Modeling Collaboration Patterns on an Interactive Tabletop in a Classroom Setting

Abigail C. Evans, Jacob O. Wobbrock and Katie Davis
Information School | DUB Group
University of Washington
Seattle, WA 98195 USA
{abievans, wobbrock, kdavis78}@uw.edu

ABSTRACT
Interaction logs generated by educational software can provide valuable insights into the collaborative learning process and identify opportunities for technology to provide adaptive assistance. Modeling collaborative learning processes at tabletop computers is challenging, as the computer is only able to log a portion of the collaboration, namely the touch events on the table. Our previous lab study with adults showed that patterns in a group’s touch interactions with a tabletop computer can reveal the quality of aspects of their collaborative process. We extend this understanding of the relationship between touch interactions and the collaborative process to adolescent learners in a field setting and demonstrate that the touch patterns reflect the quality of collaboration more broadly than previously thought, with accuracies up to 84.2%. We also present an approach to using the touch patterns to model the quality of collaboration in real-time.

Author Keywords
Collaborative learning; tabletop; modeling; “in the wild.”

ACM Classification Keywords
H.5.3. Group and organization interfaces: Collaborative computing.

INTRODUCTION
Tabletop computers have affordances for collaborative learning because of the large, shared interface that multiple people can see and interact with at once (Figure 1) [9,17,18,32,33]. However, the process of collaborative learning is complex and subject to a variety of factors external to the technology, such as students’ collaboration skills, group dynamics, and the nature of the learning activity [34]. When the collaboration process derails, students’ learning can be negatively impacted. So although tabletops may afford collaboration, small group work at a tabletop computer is just as susceptible to breakdowns in collaboration as group work using other kinds of tools. More can be done to capitalize on the affordances of collocated tabletop interaction to bring about resilient learning outcomes.

One avenue for tabletop technology to better support collaborative learning is to detect and respond to breakdowns in collaboration in real-time. With this goal in view, we model the process of small-group collaborative learning at tabletop computers. By modeling collaboration, we plan to enable the design of software that supports the collaborative process through interventions that track, measure, and reinforce beneficial group dynamics.

A significant challenge in modeling collaborative learning at an interactive tabletop is that the collaboration process is made up of verbal and gestural interactions between learners and their interactions with the tabletop, but only the latter—direct interaction with the tabletop—is visible to the computer. In our previous work [14], we found several patterns in those direct interactions that indicated the quality of content monitoring [34], a key collaborative learning process. In this paper, we present a follow-up study conducted in a classroom field setting with high school students to find out if these touch interaction patterns continue to reflect collaborative learning processes beyond the original lab setting.

Today’s tabletop computers are an emerging technology, rarely seen in classrooms. It is predicted, however, that
tabletops will see widespread adoption in the near future [28]. If tabletops are to be adopted in educational settings, the peculiarities of these environments must be taken into account [2,12,20]. This is a key motivation for our work moving from the lab to the field.

In this work, we address the following research question: can the quality of groups’ collaborative learning processes be detected using touch data alone? The contributions of this work are: (1) a refined set of touch patterns reflecting the quality of high school students’ collaborative learning processes; and (2) an approach to using these touch patterns to model collaborative learning processes in real-time that enables breakdowns to be detected using only touch data.

**COLLABORATIVE LEARNING & SOCIAL REGULATION**

Education research [5,10] shows that pedagogical approaches that give students the opportunity to discuss and collaborate with others can be particularly helpful in developing their higher-order thinking skills such as analyzing, synthesizing, and evaluating. Through working with their peers, learners are exposed to others’ problem solving approaches and are given the opportunity to better understand their own thinking by explaining it to others.

In this paper, we use a definition of collaborative learning drawn from the Learning Sciences. Collaborative learning occurs when a group works together to construct a shared understanding of a problem [11]. This definition separates collaborative learning from cooperative learning, in which learners work independently before coming together, e.g., by dividing a task among group members. Although collaborative and cooperative learning both have strengths, the focus here is activities where collaboration is desirable.

An implication of this definition of collaborative learning is that the unit of analysis is the group rather than the individual. To model the collaboration process, it is important to look at interactions among group members rather than contributions of individual members in isolation.

**Social Regulation in Collaborative Learning**

Social regulation refers to the processes that groups use to manage their learning [19,34]. When social regulation is successful, the group collectively establishes goals, monitors and develops shared understanding of the task, and monitors progress toward shared goals. Research shows that students often fail to capitalize on the benefits of collaborative learning because they do not engage in social regulation [19,34].

Social regulation involves three main processes: planning of task work and the group’s approach to task content; monitoring of the task plan, of group understanding of content, and progress; and behavioral engagement, the process of getting an off-task group member to engage with the task [34]. In our previous work [14], we found touch interaction patterns were associated with the quality of groups’ content monitoring, a sub-process of monitoring that groups use to check their understanding of the task at hand by asking questions and engaging in dialogue to confirm or challenge each other’s contributions. Of all the social regulation processes, content monitoring has been shown to be critical to groups’ development of shared understanding. According to Rogat and Linnenbrink-Garcia [34], content monitoring is high-quality when groups focus on building conceptual understanding of a task, rather than finding the right answer.

While no single concept can encompass the full complexity of collaborative learning, social regulation’s influence on collaborative cognition and group outcomes [19,34] make it an appropriate tool for determining quality of collaboration in the context of this research.

