Wetzel, J., Burkhardt, H., Cheema, S., Kang, S., Pead, D., Schoenfeld, A. H., & VanLehn, K. (2018). A preliminary evaluation of the usability of an AI-infused orchestration system. In C. Rosé, R. Martínez-Maldonado, U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (Eds.), Artificial Intelligence in Education: Proceedings of the 19th International Conference (pp. 378-383). Berlin: Springer.

A preliminary evaluation of the usability of an AI-infused orchestration system

Jon Wetzel[1], Hugh Burkhardt[2], Salman Cheema[1], Seokmin Kang[1], Daniel Pead[2], Alan Schoenfeld[3], and Kurt VanLehn[1]

[1] Arizona State University, Tempe, AZ 85287, USA
[2] University of Nottingham, Nottingham, UK
[3] University of California, Berkeley, Berkeley, California, USA
jwetzel4@asu.edu

Abstract. Artificial intelligence (AI) holds great promise for improving classroom orchestration—the teacher's management of a classroom workflow that mixes small group, individual, and whole class activities. Although we have developed an orchestration system, named FACT, that uses AI, we were concerned that usability issues might decrease its effectiveness. We conducted an analysis of classroom video recordings that classified and compared the time FACT students spent to the time spent by students using paper versions of the same lessons. FACT wasted half the time that paper did. However FACT students spent slightly more time off task and had difficulties referring to objects on shared documents.

Keywords: orchestration, usability, digital media

1 Introduction

Classroom orchestration refers to a teacher's management of classroom workflows that involve small group, individual, and whole-class activities [1, 2]. An orchestration system helps by increasing teacher awareness and facilitating management of the workflow [3]. Combining orchestration features with an intelligent tutoring system (ITS) should both facilitate orchestration and enhance adoption of the ITS [4]. We have iteratively developed, over 50 classroom trials, an AI-infused orchestration system, FACT (Formative Assessment Computing Technologies) [5-8]. Although FACT employs traditional ITS technology, its primary function is to help the teacher orchestrate a lesson. It is similar to Lumilo [9], MT Classroom [10] and Group Scribbles [11].

However, before the benefits of AI can be evaluated in any such system, the usability of the system vs. paper baseline classes must be assessed, just as Hao [12] did for Group Scribbles. Otherwise, the AI might be blamed when it is the usability of the media that is flawed. This paper presents a preliminary evaluation of the usability of FACT compared to paper.

The FACT user's experience. Students may use a desktop, laptop, or tablet with a web browser. Teachers should carry a tablet so that they can access FACTs dashboard as they circulate. When the class begins, everyone logs in, and the teacher can select a lesson and step through a series of activities. Whenever the teacher selects an activity, FACT gives the appropriate digital poster to the students. Figure 1 shows a poster with 6 cards on it. Students can edit both cards and the poster itself. They can draw, type, move, or erase. FACT supports group work as well as individual work. Once students have joined a group, they can edit the group's poster simultaneously in real time using conventions similar to other online collaborative editors like Google Docs.

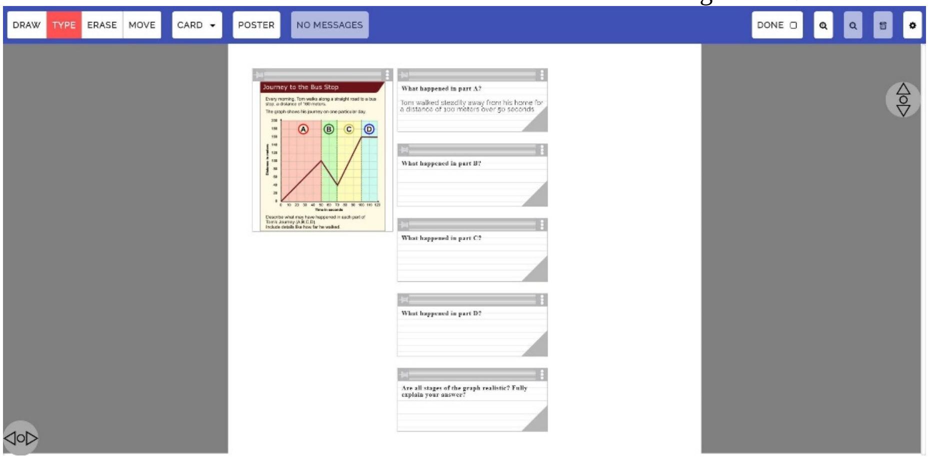

Fig. 1. A digital poster as displayed by a student's editor in FACT

FACT's intelligence. FACT monitors the students' edits and updates the teachers' dashboard in real time to show progress and alerts. The alerts are driven by a variety of technology including AI, image analysis, sketch recognition, handwriting recognition, and collaboration detection. Until recently, FACT used a human-in-the-loop policy, so all actions were taken by the teacher. However, teachers in our trials still had too many students to visit, so we have begun to experiment with human-ON-the-loop policies, where FACT sometimes takes actions by itself given the teachers' prior or concurrent approval.

