

© Copyright 2018

Abigail C. Evans

Adaptive Support for Face-to-Face Collaborative Learning at Tabletop Computers

Abigail C. Evans

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Jacob O. Wobbrock, Co-Chair

Katie Davis, Co-Chair

Cindy Hmelo-Silver, Member

Program Authorized to Offer Degree:

Information Science

University of Washington

Abstract

Adaptive Support for Face-to-Face Collaborative Learning at Tabletop Computers

Abigail C. Evans

Chairs of the Supervisory Committee:
Professor Jacob O. Wobbrock, The Information School
Assistant Professor Katie Davis, The Information School

Collaborative learning is a common practice in today's classrooms. Technology-supported collaborative learning environments are becoming increasingly sophisticated, enabling new ways for students to work together with technology. Research has shown that collaborative learning has many benefits, particularly for developing students' higher-order thinking and problem-solving skills. However, it has also been shown that students do not always know how to collaborate effectively, which can inhibit the success of collaborative learning. These findings suggest that collaboration itself is a skill that needs to be fostered and developed in the classroom.

Tabletop computers have affordances for collaborative learning because of the large, shared interface that multiple people can see and interact with at once. Despite these affordances, small

group work at a tabletop computer is just as susceptible to breakdowns in collaboration as group work using other kinds of tools. Through design-based research in classroom settings, I have investigated how tabletop computers can model social regulation—the processes that groups use to manage and monitor their collective work—in order to detect when a group of students is in need of support. While collaboration is driven by the verbal and gestural interactions between the learners, the tabletop is only able to capture direct interaction with the device.

I have identified touch patterns that reflect the quality of social regulation and can be used to detect problems in the collaborative process. To enable the real-time use of these touch patterns, I developed a machine learning-based approach for distinguishing among simultaneous users at a tabletop computer. I also present software adaptations designed to encourage more effective collaboration that are triggered when breakdowns in collaboration are detected. A classroom evaluation of these adaptations showed that they deterred disruptive behavior and reduced the length of periods of sustained, low-quality collaboration.

My dissertation demonstrates the following thesis:

Interactive tabletop software that can automatically detect breakdowns in collaboration and adapt in real-time to scaffold effective social regulation can improve secondary school students' collaboration skills.

TABLE OF CONTENTS

| | |
|---|------|
| List of Figures | ix |
| List of Tables | xiii |
| Chapter 1. Introduction | 1 |
| 1.1 Contributions..... | 4 |
| 1.2 Organization of Dissertation..... | 4 |
| Chapter 2. Theoretical Background | 6 |
| 2.1 Designing For Personalization in Educational Technology..... | 6 |
| 2.2 Collaborative Learning | 11 |
| 2.3 Social Regulation in Collaborative Learning..... | 12 |
| Chapter 3. Related Work..... | 15 |
| 3.1 Modeling Computer-Supported Collaborative Learning | 15 |
| 3.2 Distinguishing Among Users at a Tabletop Computer | 22 |
| 3.3 Supporting Collaboration and Collaborative Learning..... | 27 |
| Chapter 4. Modelling Collaborative Learning at a Tabletop Computer | 35 |
| 4.1 Study 1: An Exploratory Study of Touch Patterns Associated With Quality of Social Regulation..... | 35 |
| 4.1.1 Data Analysis | 37 |
| 4.1.2 Results: Forms and Quality of Social Regulation..... | 40 |
| 4.1.3 Results: Social Regulation in the Touch Data | 42 |

| | | |
|---|---|-----|
| 4.2 | Study 2: Touch Patterns Associated With Quality of Social Regulation in a Classroom Setting | 44 |
| 4.2.1 | Video Analysis | 48 |
| 4.2.2 | Touch Data Analysis | 51 |
| 4.2.3 | Results | 54 |
| 4.3 | Discussion | 62 |
| 4.4 | Conclusion | 65 |
| Chapter 5. Group Touch – Distinguishing Users at a Tabletop Computer | | 67 |
| 5.1 | Design and Evaluation of Group Touch | 69 |
| 5.1.1 | Creating the Touch Dataset | 70 |
| 5.1.2 | The MLP Model and Feature Selection | 73 |
| 5.1.3 | The MLP Model, Pre-processing, and Training | 75 |
| 5.1.4 | Touch Grouping Algorithm | 77 |
| 5.1.5 | Results | 80 |
| 5.2 | Discussion: Group Touch as an Approach For Distinguishing Among users | 86 |
| 5.3 | Conclusion | 91 |
| Chapter 6. Detecting Collaboration Quality in Real Time Using Group TouCh | | 92 |
| 6.1 | Determining the Optimal Value of the Probability Threshold for Each Touch Pattern | 92 |
| 6.2 | Effect of Using Group Touch on Collaboration Quality Labels | 97 |
| 6.3 | Conclusion | 100 |
| Chapter 7. Design of Adaptations to Encourage Effective Collaboration | | 101 |
| 7.1 | Design Process | 102 |

| | | |
|---|--|-----|
| 7.2 | Implementation of the Adaptations..... | 109 |
| Chapter 8. Evaluation of Software Adaptations to Support Collaborative Learning | | 113 |
| 8.1.1 | Participants..... | 114 |
| 8.1.2 | Apparatus | 115 |
| 8.1.3 | Study Design..... | 119 |
| 8.1.4 | Pilot Test | 123 |
| 8.1.5 | Data Analysis | 135 |
| 8.1.6 | Results..... | 139 |
| 8.2 | Discussion..... | 163 |
| 8.3 | Conclusion | 169 |
| Chapter 9. Conclusion..... | | 170 |
| 9.1 | Summary of Findings..... | 171 |
| 9.1.1 | Modeling Collaboration at Tabletop Computers | 171 |
| 9.1.2 | Distinguishing Users at a Tabletop Computer | 173 |
| 9.1.3 | Using Group Touch to Detect Collaboration Quality in Real Time | 175 |
| 9.1.4 | Adaptations to Encourage Effective Collaboration | 176 |
| 9.2 | Reflections and Discussion..... | 178 |
| 9.3 | Contributions..... | 182 |
| 9.4 | Future Work | 183 |
| 9.5 | Final Remarks | 185 |
| Bibliography | | 186 |
| Appendix A – Example Relationship Definition XML | | 196 |

| | |
|---|-----|
| Appendix B – Survey of Student Attitudes Toward Group Work..... | 197 |
| Appendix C – Screenshots of Learning Applications..... | 201 |

LIST OF FIGURES

Figure 4-1. A screenshot of the software used in Study 1. 36

Figure 4-2. The applications created for this work. Top left: used to conduct design critiques of early wireframes. Top right: used to brainstorm questions for usability testing. Bottom left: helped students to improve their online search skills. Bottom right: focused on Nielsen’s (1994) usability heuristics. 46

Figure 4-3. The data analysis tool I built for this work. This tool includes controls for syncing session video recordings to touch logs. Each row in the table represents a single touch. Clicking on a row in the table skips the video to the corresponding timecode and draws a trace of the touch in the visualization area. The table also includes controls for labeling touches by owner. The pattern evaluation controls were used to test the touch patterns uncovered in this work. 51

Figure 5-1. The mapping application used in the classroom. The two blue squares represent locations with available water quality data. Touching the squares opened up graphs and charts of the data. The shown toolbar provided access to additional navigation and settings. 70

Figure 5-2. The applications used in the second study setting. From left to right, top to bottom, the first three applications used the computer’s native SDK. The fourth (bottom right) application was browser-based. All applications supported multi-touch. 71

Figure 5-3. High school students interacting with the PixelSense. 73

Figure 5-4. The touch features used to predict whether a pair of touches was carried out by the same user are: (1) the difference θ between the touches’ orientations, (2) the distance d between the touch points, and (3) the time between $t1$ and $t2$ 73

Figure 5-5. A raw image captured by Microsoft PixelSense on board cameras showing two different people touching the screen. 74

Figure 5-6. An example of the algorithm to place touches into groups. A) The first touch creates the first group. B) For subsequent touches, we create touch pairs with the last touch of each group, then query the model for each pair. We create a new group if no pairs have a same-author with probability $\geq p$, set to 0.80 in this example. C) If multiple touch pairs have a

same-author probability $\geq p$, we add the touch to the group with the highest probability. 78

Figure 5-7. The ROC curves for each MLP model. The area under the curve, listed in Table 5-1, ranged from 0.89 to 0.99. 82

Figure 5-8. As the value of p increases, the accuracy of Group Touch increases. However, the median time since the last group of touches with the same user and the median duration of groups decreases. 83

Figure 5-9. Average features of touch pairs by the probability they were carried out by the same user according to the MLP model. 85

Figure 6-1. The process of automatically detecting collaboration quality: (1) Group Touch assigns raw touches to groups of touches likely to have been carried out by the same user; (2) sequences of touches are extracted from the groups using the process described in Table 4-1; and (3) sequences in 2-minute intervals are checked against the touch patterns identified in Chapter 4 94

Figure 6-2. The differences between the output of each pattern’s key metrics when using touches that were labeled manually versus touches labeled by Group Touch. The box plots show the distribution of the differences for all two-minute intervals at each probability threshold. The tighter the distribution and the closer to zero on the y-axis, the closer the Group Touch-generated outputs were to those derived from manually labeled touch data. 96

Figure 7-1. The criteria for triggering adaptations. Adaptations are triggered immediately if both touch patterns label a single interval low quality (left-red) or two consecutive intervals are labeled low or medium quality by Unrelated Touches (left-yellow to right-red). The colors in this figure indicate the color of the *group awareness icons* for each possible combination of quality labels produced by the touch patterns..... 111

Figure 8-1. Screenshots of the Snow Leopards 101 application used in the baseline data collection settings. The activity began with a warm-up puzzle (left), then students explored how the four different species of big cat were uniquely adapted to their habitat (center). Finally, students completed a graphic organizer describing each cat’s physical characteristics, habitat characteristics, and unique adaptations (right). 116

- Figure 8-2. Screenshots of the Help a Scientist application. From the home screen (left), students can see an aerial view the scientists’ study area with overlays of snow leopard home ranges. Upon selecting a camera, students are taken to a 3D simulation of the study area (center). Selecting a camera from the 3D view gives students access to the photos taken by the camera (right). The goal of the activity is to label the species captured in the photos taken by each camera..... 117
- Figure 8-3. Screenshots of the Game Challenge application, a partially-built game world that students can edit to bring about particular outcomes. Beginning with a mostly empty play field with a Player (the blue circle) and four containing walls (left), students can add collectable objects, obstacles, and enemies, all of which have editable settings (center). Students test their edits in play mode (right). 117
- Figure 8-4. Group 3 uses the tabletop. The computer was set up in a small breakout room across a hallway from the main classroom. Students had enough room to freely move around the tabletop. A single wide-angle camera was pointed at the screen. The full shot included students faces—the view is cut off in this screenshot to obscure students’ identity.119
- Figure 8-5 The *group awareness icon*. The background changes color (left) to reflect the detected quality of collaboration. Green represents high quality collaboration, yellow represents medium quality collaboration, and red represents low quality collaboration. The icon turns gray when no touches are detected. The screenshot on the right shows the placement of the icons in the software used in this study—top left and bottom right of the screen, next to the “Call help” button that students could use to summon the teacher.124
- Figure 8-6. The Calendar control is blocked by control lockout. 126
- Figure 8-7. The voting adaptation requires all students to vote on global changes to the application. 127
- Figure 8-8. *Escalate to authority* sends an alert to the teacher's mobile device via the Slack API. This intervention is silent from the students’ point of view—they are not notified that the teacher has been summoned and there are not visible changes to the UI. 128
- Figure 8-9. A prompt encourages group members to revisit task instructions, listen to each other, and coordinate their actions. 134

