VanLehn, K., Burkhardt, H., Cheema, S., Pead, D., Schoenfeld, A. H., & Wetzel, J. (2018). How can FACT encourage collaboration and self-correction? In K. Millis, D. Long, J. Magliano, & K. Wiemer (Eds.), Multi-Disciplinary Approaches to Deep Learning. New York, NY: Routledge.

9

HOW CAN FACT ENCOURAGE COLLABORATION AND SELF-CORRECTION?

Kurt VanLehn, Hugh Burkhardt, Salman Cheema, Daniel Pead, Alan Schoenfeld, and Jon Wetzel

This chapter describes some atypical mathematics instruction and our attempts to support it with an intelligent tutoring system (ITS). The instruction encourages collaboration and a particular type of deep thinking that we call self-correction. We first describe the instruction and its success. Then we describe our system and how it differs from other ITS.

The Classroom Challenges

The Classroom Challenges, which were developed by the Mathematics Assessment Project (MAP), are a set of about 100 middle school and high school math lessons. Each lesson consists of pdfs for printing student handouts, PowerPoint slides, and a pdf teacher's guide. They can be download for free from the MAP website (http://map.mathshell.org/lessons.php). Each lesson takes about 90 to 120 minutes and is often enacted over two days. Lessons do not form a sequence. Different teachers pick different lessons to fit their curriculum.

The Classroom Challenges (CCs) have a number of distinctive design features. Broadly speaking, they are aligned with the Teaching for Robust Understanding framework (Schoenfeld, 2014, 2017) which describes classrooms from which students emerge as knowledgeable and resourceful thinkers and problem solvers. The CCs provide significant opportunities for students to practice extended mathematical reasoning, problem solving, and modeling. CC students engage in collaborative activities that provide opportunities for productive mathematical discourse, which in turn provide equitable access to the content and opportunities to develop increasingly positive mathematical identities.

A year-long study involving 17 schools and 2316 students in Kentucky showed that the CCs were successful (Herman et al., 2014; Herman et al., 2015). The teachers enacted between four and six CCs during the course of the school year.

Their students' scores on the state-wide, standardized math achievement tests were reliably higher than a matched control group. The effect size corresponded to an extra 4.6 months of schooling. The evaluators concluded (p. 30): "The effect is quite dramatic given that a typical teacher spent less than a month of class time teaching Challenges."

In order to find out why the CCs were so successful, Inverness Research interviewed and surveyed CC teachers and students (InvernessResearch, 2016). Although survey results for the students were positive but "not dramatically affirmative," interviews with the teachers indicated that, "Their experiences were overwhelmingly positive." These interviews make it clear why such a short intervention had such a large impact on standardized math achievement scores: The teachers said that the CCs caused them to *modify their teaching of all their other math lessons* to make them more like the CCs.

The enthusiasm of teachers for the CCs is affirmed by other evidence. There have been over 7 million downloads of CCs as of the end of 2016. Surveys of CC teachers were overwhelmingly positive (Research_for_action, 2015). Several professional development programs use the CCs as examples of exemplary instruction (MDC, 2016). Studies of formative assessment have used the CCs as their content (FaSMEd, 2017).

Even though the CCs were already quite successful, the Bill and Melinda Gates Foundation challenged us to find a way to use ITS technology to improve their effectiveness even further.