**RELATED WORK**

Considerable work has been carried out on understanding how groups collaborate at tabletop computers, e.g. [3,29], with many researchers focusing on children, e.g. [15–17,31,32]. Previous research has also explored how interaction design can help users manage territoriality [35,37] and resolve conflicts that occur due to the shared interface [27]. Additionally, visualization tools exist to help designers to understand and evaluate how users collaborate with their software [1,38].

Most of the prior work in modeling collaboration at tabletop computers has taken place in lab settings, with groups of adults working on simulated tasks. A smaller body of work has begun to model tabletop collaborative learning specifically. Even less published research has taken place in authentic classrooms. The most extensive work in this area has come from Martinez-Maldonado et al. [25]. He and his colleagues used data mining techniques to extract application-specific sequential patterns of interaction associated with high- and low-achieving groups of elementary students. They also developed visualizations to show how much individual group members contributed in verbal utterances and touch interactions, which clearly showed the balance of participation among members but could not reveal quality [22,23]. Later, they combined these visualizations with a model task solution created by a content expert. Teachers could compare groups’ progress on a task to the expert model, enabling real-time identification of groups that needed help [21,24].

In Martinez-Maldonado et al.‘s approach, the process modeling and the software that students used were tightly integrated—the patterns detected were specific to the software. They also relied on an expert model, or clear “right answer,” which is not appropriate for open-ended or creative activities. In contrast, our goal is to develop a model of groups’ social regulation processes that can be used in a variety of contexts, including open-ended collaborative tasks. This requires an approach that is independent of the specific task or learning objectives.

Social regulation occurs in the interactions between group members as well as their interactions with the learning
environment. A tabletop computer, however, can only detect direct interactions with the interface without the use of additional sensors. To model the collaborative learning process, we need to understand how physical interactions with the computer relate to the collaboration as a whole. A small body of work has been carried out in this area.

Fleck et al. [16] explored the relationship between groups’ verbal interactions and physical actions on the tabletop. Verbal elements of successful collaboration, such as making and accepting suggestions or negotiating, were often complemented by particular actions in the software. Although Fleck et al. do not mention it directly, their markers of successful collaboration align with processes of social regulation, including content monitoring.

Pontual-Falcão and Price [30] investigated instances of “interference” that interrupted or changed the flow of students collaborating on a tabletop activity. Although it is often assumed that such interference hinders the collaborative process, Pontual-Falcão and Price found that interference could be productive or counter-productive, depending on how the group responded to it. A key finding was that interference primarily took place through physical rather than verbal interactions.

Both of these studies demonstrate that physical interactions are an integral part of the tabletop collaboration process, often directly coupled to or influencing verbal interactions. Building on this close relationship between physical and verbal collaboration processes, we identified touch patterns associated with quality of social regulation in a lab study of adults working on an open-ended poetry analysis task [14]. Although the quality of most social regulation processes was similar across groups, participants’ patterns of interaction clearly differed during low- and high-quality episodes of the regulatory sub-process content monitoring. This finding suggested that collaboration breakdowns could be detected independently of specific software or learning activities, and provided the rationale for the current work, whose main contribution beyond our previous work is an approach to using touch patterns to detect collaboration problems in real time.

**METHOD**

In this study, we sought to build upon our previous work [14] by evaluating the interaction patterns associated with quality of content monitoring in a field study with high-school students in a different learning domain. Whereas the adults in our lab study [14] remained on-task and engaged at all times, studies of students using interactive tabletops in the field have shown them to be not quite so consistently well-behaved or motivated to stay on task [13,20]. Given the stark differences between the behavior of the adults in the lab setting and that of students in the wild, it is important to test this approach to detecting collaborative processes in an authentic setting.

By studying students’ social regulation at a tabletop computer in a classroom, we determined: (1) whether the touch patterns associated with quality of content monitoring would transfer to this new, field-based context; (2) what combinations of patterns revealed about the collaboration process; and (3) whether the touch patterns could be used to detect automatically and in real-time the quality of collaboration.

**Participants**

Sixteen high school students (10 female, 6 male) participated in this study. The students were enrolled in a six-week user-centered design seminar offered as part of a summer program. The course was structured so that students learned about user-centered design while working on a group project to design and build a website around a topic of their choosing. The students were divided into four project groups with four students per group.

**Apparatus**

The participants used a Microsoft PixelSense, a vision-based tabletop computer with a 40” multi-touch screen. The computer was placed to one side of the classroom with a wide-angle video camera mounted on a tripod on a desk next to the computer, leaving enough space for students to move freely around the tabletop. The camera was angled toward the screen so that it could capture every touch to the computer and interactions among the group members.

**Procedure**

We custom-built four distinct applications designed to be integrated into the program curricula and used alongside other classroom activities and resources. Each application was used in a different class session scheduled to fit the timing of particular curriculum topics. The students carried out each of the tabletop activities in their project groups.

![Figure 2: The applications created for this work. Top left: design critique. Top right: question brainstorming. Bottom left: search. Bottom right: heuristics.](image)

Each application addressed the specific learning objectives for a class session. Three of the applications used the native Microsoft PixelSense SDK and one was web browser-based. The first application (Fig. 2, top left) supported design critiques and was used in the early stage of the group project to finalize their website layouts. A second
application (Fig. 2, top right) helped groups prepare a usability test plan. It included a question sorting component and a question brainstorming component. Most of the students had no web programming experience so a third application (Fig. 2, bottom left) developed groups’ search skills to help them find and evaluate online technical resources. The final application (Fig. 2, bottom right) was a Chrome browser plugin that enabled students to find and annotate, via drag-and-drop, real-world examples of designs that support or violate Nielsen’s usability heuristics [29].