Figure 8-10. Social regulation and collaboration codes over time for Group 1. Blue bars represent episodes of high quality social regulation/positive group work codes. Orange bars represent low quality social regulation/negative codes. Dashed bars represent interactions taking place away from the computer. Gray blocks represent interruptions by program staff, which were not coded. 141

Figure 8-11. Social regulation and collaboration codes over time for Group 2. Blue bars represent episodes of high quality social regulation or positive group work codes. Orange bars represent low quality social regulation or negative codes. Dashed bars represent interactions taking place away from the computer. Gray blocks represent interruptions by program staff, which were not coded. 142

Figure 8-12. Social regulation and collaboration codes over time for Group 3. Blue bars represent episodes of high quality social regulation or positive group work codes. Orange bars represent low quality social regulation or negative codes. Dashed bars represent interactions taking place away from the computer. 143

Figure 8-13. Group 1 triggers the prompt for the second time in their second experimental session. The two students on the left were engaged in the activity. The student at the top right disrupted the activity by hammering on the screen. The student in the bottom right was sitting back from the tabletop and using his phone..... 156

LIST OF TABLES

| | |
|--|----|
| Table 2-1. The seven principles of ability-based design (reproduced from Wobbrock et al., 2018). | 8 |
| Table 2-2. The processes and sub-processes of social regulation, adapted from Kempler Rogat and Linnenbrink-Garcia (2011). | 13 |
| Table 4-1. Rules for grouping touch event data into purposeful actions by time between touches. | 39 |
| Table 4-2. Codes used in the video analysis. Social regulation codes adapted from (Rogat & Linnenbrink-Garcia, 2011)..... | 49 |
| Table 4-3. Touch patterns associated with quality of content monitoring. Quality thresholds are derived from the results of the lab study..... | 52 |
| Table 4-4. Quality of social regulation by group, shown as the percentage of each group's social regulation episodes in all activities combined. | 55 |
| Table 4-5. Mean number of episodes of each additional code per activity by group. | 56 |
| Table 4-6. 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 the lab study and the revised pattern found in this study. | 56 |
| Table 4-7. 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..... | 59 |
| Table 5-1. Group Touch results by session. Columns (d) and (e) show how just the MLP model performed on the test data for each session. Column (f) shows the performance of the | |

grouping algorithm with threshold $p=0.80$. Column (f) is therefore the performance of Group Touch. 81

Table 6-1. The number (#) of intervals by the quality label produced using Group Touch and the percentage of intervals where that quality label matched the label generated using the manually-labeled touches..... 98

Table 6-2. The distribution of collaboration quality codes from the video data (Chapter 4, Section 4.2.3) for each possible output of the touch patterns using Group Touch to distinguish among users. Note that intervals labeled high quality by both touch patterns (first data row in table) were mostly made up of episodes coded positively during video analysis and that intervals labeled low quality by both touch patterns (last data row in table) were mostly made up of episodes coded negatively during video analysis. 99

Table 7-1. All adaptations designed during initial brainstorming. Note: this table spans multiple pages. 103

Table 8-1. Inter-rater reliability statistics by code. 137

Table 8-2. The occurrence of each video code by group for the baseline data collection sessions, in which the groups used the *Snow Leopards 101* application. 146

Table 8-3. Summary of video codes from the experimental sessions. Group 1 received the adaptations in both sessions. Group 2 was the control. Group 3 used the application without adaptations in the first session, then with the adaptations in the second session. *Group 3’s episodes of *non-collaborative interactions* were coded as *task work*-only by the video coder. I added the *non-collaborative interactions* code for task work episodes during which one student was working alone and the other two students were chatting. 151

Table 8-4. The median length of sustained periods of low-quality collaboration at the tabletop in each of Group 1’s sessions. The “pre-adaptation” row shows the median length of periods of low-quality collaboration during intervals that caused an adaptation to be triggered. “Adaptation present” refers to periods of low quality collaboration that occurred while a triggered adaptation was active. “No adaptation” covers all other intervals. Episodes of negative collaboration that took place away from the computer are not included. 158

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of many people.

I would like to thank my advisors Jacob O. Wobbrock and Katie Davis, for their mentoring, guidance, and encouragement over many years. I also thank the past and present members of my committee for their support and advice: Cindy Hmelo-Silver, Katie Headrick Taylor, Batya Friedman, Eliza Dresang, and Heather Hebard.

Several organizations were important partners in this work: the Snow Leopard Trust, UW Upward Bound, the Pacific Science Center, and Woodland Park Zoo.

I would like to thank the current and former iSchool staff who helped with all things logistics: Shannon Gilmore, Cortney Leach, Jennifer Phipps, Frankie Fowler, and Dora Tkach; my fellow doctoral students, past and present, who provided moral support (and manual labor in some cases): Rachel ‘Ivy’ Clarke, Elisabeth Jones, Chris Heaney, Ian King, Martez Mott, Jason Portenoy, Kristen Shinohara, Abdullah Ali, Katie O’Leary, Norah Abokhodair, Liz Mills, Alex Jansen, Parmit Chilana, Katherine Thornton, Shaun Kane, Jordan Eschler, Michael Lee, Jeff Huang, Shawn Walker, Sarah Evans, and Alex Mariakakis.

I thank my mother, Jeanne Webster, for always encouraging me. Even though she is no longer here, I know she would be proud. My in-laws Val and Ben Evans, and cousin, Jill Da Silva, have also helped kept me going over the last seven years.

Michael Evans deserves special thanks for following me 6,000 miles from home and showing enormous patience and love even though it’s been difficult at times.

Thank you to Lantern Brewing staff and regulars for providing the fuel and friendship that made the final push a lot easier. Thanks also goes to numerous friends not already mentioned,

especially those who have tolerated me working on this dissertation at their birthday parties and other social events.

Finally, this work would not have been possible without funding from the National Science Foundation and Microsoft Research's generous gifts of a PixelSense and PPI display.

DEDICATION

To Michael Evans

In memory of Jeanne Webster and Nigel Webster

Chapter 1. INTRODUCTION

Researchers and designers of educational technology have long sought to create personalized learning environments that can provide targeted, just-in-time support to learners (e.g. Jermann, Muhlenbrock, & Soller, 2002; Lonchamp, 2010; Marcos-Garcia, Martinez-Mones, Dimitriadis, & Anguita-Martínez, 2007; Martinez, Collins, Kay, & Yacef, 2011). Collaborative learning environments represent a particular challenge in personalization, as they must take into account interactions between multiple learners. While much has been done to bring adaptive support to asynchronous, text-based, online collaborative learning environments, face-to-face collaborative learning environments such as tabletop computers bring further challenges, as much of the learning takes place in the verbal interactions between group members that are invisible to the computer.

Tabletop computers have horizontal, multi-touch screens large enough to accommodate multiple users at one time. Featuring a large, shared interface, tabletop computers are said to afford collaboration as multiple people can see and interact with the same content at the same time (Buisine, Besacier, Aoussat, & Vernier, 2012; Dillenbourg & Evans, 2011; Higgins, Mercier, Burd, & Hatch, 2011). Examples include the Microsoft PixelSense and the SMART Table. While tabletop computers are not yet commonplace in classrooms, they are predicted to see widespread adoption in the coming years (Muller-Tomfelde & Fjeld, 2012). It is important that emerging technologies such as this are a focus of education research (U.S. Department of Education, 2010). A thorough understanding of the benefits of a new technology can help to ensure that, when it eventually arrives in schools, it improves teaching and learning and is not simply used as an expensive version of existing low-tech alternatives, or not used at all, as is often the case (Aldunate & Nussbaum, 2013).

Small group face-to-face collaborative learning commonly features in initiatives for reform and modernization of K-12 education, such as the Framework for 21st Century Skills (The Partnership for 21st Century Skills, 2009), the Next Generation Science Standards (National Research Council, 2015), and the Common Core State Standards (National Governors Association Center for Best Practices, 2010). Therefore, it is important that collaborative groups are not left out of the drive for personalized education. The process of collaborative learning is complex and many factors contribute to learning outcomes. The effectiveness of the group learning experience can be influenced by the personalities and motivations of individual students, their background knowledge, and the nature of the task. Students do not always know how to collaborate in a manner that is productive and conducive to learning for all members (Dillenbourg & Jermann, 2007; Järvelä & Hadwin, 2013; Rogat & Linnenbrink-Garcia, 2011). In the classroom, the teacher will often circulate among groups, but she can only visit one group at a time, meaning that struggling groups may not receive the help they need in a timely manner. Tabletop software that can adapt to a group's collaboration needs and abilities in real-time could prove useful in scaffolding collaboration while the teacher spends time with other groups.

While the primary objective of this research was to develop software that can adapt to help high school students make the most of small group collaboration, the ability to automatically model their process could also support teachers with longer term planning and classroom management concerns. For example, a group that consistently engages in weaker social regulation than their peers might mean that a particular combination of individuals is unable to work together effectively. Additionally, cases where many groups are unable to collaborate effectively may indicate that the assigned task is too challenging or too easy.

The goals of this dissertation were 1) to uncover what physical interactions with a tabletop computer reveal about collaborative interactions taking place above the table; and 2) to develop tabletop software that can adapt to scaffold effective collaboration. Rather than trying to determine if a group is meeting content-based learning objectives, my intention was to model and adapt to the collaboration itself. While it may not always be true in practice, an assumption in this work is that effective collaboration leads to positive learning outcomes, while ineffective collaboration hinders learning (Dillenbourg & Evans, 2011; Rogat & Linnenbrink-Garcia, 2011). The research questions addressed in this dissertation are:

- 1) What can learners' physical interactions with a tabletop computer reveal about the group's social regulation processes? With this question I sought to scope out the specific regulatory processes that can be detected from physical interactions alone.
- 2) How can a tabletop computer distinguish among users? Although this question is not directly tied to collaborative learning, it had to be answered in order to use the findings from research question 1 to model collaboration in real time.
- 3) How can tabletop software adapt to a group's social regulation processes in order to support effective collaboration in real time? Addressing this question involved the design of adaptations to be triggered in response to social regulation processes automatically detected in light of the results of questions 1 and 2.
- 4) Do the adaptations encourage effective collaboration? This question included the following sub questions: a) Are the adaptations triggered appropriately? b) How do the students respond?

