems with 1 to 32 Mice. *In Proceedings of ACM CHI'09*.
21. Nadel, J. & Camaioni, L. (1993): Introduction. In Nadel, J. & Camaioni, L. (Eds). New Perspectives in Early Communicative Development. New York: Routledge.
22. Papert, S., Idit, H. Constructionism. *Ablex Publishing Corporation*. Norwood, NJ, 1991.
23. Pawar, U.S., Pal, J., Gupta. R., and Toyama, K. (2007) Multiple Mice for Retention Tasks in Disadvantaged Schools, *In Proceedings of ACM CHI'07*, ACM Press.
24. Pawar, U. S., Pal, J., and Toyama, K. (2006) Multiple mice for computers in education in developing countries, *IEEE/ACM Conf. on Info. & Comm. Technologies for Development,* 2006.
25. Roschelle, J., Vahey, P., Tatar, D., Kaput, J., Hegedus, S. J. (2003). Five key considerations for networking in a handheld-based mathematics classroom. *Proc. of 2003 Joint Meeting of PME/PMENA*.
26. Russell, D.M., Drew, C., & Sue, A. (2002). Social aspects of using large public interactive displays for collaboration. *International Conference on Ubiquitous Computing*.
27. Scott, S., Grant, K., & Mandryk, R. (2003). System guidelines for co-located, collaborative work on a tabletop display. *European Conf. on Computer Supported Cooperative Work*.
28. Stanton, D., Neale H.R. (2003). The effects of multiple mice on children's talk and interaction. *Journal of Computer Assisted Learning*.
29. Stewart, J., Bederson, B. B., & Druin, A. (1999). Single display groupware: a model for co-present collaboration. *In Proceedings of ACM CHI'99*. p. 286-293.
30. Tharp, R. G., Gallimore, R. (1988). Rousing Minds to Life: Teaching, Learning and Schooling in Social Context. Cambridge University Press.
31. Tse, E., Greenberg, S. (2004). Rapidly prototyping single display groupware through the SDGToolkit. *Proc. Australasian User interface*.
32. Watson, A. (2000). Chorus Response in Cape Town Schools. *Proc. of the British Society for Research into Learning Mathematics*.
33. Winograd, T. Bringing Design to Software. ACM, New York, 1996.