Auto-Sending Messages in an Intelligent Orchestration System: A Pilot Study

Kurt VanLehn[1](✉), Salman Cheema[2], Seokmin Kang[1], and Jon Wetzel[1]

[1] Arizona State University, Tempe, AZ, USA
kurt.vanlehn@asu.edu
[2] Microsoft, Redmond, OR, USA

Abstract. FACT (Formative Assessment with Computational Technology) is an intelligent orchestration system. That is, because it helps the teacher manage the workflow of a complicated set of activities in the classroom, it is an orchestration system. Because it conducts tasks-specific and domain-specific analyses of the students' mathematical products and their group interactions, it is more intelligent than other orchestration systems. From analyzing videos of our iterative development trials, we realized that too many students needed help simultaneously, but the teacher could only visit one group at a time. Thus, we modified FACT to send a few messages to the students directly instead of sending all its advice to the teacher. This paper reports a successful pilot test of auto-sending.

Keywords: Classroom orchestration systems · Formative assessment · Intelligent tutoring system · Classroom evaluation

1 Introduction

Some lesson plans involve individual work, group work and whole-class discussions, and some also require that the teacher integrate workflows and ideas across all three planes of activity. "Classroom orchestration" refers to the planning and enacting of such integrated workflows [1]. A "classroom orchestration system" is intended to help the teacher with classroom orchestration [2–18].

Our system [19–23] was designed to increase the effectiveness of a particular set of mathematics lessons called the Classroom Challenges [24]. In their paper-based form, the Classroom Challenges (CCs) are known to be highly effective [25]. They exemplify teaching based on formative assessment [26], wherein teachers no longer give explanations and feedback, but instead keep students engaged in solving problems.

The CC students spend most of their time working in small groups on large posters, to which they add cards and handwriting. The posters can become extremely complicated. When teachers are circulating among the groups and they stop to visit a group, they often have only seconds to conduct a formative assessment of a complex poster.

We hypothesized that difficulties in formative assessment were preventing the CCs from being even more effective. Thus, the original goal of the FACT system was to

S. Isotani et al. (Eds.): AIED 2019, LNAI 11626, pp. 292–297, 2019.
https://doi.org/10.1007/978-3-030-23207-8_54

conduct and display formative assessments of posters in order to help visiting teachers. Thus, it was named Formative Assessment with Computational Technology (FACT).

FACT students edit an electronic document called a poster. Posters can have movable cards on them. Students can write or draw on the cards or the poster with a stylus, finger, mouse or keyboard. Students can also move the cards, pin them or resize them.

Students can edit their own individual poster or their group's poster. When editing a group poster, all the members of the group can edit simultaneously, just as one does with a shared Google document. Each student's ink is a different color, so students and teachers can tell who has contributed what.

As students work, teachers can monitor their work and control the class. They carry a tablet around the classroom that displays FACT's dashboard.

To conduct a formative assessment of students' work, FACT has many issue detectors. Most of them compare the students' work to expected work; these are called *product* detectors. FACT also has *process* detectors. These raise issues about the chronological pattern of students' edits, such as failing to collaborate. Similar collaboration detectors were quite accurate when used in a lab study [27].

FACT constantly decides which active issues are most important and shows them as alerts on the teacher's dashboard. When teachers peek at a student's work (i.e., view the student's poster on their dashboard), they see the top priority issue in a sidebar. They can scroll to view other issues in the sidebar. Each issue has both an explanation of it and questions that teachers can ask the student in order to open a visit discussing it. Alternatively, the teacher can push a Send button next to one of the questions, and it will appear in the student's inbox.

In order to help design FACT, video data from 14 trials of paper-based CCs were collected. During the iterative development of FACT, video data from 52 trials were collected. The later videos were collected as a formative evaluation rather than a summative evaluation. That is, they were collected to help us understand and redesign FACT. Nonetheless, we compared the videos of Paper and FACT trials and found:

1. FACT students wasted less time than the Paper students (5.9% for FACT vs. 10.4% for Paper; $p = 0.013$). This was clearly due to replacing paper with electronic documents.
2. FACT students spent more time off-task than the Paper students (5.7% for FACT vs. 2.9% for Paper; $p = 0.011$), probably due to the novelty of the stylus-tablet user interface.
3. FACT groups and Paper groups did not differ in how they worked together. Both FACT and Paper groups worked silently most of the time (53.8% for FACT vs. 67.7% for Paper). Groups rarely engaged in the most desirable form of collaboration, called co-construction or transactivity (2.8% for FACT vs. 4.0% for Paper).
4. FACT students self-corrected 47% of their errors, whereas Paper students self-corrected 67% of their 12 errors ($p < .001$). Other errors were either left incorrect or corrected with the aid of the teacher. This suggests that productive struggle was more frequent for Paper students, contrary to our expectations.

5. When pairs were classified according the amount of self-correction of errors, 39% of the FACT pairs were struggling productively vs. 63% of the Paper pairs.
6. The mean number of teacher visits per lesson did not differ (27.8 for FACT vs. 27.2 for Paper; p = 0.890), nor did the mean time between visit starts (4:25 for FACT vs. 5:40 for Paper; p = 0.182).