This dissertation demonstrates the following thesis:

Interactive tabletop software that can automatically detect breakdowns in collaboration and adapt in real-time to scaffold effective social regulation can improve secondary school students' collaboration skills.

1.1 CONTRIBUTIONS

The main contributions of this dissertation are:

- 1) A set of touch patterns that can be used to model and detect the quality of social regulation during small group collaborative learning at a tabletop computer.
- 2) An approach to distinguishing among users that can enable real-time detection of breakdowns in collaboration.
- 3) Application-independent interface adaptations that can scaffold effective collaboration at tabletop computers.

1.2 ORGANIZATION OF DISSERTATION

In Chapter 2, I discuss the theoretical background of my dissertation work. I describe approaches to designing for personalization in educational technology, drawing on both human-computer interaction and educational research. I also characterize collaborative learning and introduce the learning sciences concept of social regulation of collaborative learning, which guides much of the data analysis in subsequent chapters.

In Chapter 3, I cover related work in three areas important to this dissertation: modeling computer-supported collaborative learning, distinguishing among users at tabletop computers, and supporting collaboration and collaborative learning. As with the theoretical background discussed in Chapter 2, the related work described in Chapter 3 comes from both human-computer interaction and educational research, particularly the learning sciences.

Chapter 4 presents my early work on modeling collaboration at tabletop computers. This chapter describes two studies: a lab study of adults collaborating on a poetry analysis task and a classroom study of high-school students collaborating on a series of activities in a user-centered design course. In this chapter, I present a pair of touch patterns that were demonstrated to reflect aspects of the quality of collaboration at a tabletop computer.

In order to use the touch patterns presented in Chapter 4 to detect collaboration quality in real time, it is important to be able to distinguish among simultaneous users at a tabletop computer. As most commodity tabletops do not have this capability, in Chapter 5, I present Group Touch, a novel approach for distinguishing among users designed with the specific constraints of classroom settings in mind.

In Chapter 6, I bring together the findings from Chapters 4 and 5 to finalize my approach for using Group Touch with the touch patterns described in Chapter 4 to detect collaboration quality in real time, and to assess the implications of using this approach.

In Chapter 7, I present eleven potential software adaptations for encouraging effective collaboration in response to detected breakdowns in collaboration.

In Chapter 8, I present an empirical evaluation of a subset of the proposed adaptations. This evaluation, conducted in a classroom setting with high school students, shows that the adaptations have potential to increase effective collaboration by reducing disruptive behavior.

I close this dissertation with Chapter 9, a summary of the major findings and contributions of this work as well as a discussion of its limitations, the challenges of working with emerging technology, and opportunities for future work.

Chapter 2. THEORETICAL BACKGROUND

2.1 DESIGNING FOR PERSONALIZATION IN EDUCATIONAL TECHNOLOGY

The driving force behind adaptive, personalized learning software is a desire to make the learning experience more inclusive and better able to support the needs of all students. My dissertation research is grounded in two distinct but complementary approaches to the design of inclusive, computer-supported, learning environments: Universal Design for Learning (Rose & Meyer, 2002) and Ability-Based Design (Wobbrock, Gajos, Kane, & Vanderheiden, 2018; Wobbrock, Kane, Harada, Froehlich, & Gajos, 2011). I also use an additional framework for designing pedagogical learning interventions for student use (Wise, 2014) as a supplement to Universal Design for Learning and Ability-Based Design.

Based on research in educational neuroscience, Universal Design for Learning (UDL) is an approach to the design of learning environments that seeks to enable all students to become “expert learners” through flexible, differentiated instruction (CAST, 2011; Rappolt-Schlichtmann, Daley, & Rose, 2012; Rose & Meyer, 2002). According to UDL, barriers to student learning are found in the interactions between the student and the curriculum, not in the students themselves, and the learning environment is responsible for adapting to students’ needs in the face of poor performance. The “learning environment” as described by UDL is often used interchangeably with “curriculum,” encompassing the pedagogical methods, tools and materials, and assessment. A key tenet of UDL is the separation of learning objectives from the methods and materials used to achieve and assess them. Technology has an important role in UDL as a flexible tool that can support multiple pathways to reach the same learning objectives, allowing students to follow the path best suited to their abilities. My dissertation work is aligned with the UDL stance that the

learning environment, of which the computer is one piece among many, should adapt in response to poor performance.

UDL is concerned with the design of the learning environment as a whole, of which technology is just a small part. Therefore, the UDL approach focuses on selecting a variety of largely non-adaptive digital resources and stops short of realizing the potential of computers to model and adapt to students' needs and abilities. This is likely because the primary audience for UDL is education practitioners who are unlikely to have the technical expertise required to build adaptive software. Ability-based design (ABD) is a complementary approach from human-computer interaction (Wobbrock et al., 2011) that has a narrower focus on the technology itself.

Although the context of ABD is very different from that of UDL, the underlying principles have much in common. ABD is an approach to the design of accessible software that focuses on what the user can do rather than what they cannot, and puts the burden of adaptation on the system rather than the user. Traditionally concerned with software users' physical abilities, ABD adaptations have so far sought to make the physical actions required to operate a computer system easier. The approach, however, can apply to design of software that can adapt to users' cognitive abilities. The seven principles of ability-based design are given in Table 2-1.

Table 2-1. The seven principles of ability-based design (reproduced from Wobbrock et al., 2018).

| | Principle | Description |
|--|------------------|---|
| Designer Stance | Ability | Designers will focus on users' abilities, not disabilities, striving to leverage all that users <i>can</i> do in a given situation, context, or environment. |
| | Accountability | Designers will respond to poor usability by changing systems, not users, leaving users as they are. |
| | Availability | Designers will use affordable and available software hardware, or other components that are acquirable through accessible means. |
| Adaptive or Adaptable Interface | Adaptability | Interfaces might be adaptive or adaptable to provide the best possible match to users' abilities. |
| | Transparency | Interfaces might give users awareness of adaptive behaviors and what governs them, and the means to inspect, override, discard, revert, store, retrieve, preview, alter, or test those behaviors. |
| Sensing and Modeling | Performance | Systems might sense, monitor, measure, model, display, predict, or otherwise utilize users' performance to provide the best possible match between systems and users' abilities. |
| | Context | Systems might sense, monitor, measure, model, display, predict, or otherwise utilize users' situation, context or environment to anticipate and accommodate effects on users' abilities. |

My dissertation work supports a number of ABD principles. The motivation for this project draws on the *ability* and *accountability* principles, which are largely the same as the stance of UDL. I have sought to model tabletop collaborative processes (*performance*) and investigate interface *adaptation* to support more effective collaboration. The adaptations also follow the *transparency* principle to some degree.

Design of educational technology requires consideration of both pedagogy and technology. Neither ABD or UDL covers the complete picture, or even the same aspect of the complete picture.

Ability-based design addresses technology design but not the design of the learning environment as a whole. UDL addresses the larger learning environment but not technology design (at least not to the same depth as ability-based design). For example, UDL aims to accommodate all learners through a design-for-one approach to the curriculum that advocates personalized learning and differentiated instruction (Rose & Meyer, 2002). Any effort to design personalized learning necessarily involves learner modeling (*performance*) in the form of assessment and observation, then adaptation to the learner's abilities (*adaptation*) through the collection or development of differentiated materials and instructional strategies. In the UDL context, performance modeling and adaptation is typically carried out by the teacher as designer of the curriculum. Applying ABD principles, *performance* and *adaptation* in particular, to educational software could lead to software that can actively help the teacher to personalize the curriculum through dynamic modeling and adaptation. Given that the two approaches share the same stance on ability and accountability, they are inherently complementary for design of software for use within a learning environment that accounts for varying abilities among students.

When developing new approaches to modeling and adapting to support collaborative learning, it is important to consider how such support should be presented to students and situated in the larger context of a learning activity or course. Wise (2014) notes it is not enough to simply present students with interventions for learning—if students are to benefit from an intervention, they need to have both an understanding of its intent and how to engage with it. Wise presents a framework for designing pedagogical learning analytics interventions that I use to guide the deployment of adaptive supports for collaboration. The framework consists of four design principles and three processes that students need to utilize in order for an intervention to be successful:

- **Principle 1: Integration.** The use of an intervention should be an “integral part of course activity tied to goals and expectations” (Wise, 2014, p.206).
 - *Process 1: Grounding.* Within the integration principle is the process of grounding. To be effective, student use of an intervention needs to be grounded in an understanding of the purpose of the learning activity, the characteristics of productive engagement in the activity, and how the intervention relates to those characteristics. Students also need to know of any aspects of productive engagement that are *not* detected by the systems they work with to ensure that positive, productive behaviors are not devalued simply because the technology cannot detect them.
- **Principle 2: Agency.** Effective learning requires that students are actively engaged. Therefore, interventions should encourage students to have agency in their own learning.
 - *Process 2: Goal setting.* An important aspect of students having agency over their own learning is setting goals for what they want to achieve, which can then help them decide how to interpret and act on an intervention. Goal setting does not have to explicitly take place within the system itself but it should be part of the larger educational context.
 - *Process 3: Reflection.* Interventions that are triggered in response to students’ past actions within a system should provide an opportunity to help students to reflect on their actions and understand why the intervention was triggered.
- **Principle 3: Reference Frame.** When an intervention is triggered, it should provide a reference frame, or perspective, that students can use to interpret why it was triggered and how they should respond to it. There are three possible reference frames: how students’

behaviors measure up to the goals and expectations of the learning activity, how their current learning behaviors compare to their past experiences in similar activities, or the behavior of other students.

- **Principle 4: Dialogue.** Interventions should be framed as opportunities for reflection and dialogue between students and teachers rather than as systems that monitor and enforce student behavior.

Wise's framework aligns with some of the principles of ABD, particularly *adaptation* and *transparency*. The framework is a useful supplement to ABD because it provides additional guidance on how adaptive technological interventions should function in learning contexts, which have different considerations than general technology use.

2.2 COLLABORATIVE LEARNING

Collaborative learning can take many forms depending on context and the intended outcome (Dillenbourg, 1999). I use Roschelle and Teasley's (1995, p.70) definition of collaborative learning as "a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem."

I also draw on the perspective of collaborative learning as knowledge building (Scardamalia & Bereiter, 1994; Stahl, 2006). In knowledge-building classrooms, groups work together to solve a problem and produce new knowledge of value to the community of learners as a whole. Intentionality and community have central importance in the learning process. Through dialogue, referred to as knowledge-building discourse, the students negotiate and refine their individual understanding of the problem to reach a shared understanding. Social interaction is also important for coordination and management of the collaboration, as well as the cognitive aspects. The relationship between the individual and collective learning processes is symbiotic; the two

cannot be separated as each influences the other and both are important. The process is cyclical, with the students continually redefining the problem space as their understanding develops.