Each group was video recorded during each activity and the computer logged every touch. In all, 4253 touches were logged over 13 group sessions.

In order to detect the touch patterns described in our earlier work [14], it was not necessary to know exactly which on-screen objects were touched. For some of the touch patterns, however, it was important to know which objects were related in terms of their function in the learning software. For example, during high-quality social regulation, a group maintains shared focus on whatever aspect of the task it is working on. Knowing whether the objects being touched were related or unrelated makes it possible to determine if the group was maintaining shared focus. When a group was interacting with related objects, it was likely they were focused on a particular feature or aspect of the task. Conversely, when a group was interacting with unrelated objects, it was likely they were shifting focus between unrelated aspects of the task.

Object relationships were described in simple XML files packaged with three of the applications. The XML files contained a top-level node for each on-screen object. Each top-level node then contained a list of child nodes identifying any related on-screen objects. All relationships were symmetrical and unordered. Relationships were determined when the application was built, using the same criteria as in our previous study [14]: each object’s function in the context of the activity.

Consider the following example from the search application (on-screen objects are italicized): This application includes a Vocabulary object, containing optional prompt words that groups may use to formulate a search query. The students type their query into a Search Query object using the on-screen Keyboard. These three objects are therefore functionally related; while interacting with these elements the focus is creating a search query and all three objects could be expected to be involved. Once a group completes a Search Query they progress to Search Results, a separate object and the final stage of the search. As the Search Results object becomes available as a direct consequence of actions in the Search Query, these two objects are functionally related. At this stage of the search process, there is no reason for the group to interact with both the Search Results and Vocabulary, or both the Search Results and Keyboard, unless they are shifting focus from evaluating results to formulating a new query. Therefore, the Search Results and Vocabulary are unrelated, as are the Search Results and Keyboard.

Because relationships were defined in XML files, not hardcoded in the applications, they could be adjusted at any time if the designer’s assumptions were later found to be false. No adjustment was necessary in this study, however.

The browser-based application did not include a relationship file because browsing was unrestricted and therefore on-screen relationships could not be determined. Without a relationship file, one of the three touch patterns, which relies on the on-screen relationships, could not be applied to the touch data for the browser-based application.

DATA ANALYSIS
Video Analysis – Coding for Social Regulation
The videos were thematically coded for social regulation independently from the touch data, a process that took a number of weeks (see Social Regulation Processes, Table 1). The social regulation codes were drawn directly from Rogat and Linnenbrink-Garcia [34] and we relied heavily on their detailed descriptions and examples of each code. All stages of video coding were carried out multiple times by one researcher and validated with peer debriefing [7] until no further revisions were deemed necessary.

First, the videos were viewed in their entirety to gain a general sense of each group’s interactions. The videos were then transcribed and each interaction between group members was coded for social regulation, where it was occurring, using Rogat and Linnenbrink-Garcia’s framework [34]. The planning and monitoring processes both have sub-processes (listed in Table 1), referring to which aspect of the collaboration was being planned or monitored. Interactions that represented planning or monitoring were coded at the more specific sub-process level. Occurrences of social regulation were also coded for quality, again drawing on Rogat and Linnenbrink-Garcia’s established framework. Table 1 does not provide the lengthy and detailed criteria used to determine quality as they were pulled directly from Rogat and Linnenbrink-Garcia’s paper without modification (see p388 in [34]).

The coded transcripts were reviewed alongside the videos. Additional narrative observations were added for events or interactions that did not fall under social regulation but described the nature of a group’s collaboration on the assigned task, such as when students were off-task. These observations were used to extend the list of codes (Additional Codes, Table 1) through collaborative coding [36] to a total of 10 codes: 6 social regulation codes consisting of the planning and monitoring sub-processes plus behavioral engagement, and 4 additional non-social regulation codes. Referring back to Rogat and Linnenbrink-Garcia’s descriptions [34] and the video, the codes applied to the transcripts were adjusted as the subtleties of each code was teased out.
All dimensions of social regulation were included in the video coding for this study even though our previous work [14] only found touch patterns associated with content monitoring. In the previous study, all groups used consistently high-quality planning processes and remained collaborative and on task at all times, meaning that there was no need for behavioral engagement, the process of re-engaging off-task group members [34]. This meant that the touch pattern analysis was limited to monitoring and its sub-processes, the only aspect of social regulation where there were differences between groups. In contrast to the lab study, we expected high school students in a naturalistic classroom setting to go off-task and show greater variation in quality of social regulation than the adults. This expectation led us to code for all social regulation processes, not just content monitoring.

Table 1 also shows whether each code was considered to have a positive or negative impact on collaboration. Non-collaborative interactions was a code used by Rogat & Linnenbrink-Garcia in their work on social regulation [34], referring to episodes in which group members work independently. While there are many cases where non-collaborative interactions would be acceptable, none of the activities used in this study would have benefited from a divide-and-conquer approach and groups were explicitly told to complete the task collaboratively. Therefore, in this study, non-collaborative interactions were deemed negative.