2 An evaluation of FACT's usability

Our evaluation aims to answer two questions. First, how does the amount of wasted time in lessons done with FACT compare to those done on paper? Second, what are the specific usability defects of FACT, and how frequently do they occur?

We analyzed video data from 13 lessons: six done on paper and seven done with FACT. The videos were taken in classrooms in middle/high schools in Nottingham, England and the South San Francisco bay area. The grade level varied, but primarily consisted of 6th grade students. Each class we observed was working on one of eight lessons selected from the Mathematics Assessment Project (MAP, http://map.mathshell.org) Classroom Challenges.

Although the classes typically had 25 students, our data source for this study was videos taken by fixed cameras positioned over two pairs of students per class. We coded the behavior of just one student per pair, typically the student whose workspace was most clearly visible.

The coders watched for and coded the following events: **Disconnected –** Student lost their connection to FACT or the internet; **JoiningGroups –** Student was at the Join Groups screen; **LearnToResizeCard** – Student was figuring out how to resize a card for the first time; **LostCardSearch –** Student was looking for a specific card; **OtherGlitch** – Student's time was wasted by some glitch in FACT; **OtherStruggleWithMedia –** Student was having trouble using FACT but not due to a glitch.

If the medium was paper, the following codes were used: **PaperNoMath** – Student was either a) manipulating traditional media without discussing math, b) discussing traditional media, c) waiting for their partner to manipulate traditional media without engaging in math, or d) waiting for media to distributed; **LostCardSearch** – Student was looking for a specific card. Here, "traditional media" refers to the tools: paper, cards, glue, pencils/pen, etc.

When coders found one of these situations, they entered the code, start time, end time, and an optional brief description of the event in a spreadsheet. To calculate the amount of wasted time due to the medium for a student in a given lesson, we summed the durations as shown in Table 1.

Our analysis found that using FACT resulted in less time wasted due to the medium. On average 12.45% (8m09s) of the student's time was wasted manipulating or waiting for paper, while only 6.30% (3m56s) was wasted when using FACT. That is, the FACT wasted half the time that paper wasted.

Table 1. Average time spent on key event for students in lessons (m:ss format)

Code	Avg Time Used (FACT)	Avg Time Used (Paper)
LoggingIn	1:22	N/A
Disconnected	0	N/A
JoiningGroups	0:51	N/A
LearnToResizeCard	0:07	N/A
OtherGlitch	1:18	N/A
OtherStruggleWithMedia	0:18	N/A
LostCardSearch	0:01	1:07
PaperNoMath (distributing paper)	N/A	1:23
PaperNoMath (other)	N/A	5:39
Avg. wasted time (% lesson time)	3:57 (6.30%)	8:09 (12.45%)

Table 1 summarizes the sources of wasted time. Time at the login screen was between 10s and 4m49s, because some teachers had students wait there while introducing the lesson. Glitches were a close second on wasted time for FACT. All but one instance of JoiningGroups were finished within 45 seconds. During the outlier event, which took 2m20s, most students joined their group within a few seconds, but then had to wait for the rest of the class. The most notable case of OtherStruggleWithMedia occurred when a student tried for 30+ seconds to move a card with their cursor in erase mode.

We also coded both the FACT and paper videos for off-task behavior. Students using FACT spent slightly more time off task than those on paper: average time 2m34s (3.9% of class time) for paper vs 3m20s (5.3% of class time) for FACT.