2.3 SOCIAL REGULATION IN COLLABORATIVE LEARNING

In order to provide appropriate computer support for knowledge-building discourse, it is important to understand the conditions that lead to productive discourse. Social regulation refers to “the social processes groups use to regulate their joint work on a task” (Rogat & Linnenbrink-Garcia, 2011, p.377). Social regulation is an extension of the concept of self-regulation in individual learning to groups of learners in a collaborative context (Järvelä & Hadwin, 2013; Panadero & Järvelä, 2015; Rogat & Linnenbrink-Garcia, 2011; Volet, Vauras, & Salonen, 2009).

Järvelä and Hadwin (2013) describe three forms of regulation that come into play in collaborative learning: self-regulated learning of individual students; co-regulation, in which group members support each other’s regulation; and socially-shared regulation, in which the group regulates learning as a whole entity. All three forms of regulation are important in collaboration, but a number of studies have shown that socially-shared regulation in particular is correlated with better group outcomes (e.g. Khosa & Volet, 2014; Lajoie & Lu, 2012; Malmberg, Järvelä, Järvenoja, & Panadero, 2015; Volet et al., 2009) and associated with higher levels of participation from all group members (Isohätälä, Järvenoja, & Järvelä, 2017).

Järvelä and Hadwin (2013) argue that, although socially-shared regulation and joint construction of knowledge often occur together, socially-shared regulation is a distinct process, referring to metacognitive processes independent of domain knowledge. Taking this view, it becomes possible to model and respond to social regulation processes independent of the specific task or content domain. The exact nature of social regulation—when and how it manifests during a learning activity—evolves over time and is shaped by the context—the task, the learners, their

skills and past experiences, and the physical environment (Järvelä, Malmberg, & Koivuniemi, 2016; Järvenoja, Järvelä, & Malmberg, 2015; Kempler Rogat, Linnenbrink-Garcia, & DiDonato, 2013; Lee, O'Donnell, & Rogat, 2015; Malmberg et al., 2015).

Kempler Rogat and Linnenbrink-Garcia (2011) identify three dimensions of social regulation (see Table 2-2): *planning* the group's approach to a task, *monitoring* of understanding and progress, and *behavioral engagement*—efforts to get group members to engage with the task. A group's use of each dimension can vary in quality, with high-quality social regulation processes leading to socially-shared regulation, in which groups maintain joint attention on the learning activity. Conversely, low-quality social regulation processes are often associated with off-task behavior. Another example of low-quality social regulation is *other-regulation*, in which one student directs and dominates the others. Kempler Rogat and Linnenbrink-Garcia (2011) and Lee et al. (2015) note a correlation between high-quality *content monitoring* and groups' construction of shared meaning, a central component of knowledge-building discourse. Content monitoring is a sub-dimension of the monitoring process and is characterized by effort to evaluate and further the group's understanding of task content.

Table 2-2. The processes and sub-processes of social regulation, adapted from Kempler Rogat and Linnenbrink-Garcia (2011).

| Social regulation process | Description |
|---|---|
| Planning - task - content | Reading and interpreting task directions, designating task assignments, discussing how to go about solving the problems |
| Monitoring - content - plan - progress | Evaluating content understanding, the shared product, assessing progress, or plan for completing the task |
| Behavioral engagement | Encouraging an off-task group member to re-engage, reminding a group member to return to task |

Kempler Rogat and Linnenbrink-Garcia's (2011) dimensions of social regulation represent the framework guiding my research. Their framework is one of a number of published alternatives (De Backer, Van Keer, & Valcke, 2015; Järvenoja & Järvelä, 2011; Khosa & Volet, 2014; Panadero & Järvelä, 2015). Social regulation of learning is still a relatively new area of study and, although most researchers appear to agree on the nature of social regulation at a high level, there is currently no widely accepted terminology or clear distinctions between its various facets. I chose to use Kempler Rogat and Linnenbrink-Garcia's (2011) framework because they provide detailed descriptions of each social regulation process and examples that illustrate the differences between high- and low-quality versions of the processes. Elements of Kempler Rogat and Linnenbrink-Garcia's framework have informed many studies of group processes in collaborative learning (e.g. Rogat & Adams-Wiggins, 2014; Sinha, Rogat, Adams-Wiggins, & Hmelo-Silver, 2015; Xu, Du, & Fan, 2013).

Prior work has demonstrated the relationship between effective collaboration and positive learning outcomes at both the individual and group level (e.g. Barron, 2003). In this work, the term "effective collaboration" is referring to high-quality, socially-shared regulation of the collaboration rather than any specific learning outcome. For example, in this context, a group that successfully completes a learning activity but got there primarily through one group member's other-regulation of her teammates would not be considered to have demonstrated effective collaboration.

Chapter 3. RELATED WORK

The goals of my dissertation research are: 1) to develop an approach to detecting potential breakdowns in collaborative learning at a tabletop computer via task-independent modeling of social regulation processes, and 2) to design just-in-time interface adaptations that will encourage students to work together more effectively once a breakdown is detected. Although no research published thus far has shared these specific goals, a range of prior work has investigated ways to detect processes and features of collaborative learning in computer-supported collaborative learning (CSCL) environments, including tabletops. Additionally, a number of studies have explored how CSCL environments can support and scaffold effective collaboration. In this chapter, I discuss related work in each of these areas, as well as a third area—distinguishing among users at a tabletop computer—which, though not directly related to collaborative learning, represents a challenge that must be addressed in order to enable automatic detection of social regulation at tabletop computers.

3.1 MODELING COMPUTER-SUPPORTED COLLABORATIVE LEARNING¹

Before any technology can adapt to support collaboration, it needs to be able to model and detect collaboration processes to determine when adaptation would be advantageous. Many researchers have explored ways to use traces of students' interactions with technology to model various aspects of collaborative learning including regulatory processes associated with high- and low-achieving groups (Schoor & Bannert, 2012), optimal group formation (Berland, Davis, & Smith, 2015;

¹ Portions of this section reproduce text from Evans, Wobbrock, & Davis. (2016). Modelling collaboration patterns on an interactive tabletop in a classroom setting. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '16)* (pp. 860–871). New York.

Erkens, Bodemer, & Hoppe, 2016), and the nature and quantity of individual students' contributions (Dascalu, Trausan-Matu, McNamara, & Dessus, 2015; Trausan-Matu, Dascalu, & Rebedea, 2014).

Most research in this area has focused on online, text-based environments, where students communicate with each other through synchronous chat tools or asynchronous discussion boards. These environments have the advantage of being able to capture a more complete trace of student learning than is possible with non-text environments, as all students' interactions with each other and the system can be logged. Many researchers have used techniques such as text mining and natural language processing to model and detect elements of collaboration in text-based environments, e.g., (Dascalu et al., 2015; Erkens et al., 2016; Lonchamp, 2010; Mu, Stegmann, Mayfield, Rosé, & Fischer, 2012; Rosé et al., 2008; Trausan-Matu et al., 2014; Wang, Rosé, & Chang, 2011). The specific applications of these techniques have ranged from automatically analyzing and categorizing student contributions to conversation (Rosé et al., 2008) to producing visualizations that represent how participation evolves over time (Trausan-Matu et al., 2014).

Text mining has enabled some of the most sophisticated examples of adaptive support for collaboration, such as computer-based agents that can interact with learners to provide just-in-time feedback and coaching, e.g., (Wang et al., 2011). However, these approaches depend upon entirely computer-mediated, text-based communication between learners. For this reason, little practical insight can be transferred from text-based approaches to modeling collaboration to tabletop environments, where communication takes the form of verbal and physical interactions between co-located learners.

An emerging area of research that could prove useful in co-located collaboration settings is the use of sensors to model and detect physical, non-verbal indicators of collaboration quality.

Schneider & Pea (2014) used eye-tracking to identify moments in which both students in a dyad maintained joint attention on the same objects within a synchronous, distributed learning environment. Joint attention was found to be predictive of collaboration quality and would likely continue to be a relevant marker of collaboration quality in tabletop activities, where students synchronously interact with the same interface. There are precedents for using eye tracking in tabletop environments, although complicated external equipment (Yamamoto, Komeda, Nagamatsu, & Watanabe, 2010) or cumbersome wearables (Van Der Meulen, Varsanyi, Westendorf, Kun, & Shaer, 2016) are required for these systems, suggesting that eye tracking for tabletop computers is not quite ready for classroom use.

Pijeira-Diaz et al. (2016) used biosensors embedded in wristbands to monitor learners' physiological responses throughout an activity. The researchers did find physiological signals that were correlated with some measures of collaboration, such as learning gains, but the environmental conditions required for the biosensors to be effective are too sensitive for everyday classroom use. Additionally, the use of biosensors for tracking students' psychological states during learning may raise ethical concerns.

The majority of the related work discussed thus far has sought to model and/or detect various features of collaborative learning, but none have touched on social regulation specifically. Most research on social regulation of collaborative learning has had the primary aim of describing, rather than modeling, and distinguishing its various forms from one another. More recently, however, attention has turned to modeling how social regulation processes evolve over time during a single activity and over periods of sustained collaboration lasting multiple sessions (Järvelä et al., 2016; Malmberg, Järvelä, & Järvenoja, 2017; Molenaar & Chiu, 2015; Schoor & Bannert, 2012).

Schoor and Bannert (2012) used process mining techniques to identify and compare sequential patterns of regulation in high- and low-achieving dyads working in a text-based environment. The patterns were derived from chat logs manually coded for social regulation processes, and they shed light on how social regulation processes work together. For example, a common loop of processes was coordination of the task, followed by task work, followed by monitoring of group work, looping back to coordination of the task. Surprisingly, there were no differences between high- and low-achieving groups in terms of the sequences of regulatory processes used or their frequency. However, in contrast to other research (e.g., Kempler Rogat & Linnenbrink-Garcia, 2011), Schoor and Bannert did not code for the quality of the processes that occurred. For example, they coded for task planning, but did not specify whether that planning was high- or low-quality. Perhaps differences between high- and low-achieving groups might have been apparent with a coding scheme that accounted for quality as well as occurrence of social regulation processes.

Järvelä, Malmberg, and Koivuniemi (2016) analyzed chat and interaction logs from a text-based learning environment to identify sequences of self- and socially-shared regulation in high- and low-achieving groups. They found that high-achieving groups progressed from self-regulation processes among individual students to socially-shared regulation. In contrast, there were no instances of socially-shared regulation among the low-achieving groups. Additionally, students in high-achieving groups were more interactive.

Molenaar and Chiu (2015) also found that high-achieving groups, this time in a face-to-face environment where multiple students worked at a single computer, followed different sequential patterns of social regulation processes than low-achieving groups. Not only were high-

achieving groups more likely to engage in high quality discussion of the learning content, but these discussions formed longer sequences than were found in low-achieving groups.