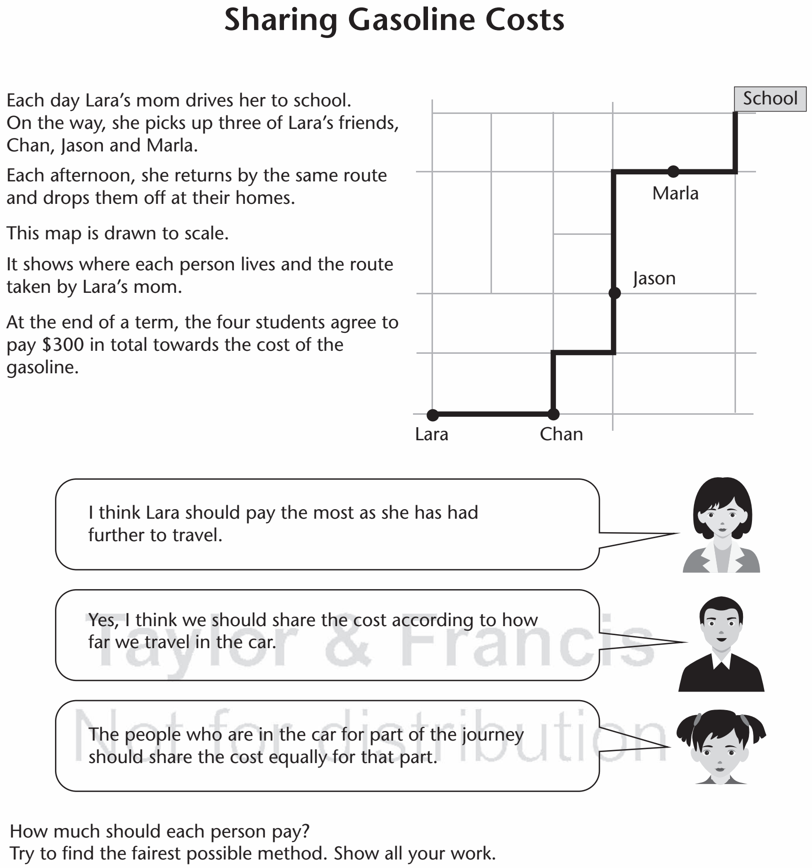
Problem-Solving CCs

There are two types of CCs: problem solving and concept development. The problem-solving lessons will be described first.

A problem-solving CC consists of a main lesson which is preceded by a brief pre-assessment and followed by a brief reflective review. During the pre-assessment, students work individually on a challenging problem, such as the one shown in Figure 9.1. The main lesson consists of a sequence of activities, such as:

1. Working in small groups, students first review each other's solutions from the pre-assessment and then formulate a group solution.
2. The whole class reviews and discusses some groups' solutions.
3. One member from each group visits another group. The visitor and the visited group discuss differences between their group's solutions.
4. Working in small groups, students critique solutions from three hypothetical students.
5. The whole class reviews and discusses some groups' critiques.

During the reflective review, students again work individually, typically by solving again the same problem and comparing their new solution to their work on the pre-assessment.



Sharing Gasoline Costs

Each day Lara's mom drives her to school.
On the way, she picks up three of Lara's friends, Chan, Jason and Marla.

Each afternoon, she returns by the same route and drops them off at their homes.

This map is drawn to scale.

It shows where each person lives and the route taken by Lara's mom.

At the end of a term, the four students agree to pay $300 in total towards the cost of the gasoline.

I think Lara should pay the most as she has had further to travel.

Yes, I think we should share the cost according to how far we travel in the car.

The people who are in the car for part of the journey should share the cost equally for that part.

How much should each person pay?
Try to find the fairest possible method. Show all your work.

FIGURE 9.1 A Problem-Solving CC's Problem

The problem solving CCs differ from a typical ITS problem-solving lesson in several ways. First, students focus on the same problem for almost 2 hours. Second, they first work on a problem *alone*, then discuss their solutions *in pairs*, then as a *whole class*. Third, a student sees many solutions: from other group members, from other groups, and from hypothetical students. Fourth, the problems afford many solutions. For instance, one solution to the problem of Figure 9.1 is just to divide the costs evenly among the four riders. As students discuss their solutions, this one is usually abandoned in favor of those that conform to the ideas expressed in the talk bubbles of the problem, not because this solution is *incorrect*, but because there are other solutions that make more sense given the context of the problem.

Concept Development CCs

Whereas a problem-solving CC has students solve a single problem many times, a concept development CC has students solve many similar problems that all use the same concepts. Unlike a problem-solving CC, there is only one correct solution to each concept development CC, although there are often multiple ways of arriving at or justifying that one correct solution.

Concept development problems are presented on cards. Sometimes students express a solution to a problem by placing the problem card on a certain region of a poster. For example, one CC gives students cards with an equation in a single variable, such as 2(x + 1) = (4 + 4x)/2, and asks them to place ~~the~~ card in one of three columns: always true, sometimes true or never true. Other concept development CCs have students make clusters comprised of two or three cards. For example, Figure 9.2 shows clusters comprised of a story card and a distance-time graph card. The story card for "Tom walked up a hill, walked across the flat top, and then slid down the steep side," is often paired with the graph card in the middle of the second row of Figure 9.2, which looks like a side view of the hill. The correct card for this story is the first one on the second row. Almost all the concept development CCs involve arranging cards on posters. This allows students to change their minds many times without having to write and erase copiously.