Off-task interactions refers to episodes in which the group was engaged with an activity other than the assigned task, such as gossiping about classmates or discussing TV shows.

Task work refers to episodes in which groups were on task but not engaged in any discussion, such as when one group member was typing while other group members looked on.

Software conflicts refers to episodes in which multiple students carried out conflicting actions in the software, bringing the activity to a halt until the issue was resolved. In these cases, the group’s attention was on the software, not the assigned task, which meant the episode could not easily be described with another code. The software conflicts code does not include conflicting actions that resemble Pontual-Falcão and Price’s [30] notion of interference – conflicts that moved task work forward, or at least allowed it to continue. Although our software conflicts code could be considered a type of off-task interaction, we decided that it described interactions that were qualitatively different from general off-task interactions and therefore merited a dedicated code. Software conflicts typically forced off-task interaction as the students had to stop what they were doing to address the fault. This is in contrast to general off-task interaction, in which groups voluntarily or spontaneously stopped working on the assigned task.

The coding process allowed the videos to be broken down into episodes. In Chi’s terms [4], an episode is “an event, or a specific activity” (p. 284), for example when a group was engaged in a particular social regulation process or another activity, such as working in silence, or off-task conversation. Episodes could have multiple codes; for example, when some students in the group were off-task while the others were engaged in task planning. Episodes varied greatly in length – from a few seconds to several minutes of interactions.

<table>
<thead>
<tr>
<th>Social regulation processes</th>
<th>Impact on collaboration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning:</td>
<td>Processes coded as high or low quality</td>
</tr>
<tr>
<td>• Task</td>
<td>High-quality processes considered to have positive impact</td>
</tr>
<tr>
<td>• Content</td>
<td>Low-quality processes considered to have negative impact</td>
</tr>
<tr>
<td>Monitoring:</td>
<td></td>
</tr>
<tr>
<td>• Content</td>
<td></td>
</tr>
<tr>
<td>• Plan</td>
<td></td>
</tr>
<tr>
<td>• Progress</td>
<td></td>
</tr>
<tr>
<td>Behavioral engagement</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional codes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Non-collaborative interactions</td>
<td>Negative</td>
</tr>
<tr>
<td>• Off-task interactions</td>
<td>Negative</td>
</tr>
<tr>
<td>• Task work</td>
<td>Positive</td>
</tr>
<tr>
<td>• Software conflicts</td>
<td>Negative</td>
</tr>
</tbody>
</table>

Table 1: Codes used in the video analysis. Social regulation codes adapted from [34].

**Touch Data Pre-processing**

Touch data analysis began with pre-processing. It is important to note that the ability to distinguish between individuals is essential to using the modeling approach described in this work. The Microsoft PixelSense used in this study does not natively distinguish between users. Therefore, the first step was to use the video to manually label each touch with its author – the group member who performed the touch. Several alternative methods for distinguishing users exist; for example, through the use of a depth cameras [6,23] or wearable sensors [26]. Additionally, the DiamondTouch [8] table natively supports user tracking. Where these options can be used, manually labeling touches would not be necessary.

Next, sequences of touches representing complete actions carried out by an individual group member were automatically extracted from the log files using the rules developed in our earlier work [14] and shown in Table 2.

<table>
<thead>
<tr>
<th>Elements touched are:</th>
<th>Part of the same sequence</th>
<th>Part of separate sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>related</td>
<td>≤ 50 seconds</td>
<td>&gt; 50 seconds</td>
</tr>
<tr>
<td>unrelated</td>
<td>≤ 15 seconds</td>
<td>&gt; 15 seconds</td>
</tr>
</tbody>
</table>

Table 2: Rules for grouping touches with the same author into sequences by time between touches, adapted from [14].

The XML relationship file for the application was queried to determine if two elements touched sequentially were related or unrelated, then the rules given in Table 2 were applied to determine if the second touch was part of the same sequence as the previous touch or the start of a new sequence. Sequences overlapped in time when multiple students were touching the tabletop simultaneously.
**Touch Data Analysis**

We used the video timecodes marking the beginning and end of episodes to find the corresponding touch data in the log files. Episode boundaries were based on group activity and interactions in the video so episodes did not always neatly align with touch sequences. Longer episodes typically contained multiple touch sequences and episode boundaries sometimes fell in the middle of a touch sequence. In these cases, touch sequences were aligned with the episode containing the bulk of the sequence as determined by the timestamps associated with the touch data and the corresponding episode timecodes.

Next, we inspected the processed touch logs to determine if the touch patterns originally identified in [14] continued to serve as indicators of the quality of content monitoring in the context of this study. Table 3 describes the touch patterns as they were applied to the data from this study. The quality thresholds in Table 3 were calculated from the results of the previous study. Episodes of touch data that fell between the thresholds were labeled “medium-quality.”

<table>
<thead>
<tr>
<th>Pattern</th>
<th>High quality</th>
<th>Low quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Unrelated Touches – Frequency of touches to unrelated objects</td>
<td>≤ 5% of touches in sequence</td>
<td>&gt; 10% of touches in sequence</td>
</tr>
<tr>
<td>2) Touch Time – Amount of time spent touching the screen per person</td>
<td>≤ 15% of time</td>
<td>&gt; 30% of time</td>
</tr>
<tr>
<td>3) Overlapping Sequences – multiple users interacting with the screen at the same time</td>
<td>≤10% of touches are “overlapping”</td>
<td>&gt;20% of touches are “overlapping”</td>
</tr>
</tbody>
</table>

Table 3: Touch patterns associated with quality of content monitoring. Quality thresholds are derived from [6].