Malmberg, Järvelä, and Järvenoja (2017) studied the sequencing, timing and frequency of self-, co-, and socially-shared regulation, as well as regulation sub-processes such as planning and monitoring, in a face-to-face undergraduate collaborative learning program lasting several months. The sequences of regulation processes were derived from videos of the students collaborating. Among the findings was that co-regulation can lead to socially-shared regulation, the ideal form of regulation in collaborative learning. This finding suggests that co-regulation may be important to fostering socially-shared regulation, which in turn has implications for designing tools to support social regulation of learning.

These studies of how regulation emerges and evolves over time are a first foray into modeling, but they stop short of enabling automatic detection of key processes. In each case, a substantial amount of manual coding of either text-based communication or video data was required before modeling could begin. For collaborative learning in text-based environments, utilizing some of the text mining techniques described at the start of this chapter (e.g., Dascalu et al., 2015; Erkens et al., 2016; Lonchamp, 2010; Mu, Stegmann, Mayfield, Rosé, & Fischer, 2012; Rosé et al., 2008; Trausan-Matu et al., 2014; Wang, Rosé, & Chang, 2011) would be a logical next step. Automatically detecting social regulation in face-to-face learning environments, such as tabletop computers, will require a different approach.

Social regulation occurs in the interactions between group members and in their interactions with the learning environment. In text-based environments, the computer can log all facets of students' interactions in their entirety. When students collaborate at a tabletop computer, however, only their direct, physical interactions with the interface can be logged. Therefore, in

order to model collaborative learning processes at tabletop computers, it is important to understand how physical interactions with the computer relate to the collaboration as a whole.

Fleck et al. (2009) explored the relationship between groups' verbal interactions and their physical actions on the tabletop. She found that verbal elements of successful collaboration, such as making and accepting suggestions or negotiating, were often complemented by particular actions in the software.

Pontual-Falcão and Price (2010) investigated instances of "interference" that interrupted or changed the flow of collaboration during a tabletop activity. Although it is often assumed that such interference hinders the collaborative process, Pontual-Falcão and Price found that the effect of the interference depended on how the group responded to it. Interference was productive when groups treated it as a contribution to the task and integrated it into their discussion. Interference was counter-productive when groups dismissed it or immediately sought to undo or reverse the change it created. Pontual-Falcão and Price found that interference primarily took place through physical rather than verbal interactions.

Both of these studies demonstrated that for collaborative learning at a tabletop computer, physical interactions are an integral part of the collaboration process, often directly coupled to or influencing verbal interactions. My initial research, described in 0, builds on the prior work of Fleck et al. (2009) and Pontual-Falcão and Price (2010) to investigate what physical interactions with a tabletop computer can reveal about a group's social regulation processes.

To date, no published research has attempted to model or detect social regulation of collaborative learning with tabletop computers. However, several studies have looked at modeling and detecting other aspects of collaborative learning at tabletop computers, including interaction patterns associated with high- and low-achieving students (Al-Qaraghuli, Zaman, & Olivier, 2011;

Martinez-Maldonado, Dimitriadis, Martinez-Monés, Kay, & Yacef, 2013; Martinez-Maldonado, Yacef, Kay, Al-Qaraghuli, & Kharrufa, 2011) and quantifying individual participation in group activities.

Martinez-Maldonado et al. used data mining techniques to extract application-specific sequential patterns of interaction associated with high- and low- achieving groups of elementary students (Martinez-Maldonado, Yacef, & Kay, 2013; Martinez-Maldonado et al., 2011) and adults (Martinez-Maldonado, Dimitriadis, et al., 2013). Using video data coded by teachers, Al-Qaraghuli, Zaman, & Olivier (2011) identified application-specific interaction patterns associated with high- and low-achieving students within groups. In all of these studies, the process modeling is tightly integrated with the software that students use—the patterns detected are specific to the software. It is unclear whether the patterns detected would be applicable with different tasks in different applications.

Martinez-Maldonado et al. (Martinez, Collins, et al., 2011; Martinez, Kay, & Yacef, 2011) also developed visualizations to show how much individual group members contributed in verbal utterances and touch interactions. External sensors were used to capture and track individual students' speech and physical interactions with the computer. The resulting visualizations clearly showed the balance of participation among group members but could not reveal the quality. In a later study, they combined these visualizations with a model task solution created by a content expert. Teachers were able to compare groups' progress on a task to the expert model, enabling them to quickly identify, in real-time, groups that needed support (Martinez-Maldonado, Kay, Yacef, Edbauer, & Dimitriadis, 2013; Martinez-Maldonado, Kay, Yacef, & Schwendimann, 2012). Although Martinez-Maldonado et al.'s approach yielded positive results, it relies on the availability of an expert model, or clear "right answer," which is not appropriate for open-ended

or creative activities. In contrast, a goal of my work is to develop a model of groups' social regulation processes that can be used in a variety of contexts, including open-ended, collaborative, knowledge-building tasks. This goal requires an approach that is *independent* of the specific task or learning objectives.

An additional contrast between Martinez-Maldonado et al.'s approach and mine is the use of external sensors to capture the verbal aspects of collaboration that are not otherwise visible to the computer. While Martinez-Maldonado et al. have successfully used additional external hardware to capture and quantify verbal interactions among group members (Martinez-Maldonado, Dimitriadis, et al., 2013; Martinez, Collins, et al., 2011), I believe that relying on external sensors will prove impractical when moving from a controlled lab setting with adults to a high school classroom. External sensors come at extra cost and require additional maintenance, increasing the complexity of the technology and therefore the risk that it will be abandoned rather than adopted (Aldunate & Nussbaum, 2013). The research findings so far do not show this extra care and expense to be worth the additional insight these tools can enable. Additionally, K-12 classrooms can be noisy, particularly during periods of small group work, casting doubt on how effective this approach will be outside of a lab setting.

3.2 DISTINGUISHING AMONG USERS AT A TABLETOP COMPUTER²

When modeling collaboration at a tabletop computer, it is often necessary to be able to track the touches of individual users (Martinez, Collins, et al., 2011). Most tabletop computers, however, are unable to distinguish among users—they can detect and respond to many touches at the same

² Portions of this section reproduce text from Evans, Davis, Fogarty, & Wobbrock. (2017). Group Touch : Distinguishing Tabletop Users in Group Settings via Statistical Modeling of Touch Pairs. *CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 35–47.

time, but they cannot tell the difference between touches carried out by one person and those of another. Although this issue is not directly related to supporting collaborative learning, it represents a significant practical challenge to modeling collaboration at an interactive tabletop.

Common tabletop computers are vision-based, using either frustrated total internal reflection (FTIR) (Han, 2005) or diffuse illumination (DI) (Schöning et al., 2009) to detect touches. Existing approaches to distinguishing among users at vision-based tabletops fall into three categories: (1) approaches that augment the tabletop with additional sensors; (2) approaches that require the user to wear or hold external sensors; and (3) approaches that use only the built-in capabilities of the tabletop hardware. Each of these categories is reviewed in turn.

Approaches that augment the tabletop with additional sensors are typically reliable, accurate, and unobtrusive to the user. These approaches make it possible to track individual users for the duration of their time at the tabletop. For example, proximity sensors have been added to the edges of tabletop computers in order to track users' locations around the screen (Annett, Grossman, Wigdor, & Fitzmaurice, 2011; Tanase, Vatavu, Pentiu, & Graur, 2008). Multiple approaches have used cameras mounted above the tabletop to track users' arms and hands (Clayphan, Martinez-Maldonado, Ackad, & Kay, 2013; Martinez, Collins, et al., 2011; Ramakers, Vanacken, Luyten, Coninx, & Schoning, 2012) or even identify individuals using biometric data (Maekawa, Masuda, & Namioka, 2016). Bootstrapper (Richter, Holz, & Baudisch, 2012) uses a depth camera mounted under the table to track users' locations by their shoes and then match touches to individuals' locations using the touch orientation detected by the tabletop's vision system.

Other approaches augment users with additional sensors, such as rings (Roth, Schmidt, & Benjamin, 2010), wristbands (Meyer & Schmidt, 2010), gloves (Marquardt, Kiemer, Ledo, Boring,

& Greenberg, 2011), or cards (Paul Jermann, Zufferey, & Dillenbourg, 2008) that communicate with tabletops' built-in sensors to identify users. Ackad et al. (2012) used a combination of users' personal mobile devices and an overhead depth camera.

However, relying on external sensors is not always practical or desirable. For example, my work is motivated by the need to distinguish among users to enable modeling of collaborative learning at tabletop computers in authentic classroom settings (Evans, Wobbrock, & Davis, 2016). Classrooms are multi-use spaces often used by multiple groups with different needs and classroom furniture often needs to be moveable and configurable, making it impractical to use external sensors such as a depth camera that would need to be fixed in place. I have encountered this issue in my own research as have other researchers studying tabletops in classroom settings (Kharrufa et al., 2013). Augmenting users (high school students) with wearable sensors was ruled out simply because such sensors require additional setup, storage, and maintenance, all of which increase time spent administering a learning activity and the burden of classroom management on the teacher. These factors go against evidence-based best practices and guidance on integrating new technologies into classrooms (Bielaczyc, 2009; Dillenbourg et al., 2011; Dillenbourg & Jermann, 2010). Therefore, to be suitable for use in classrooms, an approach to distinguishing among users should not rely on external sensors.

Several studies have investigated the use of finger orientation and hand contours, captured by vision-based tabletops' on-board cameras, to match touches to hands and users (Dang, Straub, & André, 2009; Ewerling, Kulik, & Froelich, 2012; Wang, Cao, Ren, & Irani, 2009; Zhang et al., 2012; Zhang, Zhang, & Chen, 2014). Ewerling et al. (2012) developed an image processing technique to match touches to individual hands with four users simultaneously interacting with a tabletop computer. They did not explore the application of their technique to distinguishing users.

Dang et al. (2009) and Zhang et al. (2014) both developed heuristic methods using touch orientation and distances between contact points to map touch points to hands. Mapping touches to hands could potentially improve gesture detection and thereby enhance the interactive experience. Both methods were highly effective at differentiating between the hands of single participants, but neither method was tested with groups or used to extrapolate from hands to users.

Zhang et al.'s (2012) *See Me, See You* algorithm used hand contours captured by the vision system to detect finger orientation and handedness. Using touch orientation along with a touch's coordinates, Zhang et al. trained a support vector machine to predict a user's location at the tabletop, thus distinguishing among different users. Their approach, however, does not support multi-touch input, purposefully constraining users to single-point touches in order to make identification possible.

For touchscreens with capacitive sensing capabilities, Harrison et al.'s (Harrison, Sato, & Poupyrev, 2012) capacitive fingerprinting technique takes advantage of the natural differences in electrical current passing through individual users. This approach does not restrict how users touch the screen, but a limitation of the prototype tested in Harrison et al.'s study is that it can only handle one touch at a time.

Blažica et al. (2013) enable identification of users across all types of tabletop computers using hand biometrics. However, their approach is only able to distinguish among multiple simultaneous users when they place a hand on the screen in a specific position—all five fingers splayed and touching the screen. Therefore, it is not a suitable approach for distinguishing users during “in the wild” tabletop interaction, such as in a high school classroom.