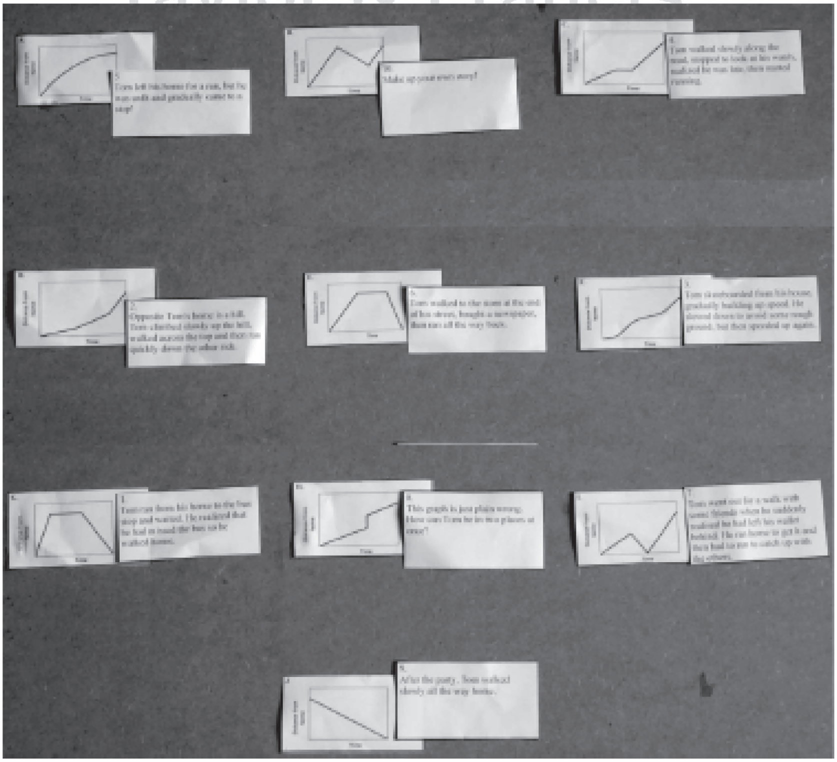


FIGURE 9.2 Solution to a Concept Development CC

How Teaching a CC Differs From Ordinary Teaching

A typical math lesson starts with the teacher explaining a concept or a problem-solving method, then having students practice applying the new ideas while the teacher circulates, offering feedback. The feedback is corrective. For instance, suppose teachers see a student making the mistake just mentioned, of pairing a story about Tom traversing a hill with a side view of his route. They would start with hints. If the hints fail, they would explain or reteach the relevant ideas. They would persist until the mistake is corrected, even if they have to reveal the correct card pairing themselves. Human tutors and ITS also give hint sequences intended to correct mistakes (Hume, Michael, Rovick, & Evens, 1996; VanLehn, 2006). When the goal is to get students to apply mathematical ideas correctly, then explanations and corrective feedback are a common instructional method.

However, the goal of the CCs is to get students to solve problems and clarify mathematical ideas *without help from teachers or other experts*. Instead of trying to comprehend their teachers' ideas, students need to draw on the ideas of their peers and their own embryonic ideas. This is the kind of problem solving and learning that they will need to do when they have left schools and teachers behind.

It is also the kind of problem solving that students do during a test, where students get neither corrective feedback nor explanations. However, unlike a normal assessment, which just provides teachers with information they need to make decisions, this kind of problem solving is intended to foster learning. Thus, it is often called formative assessment or *assessment for learning* (Black & Wiliam, 1998).

Although the *instruction* has two standard names, there is no standard name for the student *cognition* that it encourages. We will refer to it as "self-correction." This emphasizes that students needs to reflect on their thinking, explain it to others, listen carefully to others and find a correct solution as a group without resorting to any higher authority. Thus, the goal of the CCs is to engage students in self-correction. This is a kind of deep learning, which is why this chapter belongs in this book.

After posing the problem, CC teachers should, ideally, say nothing more as the students figure out solutions themselves. However, students often do not engage spontaneously in self-correction, so teachers need to guide them. Also, students can get stuck on a particular way of viewing a problem, so teachers sometimes need to help them break their mindsets and look at the problem differently.

For example, suppose a teacher visits a group and sees the mistake mentioned earlier (involving Tom and a hill). The teacher definitely should *not* give a hint sequence. Instead, the teacher should ask the group to explain their work. Typically, the student who created that card pair will respond. The teacher would then ask another group member to explain. If the second member can't explain, then the teacher has uncovered a process failure—the group has not been collaborative enough—so the teacher might remind them to work together more closely. On the other hand, suppose the second member can explain the card pair and further questions indicate widespread acceptance of the view that graphs are pictures of Tom's route seen from the side, from overhead, etc. Now the teacher has uncovered an entrenched mindset. As it turns out, the teachers' guide for this CC has a list of common misconceptions, and this is one of them. For this misconception, it suggests several questions to ask the students, such as "What would the graph look like if Tom walked in a circle around his house?" The questions are intended to challenge the students' mindset.