Based on the results of this analysis, we then looked at how combinations of patterns reflected collaboration quality, e.g., when an episode was labeled high quality according to one pattern and low quality according to another.

In the previous study [14], there were limited episodes of social regulation processes other than content monitoring so it had not been possible to determine whether the touch patterns might apply to other processes. Although there are numerous features that make each process distinct, all high-quality processes share these key characteristics: the group maintains shared focus on the learning task and the group dynamic is cohesive and inclusive of all members’ contributions. If one or both of these characteristics is lacking, the social regulation process is deemed low quality.

As the previous study had only been able to speak to quality of content monitoring, we decided to check the patterns against all video codes in order to find out if the patterns were detecting quality of content monitoring specifically or if they may be reflecting the common characteristics of social regulation more broadly. This decision was also made with a view to addressing an obstacle specific to using the touch patterns to model collaboration in real time.

With video analysis, it is possible to identify episodes of content monitoring then use the patterns to determine each episode’s quality. In real time, however, all touch data would be checked against the patterns but, without the video analysis to identify which touch data represented episodes of content monitoring, there would be no way to tell if pattern quality reflected quality of content monitoring or something else entirely. Understanding how the patterns reflect common features of quality of collaboration beyond content monitoring could address this issue.

The final step of the touch data analysis was to test an approach to using touch patterns to detect collaboration problems in real-time. In order to use the patterns in real-time, the touch data would have to be checked against the patterns while the activity was in progress. While analyzing videos allows clear episode boundaries to be identified, this would not be possible in real-time or with touch data alone. Therefore, we checked for the patterns given in Table 3 at regular intervals.

Intervals needed to be long enough to allow meaningful face-to-face interactions to take place, but short enough to allow early detection of behavior indicative of collaboration problems. In our analysis, an interval of 2 minutes was used as most (93%) of the episodes from the coded video data were under this duration. The first interval began when a log file was created at the start of an activity. Most 2-minute intervals contained at least one complete episode. To reduce the potential for incorrect classifications caused by the artificial interval length, 2-minute intervals were started every minute so that they overlapped. Each interval was automatically labeled as high, low, or medium (falling between the thresholds) quality according to each pattern in Table 3. Finally, the automatically labeled intervals were compared to the coded episodes identified during video analysis using the timecodes marking the bounds of each episode to match them to the appropriate interval(s). In cases where an episode spanned the boundary of two intervals, the episode was aligned to the interval containing the majority of the episode (in seconds).

**RESULTS**

Our analysis showed that the high school students in a classroom setting used a greater range of social regulation processes with more variation in quality, enabling us to investigate the relationship between the touch patterns and quality of collaboration more broadly than in the previous study. We found that two of the touch patterns in combination—Unrelated Touches and a modified version of Overlapping Sequences—were clearly associated with quality of collaboration in up to 84.2% of cases. Checking these patterns against intervals of touch data, rather than video episodes, proved to be a viable approach to detecting quality of collaboration in real-time.

**Social Regulation & Quality of Collaboration**

Unlike the adults in [14], whose quality of social regulation remained consistent for the duration of the activity, the high
school students regularly engaged in both high- and low-quality social regulation during an activity. Additionally, all groups of students employed behavioral engagement, a process not seen with the adult groups. Each group’s use of social regulation processes in all activities combined is shown in Table 4. The results are given as percentages to facilitate inter-group comparison as the number of episodes of social regulation in each group differed.

<table>
<thead>
<tr>
<th>Social regulation process</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning - task</td>
<td>HQ 0</td>
<td>LQ 36.8</td>
<td>HQ 0</td>
<td>LQ 6.7</td>
</tr>
<tr>
<td>Planning - content</td>
<td>0 5.3</td>
<td>14.7 5.9</td>
<td>7.7 15.4</td>
<td>10.0 10.0</td>
</tr>
<tr>
<td>Monitoring - content</td>
<td>31.6 5.3</td>
<td>32.4 17.6</td>
<td>23.1 30.8</td>
<td>16.7 33.3</td>
</tr>
<tr>
<td>Monitoring - plan</td>
<td>0 10.5 2.9 5.9</td>
<td>0 15.4 10.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral engagement</td>
<td>0 5.3 8.8 2.9</td>
<td>0 15.4 10.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>31.6 68.4 61.8 38.2</td>
<td>38.5 61.5 36.7 63.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Quality of social regulation by group, shown as the percentage of each group’s social regulation episodes in all activities combined.

Table 4 shows that each group employed social regulation processes differently, with some groups relying on particular processes more than others. For example, group 1 engaged in considerably more task planning than any other group while group 3 failed to engage in task planning at all. Content monitoring was the most heavily used process in all groups, representing 47.9% of all social regulation episodes. With the exception of group 2, overall social regulation was low quality more often than high quality.