Finally, application-based approaches take advantage of assumed or enforced social protocols. For example, an approach that automatically segments a tabletop interface into

territories belonging to individual users (Garcia-sanjuan, Jaen, & Catala, 2013) draws upon Scott et al.'s (2004) findings that groups tend to divide the tabletop into personal and shared territories without any explicit coordination. However, social protocols can vary greatly depending on the context, limiting their generalizability as a means of distinguishing users. For example, research with adults in a lab setting has shown that participants are reluctant to reach into one another's perceived personal space or territory on the tabletop (Ryall, Forlines, Shen, & Morris, 2004), supporting the notion that it is possible to distinguish users by identifying those personal territories. In contrast, children engaged in collaborative learning around a tabletop often break or ignore boundaries of personal space, which can have the effect of actually enhancing collaboration (Fleck et al., 2009; Pontual Falcão & Price, 2010). Given that our context is tabletop collaboration in authentic high school classrooms, these findings suggest that assuming young users will follow adult social protocols may be unwise.

All of the approaches described above that use only the built-in capabilities of the hardware to distinguish users also impose restrictions on how users can interact with the tabletop. Some researchers argue that there are scenarios where users will be willing to sacrifice unconstrained interaction for the ability to track users (Zhang et al., 2012). However, this will only hold true when the value added by tracking users outweighs the inconvenience of artificial constraints on interaction. When this is not the case, such as when an application requires true multi-touch input, the above approaches are not suitable. Additionally, these approaches are at risk of failure with young users because it can be difficult to get children and adolescents to comply with behavioral constraints. With Group Touch, described in 0, I address the problem of distinguishing users without utilizing external sensors and without constraining interaction.

3.3 SUPPORTING COLLABORATION AND COLLABORATIVE LEARNING

As previously stated, one of the goals of my dissertation research is to build software adaptations that provide just-in-time support for collaborative learning. The ability to distinguish among users at a tabletop computer opens the door to being able to detect collaboration behavior, but how the computer should adapt to support effective collaboration is an open question. Although prior work on supporting collaborative learning at tabletop computers is limited, there is extensive prior work on encouraging effective collaboration in other types of CSCL environments.

Järvelä and Hadwin (2013) identify three categories of support for socially-shared regulation in CSCL: structuring supports, mirroring and metacognitive awareness tools, and guiding tools. *Structuring supports* include typically fixed, non-adaptive supports, such as assigning roles to students (e.g., Kirschner, Strijbos, Kreijns, & Beers, 2004; O'Donnell, Hmelo-Silver, & Erkens, 2006; Palincsar & Brown, 1984) or asking groups to follow collaboration scripts (Clayphan, Kay, & Weinberger, 2014; Dillenbourg & Jermann, 2007; Fischer, Kollar, Mandl, & Haake, 2007). These types of support help students manage their collaboration by providing additional structure to the learning context.

An additional form of structuring support is the use of prompts—messaging or questions—that guide students through their collaboration. Prompts can be either fixed or adaptive. Harney et al. (2015) investigated the effects of task- or process-level prompts on collaboration. Task-level prompts were tied to specific actions in the software (e.g. “Can you elaborate on this idea?”) whereas process-level prompts encouraged groups to reflect on the strategies they use to manage the collaboration itself (e.g. “Can some similar ideas be merged?”). They found that groups that received process-level as well as task-level prompts not only used more sophisticated argumentation during the collaboration, but also perceived greater consensus in the group than

groups that received only task-level prompts. In Harney et al.'s study, prompts were delivered by a human facilitator but could be embedded in CSCL systems. Other forms of prompts include sentence starters in online chat and discussion tools or questions designed to encourage students to reflect on their learning, which have been shown to help groups to plan their task work and reach consensus (Hadwin, Oshige, Gress, & Winne, 2010; Järvelä & Hadwin, 2013). These findings demonstrate that prompts that focus on the process of collaborating can provide effective supports.

The second category of support for socially-shared regulation in CSCL is *mirroring and metacognitive awareness tools* that provide feedback to students, often in the form of visualizations, on how they are interacting with each other and the assigned task. The aim of these types of tools is to encourage awareness and reflection (Järvelä & Hadwin, 2013; Malmberg et al., 2015). Mirroring and metacognitive awareness tools can make individual group members aware of the nature and quantity of their contributions to the group effort in order to encourage more equitable participation and effective group regulation. For example, Lonchamp (2010) developed a tool to provide group members with real-time monitoring information derived from their actions in the software and their text-based communications showing each student how their participation compared to other group members and providing brief guidance on how to improve. Overall, students found personalized advice to be the most useful aspect of the system, while seeing their ranking against other students could motivate them to participate more. Some students, however, deliberately posted nonsensical messages in what appeared to be an attempt to push back against being so closely monitored by the system.

Janssen et al. (2011) found that the mere presence of group awareness tools did not necessarily improve collaboration but that groups who used such tools for longer did see improved

participation, coordination, and regulation of their group activities. Phielix et al. (2011) created tools that enabled group members to see real-time anonymous feedback from their peers about their cognitive and social participation in the group effort in comparison to the rest of the group. These tools helped individuals to become more aware and groups to develop shared perceptions of the social aspects of their collaboration.

Other examples of group awareness tools include: visualizations that show who is contributing to a conversation, the nature or type of their contributions (e.g., argument or response), and how their contributions connect to others (Trausan-Matu et al., 2014); a tool that shows individual students, based on self-report, how the emotions and challenges they experience while collaborating compare to those of other group members (Järvelä & Hadwin, 2013); a tool that visualizes group members' confidence in their ability to carry out a learning activity based on each member's own ratings of statements such as "I understand how to do this task" (Järvelä et al., 2014); and a widget that shows individual students periodic visualizations of their activity in a CSCL environment compared to that of other students (Scheffel, Drachsler, Kreijns, de Kraker, & Specht, 2017).

The group awareness tools described above make use of the closed nature of text-based collaborative environments, in which all group interactions can be captured and mined for patterns associated with the quality of collaboration. In face-to-face group work at a tabletop, however, the computer can only access direct interaction—the verbal and gestural interactions that learners have with each other cannot be captured without the use of external sensors. Martinez-Maldonado et al. used external sensors to capture learners' verbal and physical interactions during tabletop collaboration in order to visualize each group member's level of participation (Martinez-Maldonado et al., 2013; Martinez, Collins, et al., 2011; Martinez, Kay, et al., 2011). Although

these visualizations were aimed at teachers rather than the groups themselves, they could be presented to students as group awareness tools. Reliance on external sensors, however, makes this approach challenging to implement in a typical classroom setting.

Guiding systems make up the third category of computer-based support for collaborative learning described by Järvelä & Hadwin (2013). Guiding systems aim to actively coach students through a collaboration using adaptive feedback. For example, where group awareness tools reflect data about collaboration back to students, guiding systems would prompt students to take action based on that data. Computerized agents, a common feature of intelligent tutoring systems, are a widely used type of guiding tool. Agents interact with students to provide adaptive feedback on their actions in the learning software and guide them through an activity. Many computerized agents are content-focused or designed for single-learner environments but there are examples of agents that scaffold the process of learning in CSCL environments (Kumar & Rosé, 2011; Kumar, Rosé, Wang, Joshi, & Robinson, 2007), such as an agent that scaffolds collaborative brainstorming (Wang, Rosé, & Chang, 2011), or detects when students get stuck on a problem and helps them to move forward (Hadwin, Oshige, Gress, & Winne, 2010). These guiding systems have proved successful at scaffolding collaboration, but Järvelä & Hadwin (2013) caution that the level of support must be gradually reduced in order to encourage students to adopt the regulatory strategies that the guiding systems employ.

The three categories of support identified by Järvelä & Hadwin (2013) are not mutually exclusive. For example, some of the mirroring and meta-cognitive awareness tools described above make use of data gathered by prompts (Järvelä et al., 2014; Järvelä & Hadwin, 2013). Additionally, guiding systems may make use of supports from the other two categories, such as

coaching socially-shared regulation by prompting students to reflect on a mirroring tool, or by walking them through a collaboration script.

There is an implicit assumption in much of the research on supporting collaborative learning that tailored, adaptive support will be more effective than fixed, non-adaptive support (Patrick Jermann et al., 2002; Lonchamp, 2010; Marcos-Garcia et al., 2007; R. Martinez, Collins, et al., 2011). Walker et al. (Walker, Rummel, & Koedinger, 2011) tested this assumption in the context of a computer-mediated peer math tutoring system by comparing the effectiveness of genuinely adaptive support and fixed support that students were led to believe was adapting to their behavior. Students used the system in pairs, with one student assuming the role of tutor and the other the role of tutee as the tutee worked on algebra equations on his or her screen. Only the tutor received support from the system, the goal of which was to help them give reflective and thoughtful feedback to the tutee instead of the less effective direct hints that students typically give without training. Walker et al. (2011) found that students who received genuinely adaptive support felt that they were better tutors and also scored higher on tests of content learning than students who received fixed support. This finding supports the view that providing real-time, adaptive support where possible is likely to lead to better outcomes than fixed support alone.

To date, little research has been carried out on providing real-time adaptive support for collaborative learning at tabletop computers. The teacher dashboard by Martinez-Maldonado et al. (Martinez-Maldonado, Kay, Yacef, Edbauer, & Dimitriadis, 2013; Martinez-Maldonado, Kay, Yacef, & Schwendimann, 2012) described above remains the only known example. Text-based online collaborative environments, which have been the traditional focus of CSCL research, have received much greater attention in the literature on adaptive support for collaborative learning. Although many of the principles of supporting collaborative learning developed through research

conducted in text-based online environments will apply to tabletop computers, the tabletop computer's shared interface adds an additional challenge for collaboration that is not a factor in text-based environments. Namely, that the tabletop interface is prone to conflict among collaborators because there are no clearly defined boundaries between personal and shared spaces and multiple users are able to carry out opposing actions simultaneously. Even when users are not intentionally trying to disrupt each other, conflicts are likely to occur (Klinkhammer, Nitsche, Specht, & Reiterer, 2011; Morris, Ryall, Shen, Forlines, & Vernier, 2004). Examples of conflict include a user "stealing" an on-screen object that another user is working with, or closing an application window before others are ready to move on.

A number of tabletop interface design components have been proposed to reduce conflict, intentional or otherwise. For example, Morris, Huang, Paepcke, and Winograd (2006) found that cooperative gestures, which require input from multiple users in order to bring about a single action, could increase collaboration and help individuals maintain control over their personal space in the context of a drawing application. Morris et al. (2006) also found some of the cooperative gestures included in their application actually increased conflict when they were used in ways that were not originally intended by the researchers, suggesting that cooperative gestures need to be carefully designed and may not be suitable in every situation. Piper, O'Brien, Morris, and Winograd (2006) also implemented variants of cooperative gestures in the form of computer-enforced turn-taking and voting buttons, which required users to come to a consensus before enabling particular actions to be carried out.