In the previous illustration, both teacher-student interactions were feedback, but they were definitely not standard, corrective feedback. This kind of feedback is probably part of the strong appeal of the CCs. Teachers discover that they can indeed help students without undermining self-correction.

How Can the Classroom Challenges Be Improved?

One well-accepted theory of the relationship between instruction and learning gains is Michelene Chi's ICAP framework (Chi & Wylie, 2014). The acronym stands for Interactive, Constructive, Active, and Passive. These are major observable categories of student performance, ordered from most effective to least effective. The CCs encourage Constructive and especially Interactive performance.

However, our analysis of the videos of small groups during the CCs indicates that less than 5% of the group's time is spent on Interactive performance. Students often move the cards or write without any discussion at all; or their discussion is too shallow; or one student does all the substantive talking. So, one way to improve the CCs would be to increase the amount of Interactive performance.

Although the CCs are designed so that students can find the mathematically correct solutions by themselves, many students fail to do so. After finding an incorrect solution, they often stop, convinced that they are done. When groups examine each other's solutions, they often do not examine the differences critically; instead, they assume that both solutions are fine. Students can get stuck on an entrenched mindset. Thus, a second way to improve the CCs is to encourage self-correction by encouraging students to re-examine their solutions from different perspectives.

It is difficult for teachers to encourage Interaction and self-correction because determining when these key processes should occur and whether they did occur requires that teachers observe students over a longer period of time than they typically can afford. In particular, when the teacher walks up to a group and views their poster, the teacher cannot see whether the students created it collaboratively or not. The teacher might see some flaws in the solution, but the teacher doesn't know whether the students were engaged in self-correction or not. Thus, teachers cannot tell which students need which kind of guidance.

This suggests a role for technology. It can help teachers by monitoring the students' collaboration and self-correction. It can suggest which students to visit and what to say to them. It is often said that good teachers have eyes in the backs of their heads. Perhaps technology could give them a pair of eyes for each student group.

The FACT System

As Cuban (1993) pointed out, the only widely used technology in classrooms is the projector (transparency projectors in 1993; digital projectors nowadays). Technology that attempts to guide teachers' instruction is often abandoned as soon as it gets in the teachers' way. Projectors stay in the classroom because they don't constrain the teachers' teaching.

We divided the FACT system into two parts. The FACT Media System is intended to be like a projector in that it allows teachers to teach anything. The FACT Analysis System is a regulatory loop (Soller, Martinez, Jermann, & Muehlenbrock, 2005; VanLehn, 2016) for the CCs. It has expectations about how students should perform, detectors for sensing the actual performances, and messages that it can send to teachers and students. This cybernetic loop is intended to improve enactment of the CCs. The FACT Media System will be described first.

The FACT Media System

By definition, a system that helps teachers with all phases of their job in the classroom is called an *orchestration system* (Dillenbourg, 2013, pg. 485). Although Dillenbourg's group and others have developed novel and interesting orchestration systems for special activities, perhaps the most well-developed *general-purpose* orchestration system to date is Group Scribbles (Lin, Chen, Yang, Xiet, & Lin, 2014; Looi, Lin, & Liu, 2008; Prieto, Dlab, Abdulwahed, & Balid, 2011). The FACT Media System was inspired by Group Scribbles.

For the students, FACT looks and acts like a piece of paper. Although it can be blank, it is usually a worksheet or poster. The images can be scans of existing material, which will eventually make it simple for teachers to incorporate existing instruction or modify it. Students can write on the virtual paper with a stylus or a finger, and they can type. Although the virtual papers can be any size, we'll use "poster" for them all.

Following the successful design of Group Scribbles, students can also add cards to their ~~paper~~. They can move, resize, delete, copy, paste, and write on the cards. Posters often include some cards bearing images or text. For example, FACT has a poster bearing the cards shown in Figure 9.2.

FACT supports three types of classroom activities: small group, individual, and whole class. When students work in a small group, the group has one poster which can be viewed and edited by every group member simultaneously on their own device. When students work individually, they have a poster of their own. They can copy work among their individual posters and their group's posters.