<table>
<thead>
<tr>
<th>Code</th>
<th>Mean number of episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-collaborative interactions</td>
<td>1.7 1.3 1.3 0.3 1.2</td>
</tr>
<tr>
<td>Off-task interactions</td>
<td>2.0 2.0 3.0 0.7 1.9</td>
</tr>
<tr>
<td>Task work</td>
<td>0 2.0 0.7 1.3 1.1</td>
</tr>
<tr>
<td>Software conflicts</td>
<td>1.0 1.8 2.0 1.3 1.5</td>
</tr>
</tbody>
</table>

Table 5: Mean number of episodes of each additional code per activity by group.

The high school students also differed from the adults in [14] in terms of the behaviors described by the additional codes given in Table 1: non-collaborative interactions, off-task interactions, task work, and software conflicts. Whereas the adults remained on task and collaborative, the students frequently engaged in off-task and non-collaborative interactions. Episodes of each of the additional codes occurred in all groups, with the exception of group 1, who did not have any episodes of task work. Table 5 shows the average number of episodes of each additional code per activity for each group.

**Touch Patterns and Quality of Content Monitoring**

Table 6 shows the number of episodes coded as high- or low-quality content monitoring in the video analysis and the percentage of automatically generated quality labels that matched the video code for those episodes.

Table 6: The number (#) of video episodes coded for each pattern by quality and the percentage of episodes where the automatically generated label matched the video code. Results are shown for the three patterns found in [14] and the revised pattern found in this study.

As previously described, quality of content monitoring varied between and within groups, across activities and within activities. Although some groups (1 and 2) tended to be more effective at content monitoring, all groups had episodes of high and low quality. This is markedly different from the original study [14], in which the majority of episodes of content monitoring were high quality, with all low-quality episodes occurring in only one group.

The video analysis confirmed that the Unrelated Touches pattern, the proportion of unrelated elements in touch sequences, was associated with quality of content monitoring in the classroom setting. As in the lab study [14], a high occurrence of touch sequences involving unrelated UI elements indicated low-quality content monitoring. Of the 26 episodes of content monitoring checked against this pattern, only two were mis-categorized. Both were low-quality episodes; one was categorized as medium-quality, and the other as high quality. Fewer episodes were checked against Unrelated Touches than Touch Time and Overlapping Sequences because it requires the relationship definitions, which were unavailable for one of the applications.

Review of the videos for high- and low-quality episodes of the Unrelated Touches pattern suggested that episodes were labeled high quality when all participants interacting with the screen were engaged in focused work, whether they were actively doing task work or not. During low-quality episodes, the participants interacting with the screen were shifting focus or playing with the interface, interacting with multiple unrelated elements without a particular purpose.
The video analysis showed that Touch Time, the time spent interacting with the screen per person, did not hold as an indicator of quality of content monitoring in this study. Almost all episodes were categorized as high quality by this pattern. Therefore, Touch Time was dropped from further analysis as not generalizable beyond the original context.

Overlapping Sequences was also a poor indicator of content monitoring quality, most frequently categorizing episodes as low quality. The video analysis, however, suggested an amendment. In several cases, the video revealed that during an interval categorized as low quality by Overlapping Sequences, students were, in fact, engaged in positive collaborative interactions, such as helping each other complete a task or building on each other’s contributions. In these cases, we noticed that the overlapping touch sequences were occurring across related elements.

Accordingly, we revised Overlapping Sequences to include a qualifier: a high number of overlapping sequences indicates low-quality content monitoring only when students interacting with the screen are working with unrelated objects. Conversely, a high number of overlapping sequences indicate high-quality content monitoring when the students are working with related objects. The results for the revised version of Overlapping Sequences, renamed Overlapping Unrelated Sequences, are shown in Table 6. Overlapping Unrelated Sequences replaced Overlapping Sequences in subsequent analysis. Review of the videos for high- and low-quality episodes of Overlapping Unrelated Sequences suggested that episodes were labeled high quality when the group members interacting with the screen were working collaboratively and low quality when they were working independently.

**Touch Pattern Combinations & Quality of Collaboration**

Our video analysis showed a much greater range of collaboration processes than reported in the lab study with adults [14]. Therefore, all episodes were included in our model in order to extend the analysis beyond content monitoring and investigate the relationship between the touch patterns and quality of collaboration more broadly.

For Unrelated Touches and Overlapping Unrelated Sequences, we identified the most common video codes as well as the distribution of positive and negative collaborative processes (defined in Table 1) for each possible quality combination of the two patterns (listed in column a, Table 7). The pattern quality was determined using the video episode timecodes for a one-to-one match between the video codes and the automatically generated quality labels. Both remaining touch patterns required knowledge of the relationships between on-screen elements so data from the browser-based activity was omitted. The results are shown in columns a – d in Table 7.

