How Do Instructors Design Classroom-wide Interactive Formative Assessments? A Field Study with 18 Schools

Neema Moraveji, Roy Pea, Learning Sciences and Technology Design, Stanford University
Email: neema@stanford.edu, roypea@stanford.edu

Abstract: We present the results of a study assessing how instructors used an authoring tool called Mischief to create formative assessments. The study, run over 3.5 months with 18 schools, 50 instructors, and 3233 students, shows how instructors used different features to achieve different participation structures and pedagogical goals. When compared to individualistic activities, collaborative activities targeted higher cognitive levels of assessment. These results motivate code-free authoring environments for novices to create multiuser activities of collaborative, rather than individual, assessments to target higher cognitive levels.

A Platform for Large-group, Interactive Formative Assessment

Individual studies and meta-reviews such as (Black, Wiliam, 1998) 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 instructors 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.

Our approach to enabling instructor-designed interactive formative assessment has been to create a platform for instructors 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 a wired or wireless mouse and connecting those mice all to a single computer whose display is being projected at the front of the room. The system is described in detail in (Moraveji, *et al*, 2008). Each cursor maps to a unique cursor and multiple cursors can be on-screen simultaneously. Anywhere between 5-30 mice are used in a single classroom. When necessary, students can share a single mouse in a small group or take turns. Mischief reads Powerpoint files and, according to metadata appended by the instructor inside each slide, renders slide contents with interactive components (described in detail below). The system maintains answers and cursor identities during a classroom session. The instructor, using a 'supermouse' and her keyboard, orchestrates the presentation of instructional content where each slide is a separate activity.

This study aims to a) understand what pedagogical goals instructors 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. To answer these, we conducted a field study spanning 3.5 months, 18 schools, 50 instructors, and 3233 students.

Field Study

We deployed the technology to schools and instrumented it to log data that our team members collected manually, every several weeks. Our team contacted teaching institutions in several Southeast Asian countries. This call made the intent of the study known and included a video of the system. The study ran from Jan. 10, 2009 – Apr. 30, 2009. Schools and instructors were not compensated. We did not purchase mice or USB cables for schools. The instructors volunteered and were proficient with Powerpoint. This included 50 instructors in 18 schools (7 primary, 11 secondary). 3233 students used the system for approximately 310 classroom hours in total. Each row in each log file contained the following columns of data: time since session started (in milliseconds), unique cursor identifier, position of cursor on screen, event type (mouse movements, clicks, and keys typed), object type under the cursor, coordinates of said object, and unique Powerpoint object identifier.

Results

Based on a representative sample, 49.7% of the slides instructors created were interactive, 18% were drag-and-drop, 46.1% multiple-choice, 28.7% short-answer, and 7% drawing. There were two types of participation structures: 'Individual' activities where a group of students does individual tasks in parallel. 'Collaborative' structures are those where the students interact with the same objects towards the same overall coal.

The Pedagogical Goal of a slide was based on Bloom's Taxonomy of Cognitive Levels (Bloom, 1956). We consolidated each pair of proximal levels into three meta-levels. We used Bloom's taxonomy to identify the instructor's *intention*. It does not describe how the slide was actually *used* but, rather, how it was intended to be used. How is pedagogical goal related to participation structure? This question is interesting because if instructors 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.

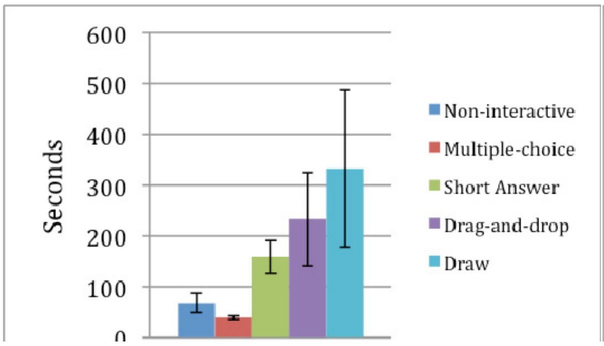
Table 1: Two-way associate table of associations between Pedagogical Goal and Participation Structure of instructor-designed formative assessments based on a representative sample.

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 2: Interface widgets instructors used to create assessment types (e.g. 2I = 'Level 2, Collaborative'). Note this table contains only interactive slides (201 of the 440 total slides in our sample).

Ped. level	Part. structure	Drag-drop		Multiple choice		Short answer		Draw		All
1	Indiv.	5	14	64	65	17	21	1	6	87
	Collab.	9		1		4		5		19
2	Indiv.	0	13	33	38	11	19	5	8	49
	Collab.	13		5		8		3		29
3	Indiv.	0	6	2	2	8	8	0	1	10
	Collab.	6		0		0		1		7
Sum		33		105		48		15		201

Based on Table 1, instructors employed significantly more collaborative activities to access higher levels of cognitive learning. Table 2 shows that, while multiple-choice is used more for Level 1 assessments, the others are more evenly distributed, $c^2(6, N=201)=17.449$, $p=.008$. Figure 1, below, illustrates the average times (in seconds) spent using different interactions. A one-way ANOVA analysis shows a significant interaction between activity type and duration, $F(4, 118)=9.71$, $p<.01$.



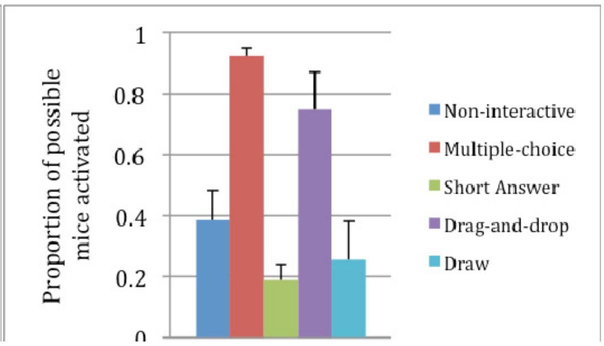


Figure 1. Average time spent using different activities. Error bars represent standard error.

Figure 2. The proportion of mice activated out of the total number available in the class, across different activities. Error bars represent standard error.

One unexpected finding from this analysis is that multiple-choice activities are not significantly longer than non-interactive activities. Non-interactive slides could be instructional material where the instructor is teaching and activating students for various reasons during instruction. Figure 6 shows the proportion of students activated across activity types. A one-way ANOVA analysis shows a significant interaction between activity type and proportion of mice activated, $F(4, 118)=47.316$, $p<.01$.

Results and Conclusion

Instructors designed far more low-level activities. For those activities, they used designs meant for individual assessment. For higher-level forms of assessment, on the contrary, they tended to create multiuser activities. This begs the design of more multiuser formative assessments. In classes where the system was used, they used it frequently and spent significantly more time on multiuser activities than on individualistic activities. Instructors were 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 instructors are creating collaborative activities that are not supported by existing features.

References

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

Bloom, B., Englehart, M. Furst, E., Hill, W., Krathwohl, D. (1956). Taxonomy of educational objectives: The classification of educational goals: Cognitive domain. Longmans, Green.

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.