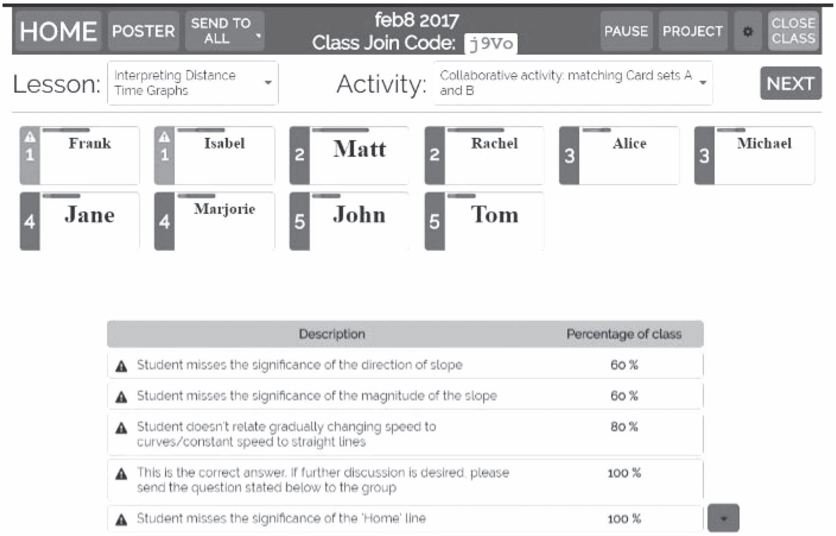
FACT provides a dashboard for teachers to carry as they circulate (see Figure 9.3). When teachers want the attention of the whole class, they can tap the Pause button on the dashboard. This causes FACT to freeze the student devices and display "Eyes on teacher." The dashboard has a tile for every student. The tile displays the students' name, the number of the group that the student is in, and a progress bar indicating how much work the student has done on the current poster.

Tapping on a tile allows the teacher to see the student's poster. This capability, called "peeking," helps the teacher decide whether to visit the student or the student's group. When teachers visit students, they often peek at the students' poster as they talk with the students. This allows them to zoom and scroll so that they can see critical parts of the poster without bending over.

For FACT, a lesson plan is simply a list of activities. An activity consists of a poster for the students to edit either as a group or individually, and some posters for the teacher to display, e.g., for introducing the activity. Teachers move the class along to another activity by tapping on a Next button or selecting from the activity menu.

The Project button allows the teacher to show their screen or any student's poster to the whole class via the classroom digital projector. The "send to all" button allows the teacher to send a message to the whole class such as "5 minutes remaining."

When class is over, the teacher can log into FACT from their office or home and review all the posters. They can write on them or paste cards on them.



Lesson: Interpreting Distance Time Graphs

Activity: Collaborative activity: matching Card sets A and B

Description	Percentage of class
▲ Student misses the significance of the direction of slope	60 %
▲ Student misses the significance of the magnitude of the slope	60 %
▲ Student doesn't relate gradually changing speed to curves/constant speed to straight lines	80 %
▲ This is the correct answer. If further discussion is desired, please send the question stated below to the group	100 %
▲ Student misses the significance of the 'Home' line	100 %

FIGURE 9.3 Teacher's Dashboard

Teachers can type onto a blank card or select a card from a menu of suggestions from the teacher's guide.

FACT tries to keep the usual overhead of computer usage as low as possible. Posters persist on the FACT server without students needing to remember to save them or name them. If a student's device crashes or the internet goes down, the students lose at most a few seconds work. Logging in and managing group membership is simple. FACT does not need to be installed or maintained because it runs inside a browser. FACT is designed to run on low-cost devices, such Chromebooks and tablets. Although FACT runs on smartphones, it is too difficult to use on such small screens.

FACT was developed over a four-year period with approximately 10 classroom trials per year. The initial designs were too elaborate and hard to use. Our trials taught us to simplify, simplify, and then simplify. For example, we initially let students select a group of cards then move or delete the whole group; students rarely discovered this feature and seldom used it. All features that remain are easily discovered and routinely used. Students need no user interface training, and teachers get by with just a few minutes of training. Teachers praise the Pause button and the ability to instantly display a student's poster to the whole class. These save considerable time compared to gaining the class's attention, calling on a student, having them walk their poster up to the document camera, and then continually reposition the poster to show various parts of it to the class. Both teachers and students are delighted not to have to cut out cards nor use tape, glue, or post-its. Teachers also like being able to switch activities instantly, which saves considerable time over collecting one worksheet and handing out another.

Cuban (1993) argued that projectors became ubiquitous because they allowed teachers to prepare transparencies or slides in advance and to write legibly while facing the class. Those were tiny advantages, but overhead projectors did not constrain the instructors, so they became ubiquitous. We believe that general-purpose orchestration systems will become as ubiquitous as the overhead projectors, because they have exactly the same design goals: to make the teachers' job a tiny bit easier.