The figures above indicate that many groups were not productively struggling and almost all were not collaborating properly. Yet teachers visited few groups (one per 4 or 5 min). Thus, when a teacher finished one visit and was deciding whom to visit next, *almost every group in the class needed to be visited*. Even if FACT helps the teacher make an optimal choice of whom to visit and what to say, there are many other groups left without a visit. Perhaps it would help if FACT could "visit" groups, too.

2 Auto-Sending and Its Pilot Test

As mentioned earlier, when teachers Peek at a student, they see a sidebar that shows questions that the teachers can use to initiate a visit. Teachers can also push a Send button to send a question directly to students. It then appears as a message in the student's inbox.

In order to increase its effectiveness, FACT was modified to, so to speak, push the Send button pushes itself. After an activity began, it waited 5 min so students could get well started. It would then send students a message from their highest priority issue. It would always wait at least 2 min between sending messages to a group. We called this policy "auto-sending."

As a pilot test of auto-sending, we conducted an AB evaluation in the middle school math classes of 2 teachers. Three classes had the full FACT system. Two classes had FACT with its detectors turned off, which meant that the teachers saw alerts neither on the dashboard nor when Peeking, and FACT auto-sent no messages.

We used the same methods and measures as in the formative evaluation reported earlier. The pattern of results during this pilot test were similar to those reported earlier, except for the most critical outcome, productive struggle, so we report just those results.

In order to help determine what encouraged students to be differentially productive, we divided all errors into four categories:

- The teacher visited the group when the error was being corrected or within the preceding 30 s.
- The students read a message in their inbox during the 30 s preceding correction of the error. The message could have been sent either by the teacher or by FACT.
- The students corrected the error without having consulted their inbox or the teacher during the preceding 30 s.
- The error had not yet been corrected when the activity ended.

Table 1 shows the error counts per condition per category. Comparing the On vs. Off conditions, the error distributions are reliability different (Chi-square, $p < .001$). Students in the detectors On condition corrected significantly more errors without help from FACT or the teacher.

As in the formative evaluation, we classified pairs that corrected more than 50% of their errors by the end of the activity as productive. By this somewhat arbitrary criterion, all 7 pairs in the Analysis On classes were productive, while in the Analysis Off classes, only 3 of the 6 pairs were productive. This difference is reliable (Chi-sq, p = 0.004). This is consistent with the hypothesis that turning the detectors on increased productive struggle.

3 Discussion

Summary: While iteratively developing FACT with aid of teachers, students and classroom observers, we recorded videos of 52 FACT classes and 14 paper-based CC classes. Video analyses suggest that although FACT made the workflow more efficient, there appeared to be little change in group interaction and teacher behavior. Contrary to our ambitions, FACT decreased productive struggle in the groups. The problem appeared to be simply that there weren't enough teacher visits to students because there is only one teacher but almost all groups need visits. Thus, we modified FACT to automatically send the messages that teachers could send. We compare two versions of FACT, with detectors turned either On or Off. Groups in the On condition more frequently struggled productively than groups in the Off condition. This is consistent with our hypothesis that the bottleneck in our classes is that more groups need to be visited, and that FACT's auto-send feature can at least partially fill the gap.

Although we refer to the conditions as detectors On vs. Off, many other factors co-varied with the manipulation including the classes, the time of day and the familiarity of the teachers with FACT. Thus, we cannot conclude that turning the detectors on *caused* students to correct more errors. Better-controlled experiments with more classes and teachers are needed.

A second problem is that errors are only one sign of struggle. Struggle could also show up as slow speed or extensive discussion.

In future work, the teacher's visits and the system's messages should be coordinated closely in order keep the teacher in charge of the class and yet maximize the impact on students. FACT will need a new kind of intelligence in order to support this sort of coordination. A larger, better-controlled evaluation would also be important.

Table 1. Errors

Correction type	Off	On
Teacher-assisted	0	0
FACT-assisted	0	4
Self-corrected	3	13
Uncorrected	4	1

Acknowledgements. This research was supported by the Bill and Melinda Gates Foundation under OPP1061281, the Diane and Gary Tooker Chair for Effective Education in Science, Technology, Engineering and Math and NSF grant 1840051. We gratefully acknowledge the contributions of all the members of the FACT project, past and present.