Table 2. Counts of Instances of co-referring by students in FACT lesson

Student	ReferPoint	ReferID	ReferOther	ReferShow
1	11	3	1	1
2	3	0	1	0
3	3	0	6	0
4	8	0	0	2
5	3	0	0	1
6	0	17	0	0
7	7	0	0	1
% of instances	51.47%	29.41%	11.76%	7.35%

Hao [12] and others have noticed that when students want to refer to an object in a shared document, they often try to point to their own screen and their partners cannot see what they are pointing at. This could impede collaboration. To evaluate this issue, we coded: **ReferPoint –** The student being coded pointed at the other student's tablet; **ReferShow –** The student showed their own tablet to the partner and pointed to it; **ReferID** – The student orally mentioned the number or letter label on the card; **ReferOther –** The student used some other method. Our analysis, summarized in Table 2, found that pointing at the other person's tablet was the most frequent method. References using the ID on the card were less frequent (and all cards had IDs except in Lesson 7). On average, students spent 61 seconds per lesson getting their partners to understand what they were referring to. Presumably, referring would take much less time if students were using a paper poster instead of a shared electronic poster.

3 Conclusion

While our sample size is small, our analysis leaves us optimistic about FACT's usability. From our observations so far, we see FACT wastes less time than paper, and we identified several ways to lower its wasted time. We are addressing the largest time-waster, glitches, by continuing to fix bugs in FACT.

FACT students tend to go off-task more than paper students. It is not clear why and could be just a sampling artifact.

Referring to objects in a shared document seems to occupy a relatively short time per lesson, so we hypothesize that collaboration is unhindered. However, confirmation requires a closer study of this issue which also coded students' references to objects on paper posters.

Acknowledgements. The FACT project is sponsored by the Bill and Melinda Gates Foundation under grant OPP10612881. We also thank all the students at ASU who have worked on FACT and the video analyses. Finally, we thank the Silicon Valley Mathematics Initiative and the Shell Center for conducting the classroom testing.

References

1 Dillenbourg, P.: ‘Design for classroom orchestration’, Computers & Education, 2013, 69, pp. 485-492

2 Roschelle, J., Dimitriadis, Y., and Hoppe, U.: ‘Classroom orchestration: Synthesis’, Computers & Education, 2013, 69, pp. 523-526

3 Prieto, L.P., Dlab, M.H., Abdulwahed, M., and Balid, W.: ‘Orchestrating technology enhanced learning: A literature review and conceptual framework’, International Journal of Technology Enhanced Learning, 2011, 3, (6), pp. 583-598

4 Holstein, K., McLaren, B., and Aleven, V.: ‘Intelligent tutors as teachers' aids: Exploring teacher needs for real-time analytics in blended classrooms’. Proc. Learning Analytics and Knowledge: LAK '17, Vancouver, BC, Canada2017 pp. Pages

5 Cheema, S., VanLehn, K., Burkhart, H., Pead, D., and Schoenfeld, A.H.: ‘Electronic posters to support formative assessment’. Proc. CHI'16: Extended Abstracts2016 pp. Pages

6 VanLehn, K., Cheema, S., Wetzel, J., and Pead, D.: ‘Some less obvious features of classroom orchestration systems.’, in Lin, L., and Atkinson, R.K. (Eds.): ‘Educational Technologies: Challenges, Applications and Learning Outcomes’ (in press)

7 Viswanathan, S.A., and VanLehn, K.: ‘Using the tablet gestures and speech of pairs of students to classify their collaboration’, IEEE Transactions on Learning Technologies, in press

8 VanLehn, K., Burkhardt, H., Cheema, S., Pead, D., Schoenfeld, A.H., and Wetzel, J.: ‘How can FACT encourage collaboration and self-correction?’, in Millis, K., Long, D., Magliano, J., and Wiemer, K. (Eds.): ‘Multi-Disciplinary Approaches to Deep Learning’ (Routledge, in press)

9 Holstein, K., Hong, G., Tegene, M., McLaren, B., and Aleven, V.: ‘The classroom as a dashboard: Co-designing wearable cognitive augmentation for K-12 teachers’. Proc. Proceedings of the International Conference on Learning Analytics and Knowledge, Sydney, Australia2018 pp. Pages

10 Martinez-Maldonado, R., Yacef, K., and Kay, J.: ‘TSCL: A conceptual model to inform understanding of collaborative learning processes at interactive tabletops’, International Journal of Human-Computer Studies, 2015, 83, pp. 62-82

11 Looi, C.-K., Lin, C.-P., and Liu, K.-P.: ‘Group Scribbles to support knowledge building in a jigsaw method’, IEEE Transactions on Learning Technologies, 2008, 1, (3), pp. 157–164

12 Hao, C.F.: ‘A comparative study of collaborative learning in Paper Scribbles and Group Scribbles’, Australasian Journal of Educational Technology, 2010, 26, (5), pp. 659-674