The FACT Analysis System

The other half of FACT is called the FACT Analysis System. It has two kinds of alerts: *process* alerts and *product* alerts. Process alerts, which are only available for group activities, analyze the temporal pattern of students' edits and report poor processes such as:

- Edits are made so quickly that thorough discussion of the edits is unlikely.
- Two people are working simultaneously on different parts of the poster, so they are probably not discussing their work with each other.

- One person is doing all the editing.
- Some members of the group are doodling, maliciously erasing other's work, or engaging in other apparently off-task edits.

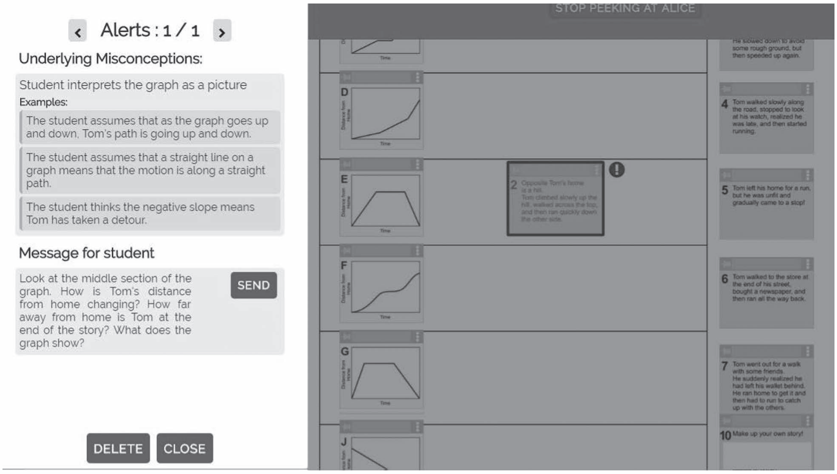
On the other hand, *product* alerts spot and report features in the students' solutions that represent possible misconceptions. For instance, there is a product alert for the misconception mentioned earlier, and it would be triggered when Tom's journey is paired with a side view of the hill. Different posters have different product alerts, whereas the process alerts are more general and apply to any poster edited by groups.

The alerts are part of a regulative loop (Soller et al., 2005; VanLehn, 2016). That is, FACT analyzes the students' work in real time, detects when there is an opportunity for improvement and alerts either the teacher or the students. Its goals are to encourage collaboration and to encourage self-correction. Let us first describe how it interacts directly with students, without involving the teacher.

In order to encourage collaboration, FACT uses its process alerts. Process alerts send messages directly to students without asking the teacher first. FACT's messages are brief and use the same wording as the posters used to launch the activity. For instance, if the poster says, "Take turns. Discuss each card placement when you make it," then one message might be "Are you taking turns?" and another message might be, "Are you discussing each card placement when you make it?" To avoid irritating students when they are just getting into an activity, process messages start appearing only after the students have completed 25% of the current activity. To avoid nagging, FACT will not give more than one process message per minute.

In order to encourage self-correction, FACT uses its product alerts. Although a standard ITS would use product alerts to point out student errors immediately and start down a hint sequence, such direct assistance thwarts self-correction. Thus, as long as students are still working on the current activity, they get no product alerts. However, when they press the Done button, FACT analyzes their poster and determines if it has any issues that provide opportunities for self-correction. If there are none, then it suggests an enrichment activity, such as writing brief justifications for each of the card placements. If there are issues, then it picks the highest priority one, draws a box around the part of the poster that raised the issue, and sends a message that is intended to break the students' mindset and push their thinking a little further. Figure 9.4 provides an example.

So far, we have described only how the FACT Analysis System interacts directly with students. However, it also interacts with teachers. Whenever the teacher peeks at a poster by tapping its tile on the dashboard, if there are alerts pending for that poster, the highest priority alert appears as a red box, a sidebar that explains the alert and a message with some questions for the students (see Figure 9.4). If the teacher agrees with FACT's analysis and decides to act on it, the teacher can walk over to that student or group armed with good initial questions for them. That is, FACT analyzes the students' performance and suggests an initial



Underlying Misconceptions:

Student interprets the graph as a picture

Examples:

- The student assumes that as the graph goes up and down, Tom's path is going up and down.
- The student assumes that a straight line on a graph means that the motion is along a straight path.
- The student thinks the negative slope means Tom has taken a detour.