References

1. Dillenbourg, P., Jermann, P.: Technology for classroom orchestration. In: Khine, M.S., Saleh, I.M. (eds.) New Science of Learning: Cognition, Computers and Collaboration in Education. Springer, New York (2010). https://doi.org/10.1007/978-1-4419-5716-0_26
2. Prieto, L.P., Dlab, M.H., Abdulwahed, M., Balid, W.: Orchestrating technology enhanced learning: a literature review and conceptual framework. Int. J. Technol. Enhanced Learn. **3** (6), 583–598 (2011)
3. Holstein, K., McLaren, B.M., Aleven, V.: Student learning benefits of a mixed-reality teacher awareness tool in ai-enhanced classrooms. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10947, pp. 154–168. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93843-1_12
4. Molenaar, I., Knoop-van Campen, C.A., Hasselman, F.: The effects of learning analytics empowered technology on students' arithmetic skill development. In: Proceedings of the Learning Analytics and Knowledge LAK 2017, Vancouver, BC, Canada, pp. 614–515 (2017)
5. Molenaar, I., Knoop-van Campen, C.A.: Teacher dashboards in practice: usage and impact. In: Proceedings European Conference on Technology Enhanced Learning: EC-TEL 2017, pp. 125–138 (2017)
6. Håklev, S., Faucon, L., Hadzilacos, T., Dillenbourg, P.: FROG: rapid prototyping of collaborative learning scenarios. In: Proceedings of the EC-TEL 2017 (2017)
7. Haklev, S., Faucon, L., Hadzilacos, T., Dillenbourg, P.: Orchestration graphs: enabling rich social pedagogical scenarios in MOOCs. In: Proceedings of the Fourth (2017) ACM Conference on Learning@Scale, pp. 261–264 (2017)
8. van Alphen, E., Bakker, S.: Lernanto: using an ambient display during differentiated instruction. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2334–2340. ACM (2016)
9. Mercier, E.: Teacher orchestration and student learning during mathematics activities in a smart classroom. Int. J. Smart Technol. Learn. **1**(1), 33–52 (2016)
10. Martinez-Maldonado, R., Yacef, K., Kay, J.: TSCL: a conceptual model to inform understanding of collaborative learning processes at interactive tabletops. Int. J. Hum Comput Stud. **83**, 62–82 (2015)
11. Martinez-Maldonado, R., Clayphan, A., Yacef, K., Kay, J.: MTFeedback: providing notifications to enhance teacher awareness of small group work in the classroom. IEEE Trans. Learn. Technol. **8**(2), 187–200 (2015)
12. Berland, M., Davis, D., Smith, C.P.: AMOEBA: designing for collaboration in computer science classrooms through live learning analytics. Int. J. Comput. Support. Collaborative Learn. **10**, 425–447 (2015)
13. Prieto, L.P., Asensio-Perez, J.I., Munoz-Cristobal, J.A., Jorrin-Abellan, I.M., Dimitriadis, Y., Gomez-Sanchez, E.: Supporting orchestration of CSCL scenarios in web-based distributed learning environments. Comput. Educ. **73**, 9–25 (2014)
14. Balestrini, M., Hernandez-Leo, D., Nieves, R., Blat, J.: Technology-supported orchestration matters: outperforming paper-based scripting in a jigsaw classroom. IEEE Trans. Learn. Technol. **7**(1), 17–30 (2014)
15. Higgens, S., Mercier, E., Burd, E., Joyce-Gibbons, A.: Multi-touch tables and collaborative learning. Br. J. Educ. Technol. **43**(6), 1041–1054 (2012)
16. Do-Lenh, S.: Supporting reflection and classroom orchestration with tangible tabletops. Ecole Polytechiuqe Federale de Lausanne (2012)

17. Alavi, H., Dillenbourg, P.: An ambient awareness tool for supporting supervised collaborative problem solving. IEEE Trans. Learn. Technol. **5**(3), 264–274 (2012)
18. Looi, C.-K., Lin, C.-P., Liu, K.-P.: Group scribbles to support knowledge building in a jigsaw method. IEEE Trans. Learn. Technol. **1**(3), 157–164 (2008)
19. VanLehn, K., Cheema, S., Wetzel, J., Pead, D.: Some less obvious features of classroom orchestration systems. In: Lin, L., Atkinson, R.K. (eds.) Educational Technologies: Challenges, Applications and Learning Outcomes, pp. 73–94. Nova Scientific Publishers (2016)
20. VanLehn, K., et al.: The effect of digital versus traditional orchestration on collaboration in small groups. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10948, pp. 369–373. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93846-2_69
21. VanLehn, K., Burkhardt, H., Cheema, S., Pead, D., Schoenfeld, A.H., Wetzel, J.: How can FACT encourage collaboration and self-correction?. In: Millis, K., Long, D., Magliano, J., Wiemer, K. (eds.) Multi-Disciplinary Approaches to Deep Learning, pp. 114–127. Routledge (2018)
22. Wetzel, J., et al.: A preliminary evaluation of the usability of an ai-infused orchestration system. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10948, pp. 379–383. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93846-2_71
23. VanLehn, K., et al.: Can an orchestration system increase collaborative, productive struggle in teaching-by-eliciting classrooms?. Interactive Learning Environments, in press
24. http://map.mathshell.org/index.php
25. Herman, J., et al.: The implementation and effects of the Mathematics Design Collaborative (MDC): early findings from Kentucky ninth-grade algebra 1 courses (CRESST Report 845). University of California at Los Angeles, National Center for Research on Evaluation, Standards and Student Testing (2015)
26. Black, P., Wiliam, D.: Assessment and classroom learning. Assess. Educ. Principles Policy Pract. **5**(1), 7–75 (1998)
27. Viswanathan, S.A., VanLehn, K.: Using the tablet gestures and speech of pairs of students to classify their collaboration. IEEE Trans. Learn. Technol. **11**(2), 230–242 (2018)