Column d in Table 7 shows how each combination of automatically generated quality labels aligned with the episode codes resulting from the video analysis. Most combinations of the two patterns aligned with episode codes indicating processes considered to have negative impact on collaboration e.g. low quality content monitoring or off-task behavior. There were only two combinations for which this was not the case: 1) the high-high combination, which was most often associated with episodes with

<table>
<thead>
<tr>
<th>(a) Pattern quality label</th>
<th>(b) Overlapping Sequences</th>
<th>(c) Most common episode code(s)</th>
<th>(d) Distribution of positive / negative codes (episodes)</th>
<th>(e) Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrelated Touches</td>
<td>Overlapping Sequences</td>
<td></td>
<td>Positive codes</td>
<td>Mixed codes</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>28</td>
<td>100% positive</td>
<td>8</td>
</tr>
<tr>
<td>High</td>
<td>Med.</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>17</td>
<td>100% negative</td>
<td>1</td>
</tr>
<tr>
<td>Med.</td>
<td>High</td>
<td>1</td>
<td>100% positive</td>
<td>1</td>
</tr>
<tr>
<td>Med.</td>
<td>Med.</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Med.</td>
<td>Low</td>
<td>7</td>
<td>100% negative</td>
<td>-</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>8</td>
<td>87.7% negative</td>
<td>-</td>
</tr>
<tr>
<td>Low</td>
<td>Med.</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>19</td>
<td>84.2% negative</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7: Collaboration codes associated with all possible quality combinations of Unrelated Touches and Overlapping Unrelated Sequences. The two original patterns that did not transfer from the original setting, Touch Time and Overlapping Sequences, are not included. Column a) combination of automatically generated quality labels; b) number of episodes labeled with given combination; c) most common video code for the given combination; d) distribution of positive/negative video codes for episodes labeled with the given combination; e) distribution of positive/negative episodes in intervals labeled with the given combination.
positive codes, particularly high-quality content monitoring; and 2) the combination of high for Unrelated Touches and low for Overlapping Unrelated Sequences, which was associated with a greater mix of positive and negative collaboration episodes than all other combinations.

The fact that all but one of the combinations (high-low) tended to be primarily associated with either positive or negative collaboration behaviors, not just content monitoring, suggests that these patterns in combination reflect the shared characteristics of high- and low-quality social regulation in sum rather than the specific features of content monitoring. For example, when an episode is labeled high quality according to both patterns, we can predict that positive collaboration behavior is occurring, as was the case in 75% of occurrences in this study. When an episode is labeled low quality according to both patterns, we can predict that negative collaboration behavior is occurring, which was true for 84.2% of occurrences.

Column b in Table 7 shows that the most frequently occurring combinations were 1) high quality for both Unrelated Touches and Overlapping Unrelated Sequences, which was associated with positive collaboration codes; 2) low quality for both patterns, which was associated with negative codes, and 3) high quality for Unrelated Touches combined with low quality for Overlapping Unrelated Sequences, which was associated with slightly more negative collaboration codes than positive. These results indicate that the high-high and low-low combinations are the clearest predictors of collaboration quality, with the high-low combination being more ambiguous. Other quality combinations were infrequent or not seen at all. Although several of the infrequent combinations were strongly associated with negative collaboration codes (87.7% - 100% of cases), we are cautious about their association with collaboration quality due to their rarity in this study.

**Testing the Interval Approach to Detecting Collaboration Quality**

Column e in Table 7 shows how each quality combination compared to the video collaboration codes when the patterns were checked against intervals of touch data rather than individual episodes. The values in column e are the number of intervals containing primarily positive collaboration codes, the number containing a relatively even mix of positive and negative codes, and the number containing primarily negative episodes. For the interval approach to be useful for detecting quality of collaboration in real-time, the intervals’ distribution of positive or negative collaboration processes should closely resemble the category of the episodes contained in the interval. For example, as the high-high combination was associated with mostly positive collaboration episodes, we would expect that intervals labeled as high quality by both patterns would contain mostly positive collaboration episodes.

When applied to intervals of touch data, the most commonly occurring quality combinations showed a similar distribution of positive and negative collaboration codes as when applied to individual episodes drawn from the video, e.g., pattern combinations primarily associated with episodes with negative collaboration codes continued to be primarily negative when applied to intervals of touch data. Of the three most common quality combinations, the low-low combination had the closest match between intervals and episodes, associated with negative collaboration codes in 85.7% of the intervals compared to 84.2% of episodes. Because the interval approach yielded collaboration quality labels close to those of the video episodes, these results suggest that applying the patterns to intervals of touch data would make it possible to model collaboration quality in real-time. A primary goal of this work is to enable real-time detection of collaboration breakdowns so the ability to detect negative behavior is particularly useful.

**DISCUSSION**

The quality of social regulation and the other behaviors observed in this study differed greatly from previous work investigating adults’ use of social regulation during a tabletop collaborative learning activity. These differences could be due to the age of the participants, or differences between the classroom setting and the lab. Regardless, the prevalence of low-quality social regulation is in line with the Learning Sciences literature that shows that students frequently do not know how to collaborate effectively [19,34]. Given that the same Learning Sciences literature has also demonstrated that ineffective or poor social regulation negatively impacts learning outcomes, we believe that technology for collaborative learning should take into account students’ developing collaboration skills as well as content learning objectives in order to bring about the best possible learning outcomes.

At the end of our lab study [14], it was not possible to state with any certainty that the patterns established as indicators of quality of collaborative processes (Table 3) in a lab study with adults would be valid outside the lab, with younger users, or with different software. The results from this study validate the Unrelated Touches pattern and refine the original Overlapping Sequences pattern to Overlapping Unrelated Sequences. While these results support the notion that collaboration quality can be detected using touch data, further validation would be needed before the patterns identified in this study could be considered generalizable.