Message for student

Look at the middle section of the graph. How is Tom's distance from home changing? How far away from home is Tom at the end of the story? What does the graph show?

2 Opposite Tom's home is a hill. Tom climbed slowly up the hill, walked across the top, and then ran quickly down the other side.

He slowed down to avoid some rough ground, but then speeded up again.

4 Tom walked slowly along the road, stopped to look at his watch, realized he was late, and then started running.

5 Tom left his home for a run, but he was unfit and gradually came to a stop!

6 Tom walked to the store at the end of his street, bought a newspaper, and then ran all the way back.

7 Tom went out for a walk with some friends. He suddenly realized he had left his wallet behind. He ran home to get it and then had to run to catch up with the others.

10 Make up your own story!

FIGURE 9.4 The Teacher Is Peeking at a Student Poster and FACT Is Showing a Product Alert

message to the students; the teacher then takes it from there, working with the students to encourage deeper thinking.

On the other hand, if the teacher thinks that this group is predisposed to take the questions seriously and that an actual visit isn't necessary, then the teacher can tap "Send." This causes the message and its red box to appear on the students' screens. Even if the teacher chooses to send the message to the students instead of visiting the students, the teacher still *owns* the message. We often see teachers call across the room, "Did you get my message?" Thus, even messages sent by the teacher count in the students' mind as coming from the teacher.

When looking at an alert, if the teacher disagrees with it, they can tap Delete. If the teacher agrees with the alert, but doesn't want to deal with it now, they teacher can tap Close. If the students fix the problem themselves, the alert disappears.

When teachers are supervising individual or small group work, they may appreciate advice on which individual or group to visit next. Since there are usually many alerts pending, FACT selects the N highest priority alerts and colors tiles on the teacher's dashboard red. The teacher can set N; its default is N = 1. For example, in Figure 9.4, group 1 has an alert.

At the bottom of the dashboard, FACT displays the most common current alerts along with the proportion of the posters in the class to which the alert applies. When most of the class is stuck on the same point, then the teacher may want to Pause the class and talk about that issue.

FACT's analysis system was field tested for the first time in early 2016. Although it interacts with teachers as described previously, it does not currently interact directly with students. Instead, all messages are sent by the teacher. The policies described earlier may change as field testing continues.

Technology

Although the FACT system grew out of tutoring system technology, it differs from most tutoring systems in several ways. Most notably, it is an orchestration system rather than a system that interacts with a student solving a problem individually. However, the really important difference is that FACT does not give corrective feedback. Its messages are intended to deepen students' reasoning by encouraging a *process*—collaborative, self-correction—rather than a mathematically correct *solution*. This has shaped the design of the whole Analysis system.

Whereas a typical tutoring system can generate or recognize all the correct solutions to a problem, the Analysis system is only concerned about raising alerts. Although process alerts use a temporal pattern matching approach that applies to all posters, different product alerts must be defined for different posters. A product alert has a detector, a description of the issue for the teacher to read, and several messages to send to students (see Figure 9.4). FACT lets developers replay a student's log file and mark episodes where an alert fired inappropriately or a new alert needs to fire. This facilitates creating and adjusting product detectors.

While it is simple to detect misplaced cards, more complex detectors are required for activities that involve sketching, handwriting, and typing in unconstrained natural language. Although we have not yet implemented sketch recognition, we use a commercial handwriting recognition system and our own segmentation algorithms to convert digital ink to text. We then use a combination of keyword spotting and deep learning for the natural language understanding. When a detector expects a mathematical expression, we use the Sympy library to recognize all mathematically equivalent expressions.

Evaluations

Our evaluations are based on classroom videos recorded from several cameras. One camera follows the teacher while a lapel radio microphone records the teachers' voice on the camera's soundtrack. For recording a pair of students, we use a camera mounted on a tall tripod looking down on the tablets of a pair of students and recording the pair's voices with a boundary microphone placed on their table. This design does a reasonable job of rejecting classroom noise. We typically record two pairs per class.

The videos are analyzed by human coders using Chi's code-and-count methods (Chi, 1997). Our summative evaluation, which has not yet taken place, will use two main dependent measures. One coding is based on Chi's ICAP codes (Chi & Wylie, 2014), extended to include off-task behavior. The other coding involves counting episodes of self-corrections, in which students edit their poster to remove an incorrect piece of work and replace it with a correct one. We will separately count spontaneous self-corrections and prompted self-corrections, where the latter occur shortly after the teacher or FACT have asked questions such as the ones shown in Figure 9.4.