Klinkhammer et al. (2011) developed an application to support territoriality in an interactive tabletop exhibit in a museum setting. Each user had access to a visually demarcated personal space on the tabletop, in which only the owner could interact with objects and make

changes, as well as a shared group space that was accessible to all. Objects could be dragged to and from the shared and personal spaces. Although Klinkhammer et al.'s approach avoided conflict between users, it was designed for parallel use of a shared interface and was not tested in highly collaborative use cases.

Morris et al. (2004) proposed a framework of coordination policies—mechanisms for controlling or shaping interaction with tabletop software that can help to reduce or resolve conflict among users. The coordination policies are based on two factors: 1) the conflict type—affecting the entire application, a whole element, or part of an element; and 2) the initiative—whose actions are used to determine the outcome: the user who owns the element in conflict or who initiated a global change versus the users affected by a conflicting action, or a mix of all users. Morris et al. (2004) provide illustrative examples of each policy. An example of a whole element, mixed initiative conflict resolution would be a tearing effect that occurs when multiple users try to drag an element in different directions. In contrast, an initiative favoring the owner of the element would only apply the touches carried out by him or her. While Morris et al. present these policies as always-on, built-in controls, it seems likely that some could also be made adaptive and only triggered as the need arises. This could be beneficial for policies that would be cumbersome or artificially restrictive if used at all times—voting, for example, could become tedious if required for every potential global action. Instead, these coordination policies could be employed only if group members frequently come into conflict.

Much research has investigated what collaboration looks like at a tabletop computer and how static interface design can help to make groups more effective. The area of adaptive support for tabletop collaboration, however, is still wide open. My dissertation research begins to address this issue.

Chapter 4. MODELLING COLLABORATIVE LEARNING AT A TABLETOP COMPUTER³

The work described in this chapter addresses RQ 1: What can learners' physical interactions with a tabletop computer reveal about the group's social regulation processes? I conducted two studies: (1) an exploratory study in a lab setting, which sought to identify patterns in how groups interacted with the computer that could indicate the quality of their social regulation, and (2) a follow-up study in an authentic classroom setting that aimed to validate and refine the patterns found in the lab study.

4.1 STUDY 1: AN EXPLORATORY STUDY OF TOUCH PATTERNS ASSOCIATED WITH QUALITY OF SOCIAL REGULATION

Eleven adults (7 female, 4 male) worked in small groups (two dyads, a group of three, and a group of four) to analyze and compare two poems, *Birches* by Robert Frost (1969), and *Fern Hill* by Dylan Thomas (2003). Participants were recruited via campus mailing lists and were grouped based on their availability for study sessions. Ten of the eleven had an undergraduate degree and seven had also completed graduate level coursework. Most participants did not know their group members prior to the study although a handful were acquaintances.

I chose poetry analysis for this study because many students find it challenging, and yet it requires little background knowledge to get started, making it an authentic learning activity for most participants. Poetry analysis is interpretative, which also makes it an ideal activity for small

³ Portions of this section reproduce text from Evans & Wobbrock. (2014). Filling in the gaps: capturing social regulation in an interactive tabletop learning environment. In *Proceedings of the International Conference of the Learning Sciences (ICLS '14)* (pp. 1157–1161). Boulder, Colorado: ISLS.

group work as students can reach deeper levels of understanding by externalizing their own thoughts and building on the ideas of others. The participants were given 30 minutes to work on the following task adapted from McMahan, Funk, & Day (1998):

Compare and contrast two poems, *Fern Hill* by Dylan Thomas and *Birches* by Robert Frost. Answer the following questions and support your answers with evidence from the text. As you answer each question, consider how the two poems are similar and how they are different.

- 1) What is the **theme** (the central idea) of each poem?
- 2) Who is the **speaker** in each poem? How would you describe the speaker?
- 3) What **imagery** does each poet use? How do the images relate to each other and contribute to the theme?



Figure 4-1. A screenshot of the software used in Study 1.

The participants used a Microsoft PixelSense, which has a 40” multi-touch screen and can comfortably seat four adults around it. The software (see Figure 4-1), created specifically for this study, was intended to be used as a tool to support discussion rather than for creating a product to be turned in. At the end of the activity, the participants were asked to verbally summarize their

findings to a researcher as if they were presenting to the rest of a class after a period of small-group work. This structure meant that the participants were relatively unconstrained in how they tackled the task and were free to make use of the software in the ways they found most useful.

The text of both poems was presented on-screen and supplemented by audio recordings of readings by the poets. Participants could take notes by clicking buttons to add “notecards” to the screen. Notecards were color-coded according to each element of the poems they were asked to discuss (e.g., red for theme, green for speaker, and yellow for imagery) and participants could add as many notecards as they wanted. Each notecard contained a text field in which participants could type an observation about the text, and an area for collecting supporting evidence by dragging and dropping lines from the poems onto the notecard. Once dragged onto a notecard, individual words could also be highlighted. Participants could annotate the poems directly by drawing on them with their fingers. The software also included a set of optional prompt questions to stimulate discussion. Each tool described so far could be moved around the screen freely. In addition, fixed pieces of virtual “scratch paper” were provided at each corner to enable each participant to take their own notes.

Two video cameras recorded the discussions. Interactions with the tabletop were recorded in log files detailing when and where a touch happened, and what action was performed. The researcher left the room while the participants worked to encourage them to give an authentic summary at the end of the activity.

4.1.1 *Data Analysis*

The videos were thematically coded for quality of social regulation processes using the codes and sub codes described in detail by Kempler Rogat and Linnenbrink-Garcia (2011) and listed in Table 2-2. The coding process allowed the videos to be broken down into episodes. In Chi’s terms, an

episode is “an event, or a specific activity,” 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 (Chi, 1997, p284). Episodes could have multiple codes; for example, when some group members 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.

In order to draw useful insights from the log files of touch data, individual touches needed to be grouped into sequences that represented purposeful actions—actions that served a particular purpose. For example, imagine a participant wants to read one of the poems, which is currently located closer to another participant on the other side of the screen. He makes one touch to move the poem closer to himself, followed by a second touch to rotate it so it is oriented properly from his perspective. In this case, the two touches together represent a single purposeful action. Three features are needed to make that determination: (1) who carried out the touch (the owner); (2) the object that was touched; and (3) the timing of the touches. Items 2 and 3 are readily available in the logs but, as the PixelSense cannot natively differentiate individual users, for this study, the owner was manually labeled by matching the timestamp of the touch to a frame from the video.

For the first group, touches were grouped into meaningful actions by close inspection of the log file alongside the video. The timing between touches was studied to determine the distinction between sequences of touches that represented purposeful actions, and sequences that should be considered separate actions. Clear rules emerged for sequences of touches owned by a single person in terms of object type and length of time between touches (see Table 4-1).

Table 4-1. Rules for grouping touch event data into purposeful actions by time between touches.

| <i>Objects touched are:</i> | Purposeful action sequence | Separate actions |
|-----------------------------|-----------------------------------|-------------------------|
| <i>Related</i> | ≤ 50 seconds | > 50 seconds |
| <i>Unrelated</i> | ≤ 15 seconds | > 15 seconds |

The object types *related* and *unrelated* refer to the objects' function in the context of the activity. For example, the time between a touch on the Thomas poem followed by a touch on a line from the Thomas poem that has been dragged to a notecard would be categorized as *related*, while a touch on the Frost poem followed by a touch on the instructions would be *unrelated*. Objects could be related either by poem, e.g., a notecard containing lines from Thomas is related to the Thomas poem; or by task sub-question, e.g., a theme notecard containing lines from Frost is related to the Thomas poem in “Annotate Themes” mode.

For groups 2, 3, and 4, logged touches were automatically grouped according to the rules given in Table 3. The rules were then validated by repeating the video inspection for these groups and, for each category, calculating the percentage of touches that were correctly categorized. The rules proved to be highly accurate, ranging from 87.2% to 100.0% correct.

The screen of the PixelSense is highly sensitive and as a result, the touch logs contained a lot of noise, often due to participants catching on-screen objects with their elbows while reaching for another object. Another source of noise came from multi-touch input—using more than one finger to carry out a single action, such as rotating an object using multiple fingers. Simple heuristics were used to filter out these types of noise: (1) touches overlapping (in time) by the same person on the same object were combined, (2) when there were overlapping touches by the same person on different objects and one touch was close to the edge at which the person was sitting, the edge touch was removed, and (3) when there were overlapping touches on multiple objects located directly on top of each other, the top touch was kept and all others were removed.

The third stage of analysis brought together verbal and physical interactions. The goal of this stage was to find patterns of touches that reflected social regulation processes used by each group, as well as their quality. Therefore, we identified the processes for which there were clear differences among the groups in terms of quality. This substantially narrowed the focus of the subsequent analysis as the groups had more similarities than differences in the quality of their collaboration, an unexpected outcome of this study. Timestamps from the video that delineated episodes where a group was engaging in a particular regulatory process were applied to the pre-processed touch data. The episodes were then compared to see whether patterns emerged. Only touches to the shared objects (i.e., everything but individuals' separate "scratch paper") were considered at this stage of the analysis.

4.1.2 *Results: Forms and Quality of Social Regulation*

All groups spent around ten minutes of the activity silently reading the poems and listening to the recordings before starting discussion. The nature of the task demanded relatively little planning, and all groups were coded as high quality in this area.

After the initial period of preparation, however, each group took a different approach in their discussions and their use of the software. Group 1 (three participants) structured their discussion around the task's three sub-questions, comparing the two poems as they considered the themes, speakers, and imagery in order. They took notes using the notecards and copied lines from each poem to support their observations. Group 1 made the most effective use of the allotted time, covering each aspect of the task in depth. Their monitoring processes—content monitoring, monitoring the plan, and monitoring progress—were consistently high quality across all three sub-codes.

Group 2 (four participants) did a great deal of comparing and contrasting of details in the poems. They were the only group to make use of the optional discussion prompts. They annotated the poems directly and wrote notes on their individual virtual notepaper. Group 2 did not follow a clear structure. They covered each sub-question, but not in a linear fashion. They exhibited consistently high-quality monitoring of the plan and progress. Although their content monitoring was mostly high quality, there were some low-quality instances, largely due to one group member's occasionally dismissive and unresponsive treatment of others' contributions.

After reading both poems, Group 3 (a dyad) decided to tackle one poem at a time. They started with the Thomas poem and covered each task sub-question in turn with just that poem, taking notes using the notecard feature for the theme and speaker before switching to directly annotating the imagery in the poem. They did not get to the Frost poem until near the end of the activity, eventually running out of time having only considered its theme and not its speaker or imagery. As a result, the group was not able to make many comparisons between the two poems, although their analysis of Fern Hill was exceptionally thorough. In terms of social regulation, their discussion was focused on content monitoring, all of which was high quality. In contrast, progress monitoring was largely missing and monitoring of the plan was very limited, though generally high quality.