Two of the original patterns, Touch Time and Overlapping Sequences, did not indicate quality of content monitoring in this study. We expect that those patterns did not hold because they were a product of the specific interface used in the previous study. Differences in the interaction styles of adults and teenagers could also be a factor. For Touch Time, almost all episodes were labeled as high quality. Because high-quality Touch Time means less time spent interacting with the screen, this result suggests that, even during low-quality collaboration, the teenagers in our study spent less time touching the screen than the adults in [14].
In the case of Overlapping Sequences, the failure to transfer could be due to differences between teenagers and adults. The “helping” behavior seen in all groups of teenagers, that led Overlapping Sequences to be revised as Overlapping Unrelated Sequences, was not seen among the adults. An additional factor could be the level of familiarity between participants – the high school students knew each other well and were friends in many cases. The adults, for the most part, were strangers interacting with each other for the first time. The helping behavior seen with the high school students generally required reaching into each other’s personal space, which may be uncomfortable for strangers.

The results for Unrelated Touches and Overlapping Unrelated Sequences in combination show that particular combinations are often associated with the quality of collaboration processes. When touch data is labeled as low quality according to both Unrelated Touches and Overlapping unrelated Sequences, it is likely that behaviors with a negative impact on the collaboration are occurring. Groups were engaged in negative collaboration behavior in 84.2% of episodes with this quality combination. Conversely, when touch data is labeled as high quality according to Unrelated Touches and Overlapping Unrelated Sequences, it is likely that behaviors with a positive impact are occurring. Groups were collaborating positively in 75% of episodes with this quality combination. Applying these patterns to two-minute intervals of touch data yielded similar results as applying the patterns to the episodes drawn from the video. This result suggests that the interval approach will enable these patterns to be used to detect and respond to collaboration quality in real-time by identifying intervals likely to represent negative collaboration processes as they occur and triggering changes to the interface designed to encourage positive collaboration.

The nature of such adaptations remains to be determined. The data from this study revealed different ways for collaboration to break down, with differing impacts on overall outcomes. For example, group 1 spent an excessive amount of time on low-quality task planning and accomplished relatively little measurable task work despite engaging in high-quality content monitoring. Group 3 combined long periods off-task with frequent low-quality behavioral engagement, leading to a disjointed group dynamic with little cohesion. Group 4 had the fewest episodes of non-collaborative and off-task interactions but two of the students dominated task work to the exclusion of others. These very different collaboration issues suggest that a range of interactive interventions would be needed.

Additionally, it is important to note that as our approach uses only touch data, it is only able to detect collaboration processes while students are actively interacting with the computer. Software adaptations designed to respond to the touch patterns described in this paper will need to take into account the fact that our modeling approach provides insight into groups’ real-time collaborative processes but does not provide the complete picture.

Our video analysis showed episodes of software conflict to be particularly disruptive to groups’ collaborative processes and should therefore be a primary concern of software that can detect and respond to collaboration quality. Software conflicts were generally the result of three types of actions: (1) accidental touches, such as a misplaced elbow or a notebook resting on the screen; (2) intentional disruption – the intentional action of one student, e.g., flicking an on-screen element across the screen to derail other group members’ work; or (3) lack of awareness – occurring when group members who were working independently, unaware or dismissive of what their peers were doing, carried out actions that were in opposition to other group members’ activities, such as clearing the screen while another group member was actively typing.

The ability to distinguish between individuals is essential to using the touch patterns described in this work, most tabletop computers’ inability to natively distinguish between users is an obstacle to utilizing our approach. However, this is an active area of research at the time of writing and a number of approaches to this problem already exist e.g. [6,23,26]. We expect innovation in this area to continue, particularly as tabletop technology moves toward widespread adoption [28].

An additional requirement for this approach is the relationship definition for on-screen objects. While these relationship definitions are simple – elements can only be related or unrelated – relationships have to be defined within the application itself.

**FUTURE WORK**

In our future work, we intend to explore how tabletop software can adapt to encourage positive collaboration learning processes in response to detected breakdowns in collaboration.

As well as helping groups of students to improve their collaboration processes, the approach to detecting collaboration quality described in this paper could be used to help teachers to monitor small group work in their classroom. Providing teachers with real-time information about groups’ collaboration processes could help them target groups that need their help the most.

In this study, we saw many differences between our teenaged participants and the adult participants in our previous study [14]. The teenagers differed from the adults in their interactions with each other, their interactions with the tabletop computer, and their engagement in collaborative work. It seems likely that younger students would interact differently than teenagers and adults. Exploring younger students’ social regulation in tabletop collaborative learning is another area for future work.
CONCLUSION
We have described small groups of students’ use of social regulation processes during collaborative learning activities at a tabletop computer in an authentic classroom setting. Our results show that high school students do not always have the skills to effectively regulate collaborative work, demonstrating the need for interventions to support the development of these skills.

We have identified two touch patterns—Unrelated Touches and Overlapping Unrelated Sequences—that reflect the quality of groups’ collaborative learning processes with up to 84.2% accuracy. We have also demonstrated an approach using these touch patterns in combination to detect the quality of collaborative learning processes in real-time. Our approach to modeling collaborative learning is the first to look at metacognitive processes rather than simply quantifying participation or cognitive performance.

Our empirical results show that the touch patterns we have identified are frequently associated with important social regulation processes. This work represents a significant step towards enabling interactive tabletops to intelligently support and reinforce high-quality collaborative learning.

ACKNOWLEDGMENTS
This work was supported in part by the National Science Foundation under grant IIS-0952786. Any opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect those of the National Science Foundation.

REFERENCES