Our evaluations so far have been formative and critically important for improving FACT. For example, we found that students would try to move a card while another group member was writing on it, so we added card locking to prevent this. This mystified students: why would cards sometimes move and sometimes refuse to move? We fixed this by adding a translucent padlock image atop cards that were locked. Teachers wanted to tell which students made which inscriptions, so we now assign each student in a group a different color ink. Messages to students from product alerts were originally displayed as cards placed next to the region of the poster that they referenced. However, if none of the students were looking at that part of the poster, they would not notice that a message had arrived. We added features that allow teachers to deal with students who arrive late or were absent during the first day of a lesson. It took roughly 40 classroom trials to converge on a robust, simple set of features.

Since then, we have been measuring collaboration and self-correction via the video analyses described earlier. We found that teachers simply could not visit students quickly enough to remedy all their process and product issues. This motivated our recent decision to have FACT start sending messages directly to students.

Although the results of the summative evaluation are forthcoming, we have already experienced some of the benefits of the FACT project. The development of the FACT Analysis System is forcing us to explicate what we expect of the students and the teachers. This is a necessary component of the development of any regulative loop system. As many ITS developers have noted, sometime making tacit pedagogical knowledge explicit is the biggest benefit of all. This chapter is our first attempt to express that formerly tacit knowledge.

Acknowledgements

This work was supported by OPP1061281 from the Bill and Melinda Gates Foundation. We gratefully acknowledge the contributions of the many graduate students, teachers, and staff whose talents were essential to the success of the FACT project.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, *5*(1), 7–75.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, *6*(3), 271–315.

Chi, M. T. H., & Wylie, R. (2014). ICAP: A hypothesis of differentiated learning effectiveness for four modes of engagement activities. *Educational Psychologist*, *49*(4), 219–243. doi:http://dx.doi.org/10.1080/00461520.2014.965823.

Cuban, L. (1993). Computers meet classroom: Classroom wins. *Teachers College Record*, *95*(2), 185–210.

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, *69*, 485–492.

FaSMEd. (2017). Improving Progress for Lower Achievers through Formative Assessment in Science and Mathematics Education (FaSMEd). Retrieved from https://research.ncl.ac.uk/fasmed/.

Herman, J., Epstein, S., Leon, S., La Torre Matrundola, D., Reber, S., & Choi, K. (2014). *Implementation and effects of LDC and MDC in Kentucky districts (CRESST Policy Brief No. 13)*. Los Angeles, CA.

Herman, J., La Torre Matrundola, D., Epstein, S., Leon, S., Dai, Y., Reber, S., & Choi, K. (2015). *The implementation and effects of the Mathematics Design Collaborative (MD): Early findings from Kentucky ninth-grade algebra 1 courses (CRESST Report 845)*. Retrieved from Los Angeles:

Hume, G., Michael, J., Rovick, A., & Evens, M. (1996). Hinting as a tactic in one-on-one tutoring. *Journal of the Learning Sciences*, *5*(1), 23–49.

Inverness Research. (2016). Mathematics Assessment Program (MAP): Project Portfolio. Retrieved from http://inverness-research.org/mars_map/1_welcome.html.

Lin, C.-J., Chen, W., Yang, S.-J., Xiet, W., & Lin, C. C. (2014). Exploring students' learning effectiveness and attitude in group scribbles-supported collaborative reading activities: A study in the primary classroom. *Journal of Computer Assisted Learning*, *30*, 68–81.

Looi, C.-K., Lin, C.-P., & Liu, K.-P. (2008). Group scribbles to support knowledge building in a jigsaw method. *IEEE Transactions on Learning Technologies*, *1*(3), 157–164.

MDC. (2016). The Mathematics design collaborative. Retrieved from http://k12education.gatesfoundation.org/student-success/high-standards/literacy-tools/mathematics-design-collaborative/.

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

Research_for_action. (2015). *MDC's Influence on Teaching and Learning,* Philadelphia, PA. Retrieved from www.researchforaction.org/publications/mdcs-influence-on-teaching-and-learning/.

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? *Educational Researcher*, *48*(8), 404–412.

Schoenfeld, A. H. (2017). Teaching for Robust Understanding (TRU-Math). Retrieved from http://map.mathshell.org/trumath.php.

Soller, A., Martinez, A., Jermann, P., & Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence and Education*, *15*, 261–290.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence and Education*, *16*, 227–265.

VanLehn, K. (2016). Regulative loops, step loops and task loops. *International Journal of Artificial Intelligence in Education*, *26*(1), 107–112. doi:10.1007/s40593–015–0056-x.