Group 4 (a dyad) were more focused on organizing the virtual workspace and collecting lines than on actually engaging with the poems. Their discussion of both poems, though valid, was mostly superficial and more of a summary than an analysis. Most instances of content monitoring were of low quality, although they did begin to improve near the end of the discussion. They were easily distracted by superficial concerns, which meant that thoughtful contributions were often unheeded—an indicator of low-quality progress monitoring. They did, however, also engage in

high-quality progress monitoring, regularly tracking what remained to be done. Their plan monitoring was generally moderate as, although they frequently sought to clarify the instructions and evaluate their plan, their continual monitoring of the plan hindered their ability to enact it.

4.1.3 *Results: Social Regulation in the Touch Data*

As described above, the groups differed in quality along the monitoring dimension of social regulation, so the three monitoring sub-codes were used to identify patterns of touch data that reflected quality of monitoring processes. Results were normalized by group size to facilitate comparison.

Monitoring Content. Monitoring content was the most-used process of social regulation, with noticeably longer duration than the other processes. Quality variations in content monitoring revealed three distinctive touch patterns:

1. ***Unrelated Touches.*** Low-quality episodes were characterized by sequences of purposeful touches to unrelated objects. All of these episodes occurred in Group 4's session. Group 4, which was consistently low quality, averaged 23.5 such sequences, more than double that of the high-quality groups. Group 1 averaged just 2.3, while group 2 averaged 10.5, and group 3 averaged 11.0.
2. ***Touch Time.*** During low-quality episodes, sequences of touches were also much temporally denser—many touches in quick succession. Overall, group 4 spent more than twice as much time, 476 seconds per person, touching the screen than any other group. The high-quality groups were similar to each other: group 1 = 199s, group 2 = 136s, and group 3 = 190s.
3. ***Overlapping Sequences.*** Finally, high-quality episodes were characterized by only one person interacting with the screen at a time, moving from one person to another

with little overlap, suggesting turn-taking. In groups 1, 2, and 3, overlapping touches by two or more people made up 10.5%, 10.7%, and 9.6% of their total touches to shared objects, respectively. Conversely, low-quality episodes often featured overlapping touches. For group 4, overlapping touches made up 23.6% of their total touches to shared objects.

Monitoring the Plan. The log data contained indicators that plan monitoring was occurring. Participants sometimes touched the instructions on-screen when revisiting the task and re-evaluating their plan and otherwise did not interact with it. However, the variations in quality among the groups did not present distinctive patterns.

Monitoring Progress. Episodes of progress monitoring did not show any distinctive patterns in the touch data, regardless of quality. Progress monitoring episodes were generally brief, predominantly verbal interactions that did not generate much touch data. While deictic gestures did appear in the video, they were generally above the screen and so were not captured by the log data.

Despite the finding that at least some aspects of social regulation were reflected in groups' interaction patterns, it was clear, however, that further investigation was needed. Whereas the adult participants in the lab study remained on-task and engaged at all times, studies of younger students using interactive tabletops in the field have shown them to be not quite so consistently well-behaved or motivated to work on the assigned task (Do-Lenh, 2012; Kharrufa et al., 2013). Additionally, all but one of the groups of adults engaged in primarily high-quality, socially-shared regulation, limiting the available data on ineffective regulation and therefore my ability to identify differences in the interaction patterns of high- and low- performing groups. As almost all of the episodes of low-quality social regulation came from a single group, there was also a concern that

the patterns identified could be a reflection of the particular style of that group rather than low-quality social regulation itself. Given the stark differences between the behavior of the adults in a lab setting and that of students in the field, as well as the limited occurrences of low-quality social regulation, I believed it to be important to test this approach to detecting collaborative processes in an authentic setting.

4.2 STUDY 2: TOUCH PATTERNS ASSOCIATED WITH QUALITY OF SOCIAL REGULATION IN A CLASSROOM SETTING⁴

In this follow-up study, I sought to build upon the previous lab study by evaluating the interaction patterns associated with quality of content monitoring in an authentic classroom environment with high school students working in a different learning domain. By studying students' social regulation at a tabletop computer in a classroom, I 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.

Sixteen 9th to 11th grade students (10 female, 6 male) from three area high schools participated in this study. The students were enrolled in a six-week user-centered design seminar offered as part of a college preparation program serving low income students who would be the first generation in their family to earn a four-year college degree. The students in the program attended classes full time for six weeks during the summer. All students in the program had to take several required courses depending on their grade but they could also choose from several

⁴ Portions of this section reproduce text from Evans, Wobbrock, & Davis. (2016). Modelling collaboration patterns on an interactive tabletop in a classroom setting. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '16)* (pp. 860–871). New York.

electives. The course used in this study was offered as an elective and 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 based on their shared interest in a website topic. Each group was made up of a mix of students who did not know each other and some who were already acquainted, either because they were returning to the program for a second summer or because they attended the same school.

As in the lab study, the participants used a Microsoft PixelSense. 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.

In a major engineering effort, I custom-built four distinct applications (see Figure 4-2) 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 4-2. The applications created for this work. Top left: used to conduct design critiques of early wireframes. Top right: used to brainstorm questions for usability testing. Bottom left: helped students to improve their online search skills. Bottom right: focused on Nielsen’s (1994) usability heuristics.

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 (Figure 4-2, top left) supported design critiques and was used in the early stage of the group project to finalize their website layouts. A second application (Figure 4-2, 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 (Figure 4-2, bottom left) developed groups’ search skills to help them find and evaluate online technical resources. The final application (Figure 4-2, 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 (Nielsen, 1994).

Each group was video recorded during each activity and the computer logged every touch. In all, 4,253 touches were logged over 13 group sessions.

In order to detect the touch patterns identified in the lab study, 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 jointly 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 (see Appendix A) 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 determined using the same criteria as in the previous study: each object's function in the context of the activity. All relationships were symmetrical and unordered.

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. *Unrelated Touches*, one of the three touch patterns identified in the previous study, relies on the on-screen relationships. Therefore, it was not possible to apply that pattern to the touch data for the browser-based application.

4.2.1 *Video Analysis*

As in Study 1, the videos were thematically coded for social regulation using Kempler Rogat and Linnenbrink-Garcia's (2011) framework. All stages of video coding were carried out multiple times and validated with peer debriefing (Creswell, 2003) until it was determined that codes had been applied to episodes accurately and consistently according to Kempler Rogat and Linnenbrink-Garcia's (2011) descriptions.

First, the videos were viewed in their entirety to gain a general sense of each group's collaboration. The videos were then transcribed, split into episodes (Chi, 1997) and coded for social regulation, where it was occurring. 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 4-2) through collaborative coding (Smagorinsky, 2008) 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 Kempler

Rogat and Linnenbrink-Garcia's descriptions (Kemppler Rogat & Linnenbrink-Garcia, 2011) and the video, the codes applied to the transcripts were adjusted as the subtleties of each code were teased out.

Table 4-2. Codes used in the video analysis. Social regulation codes adapted from (Rogat & Linnenbrink-Garcia, 2011).

| Social regulation processes | Impact on collaboration |
|---|--|
| Planning: <ul style="list-style-type: none"> • Task • Content Monitoring: <ul style="list-style-type: none"> • Content • Plan • Progress Behavioral engagement | All processes were coded as high or low quality High-quality processes considered to have positive impact Low-quality processes considered to have negative impact |
| Additional codes | |
| <ul style="list-style-type: none"> • Non-collaborative interactions • Off-task interactions • Task work • Software conflicts | <ul style="list-style-type: none"> • Negative • Negative • Positive • Negative |

All dimensions of social regulation were included in the video coding for this study even though the lab study 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 (Kemppler Rogat & Linnenbrink-Garcia, 2011). 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 4-2 also shows whether each code was considered to have a positive or negative impact on collaboration. *Non-collaborative interactions* was a code used by Kempler Rogat and Linnenbrink-Garcia in their work on social regulation (Kempler Rogat & Linnenbrink-Garcia, 2011), 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 (2010) 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 interactions, in which groups voluntarily or spontaneously stopped working on the assigned task.

4.2.2 Touch Data Analysis

Touch data analysis began with pre-processing using an application (Figure 4-3) I custom built for this purpose. The first step was to use the video to manually label each touch with its author—the group member who performed the touch. 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 the lab study and shown in Table 4-1. 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 4-1 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.

The screenshot shows the 'MainWindow' application with the following components:

- Sync controls:** Buttons for 'Find Surface log', 'Find Chrome log', 'Find Video', 'Play', 'Pause', '< 1 s', '1 s >', '365604', 'Go', 'Get Elements', 'Vid start: 13:15:27:434', 'Sync', 'First: 16786409', 'Last: 16788586', 'Save XML', 'Experiment', '<<', '>>'.
- Session video:** A video player showing a classroom scene with students at a table.
- Touch visualization:** A large empty white area with a red arrow pointing to it.
- Pattern evaluation controls:** 'Clear', 'Purposeful unrelated', 'Time period (s): 60', 'Save Transcript', 'Overlap', 'Touch Time', 'Time period (s): 60'.
- Labeling controls:** 'By ID', 'By Gaps', and a table with columns for ID, Start Time, o mean, x, y, Side, Student, Element, and Transcript.
- Touch data:** A table listing touch events with columns: ID, Start Time, o mean, x, y, Side, Student, Element, and Transcript.

| ID | Start Time | o mean | x | y | Side | Student | Element | Transcript |
|-----|------------|--------------|----------|------|------|---------|---------|------------|
| 293 | 16787882 | 13:21:14:482 | 1.00302 | 512 | 315 | T-L | ? | |
| 294 | 16787892 | 13:21:22:318 | 3.05752 | 977 | 31 | T-R | ? | |
| 295 | 16787893 | 13:21:24:553 | 2.13111 | 693 | 600 | T-R | ? | |
| 296 | 16787894 | 13:21:25:307 | 2.25357 | 639 | 608 | T-R | ? | |
| 297 | 16787895 | 13:21:27:687 | 2.13963 | 635 | 591 | T-R | ? | |
| 298 | 16787896 | 13:21:29:856 | 2.15057 | 629 | 586 | T-R | ? | |
| 299 | 16787897 | 13:21:31:877 | -2.19905 | 127 | 991 | B-L | ? | |
| 300 | 16787898 | 13:21:35:761 | -1.81931 | 977 | 10 | T-R | ? | |
| 301 | 16787899 | 13:21:36:896 | 2.52769 | 1494 | 313 | T-R | ? | |

Figure 4-3. The data analysis tool I built for this work. This tool includes controls for syncing session video recordings to touch logs. Each row in the table represents a single touch. Clicking on a row in the table skips the video to the corresponding timecode and draws a trace of the touch in the visualization area. The table also includes controls for labeling touches by owner. The pattern evaluation controls were used to test the touch patterns uncovered in this work.

I 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, I inspected the processed touch logs to determine if the touch patterns originally identified in the lab study continued to serve as indicators of the quality of content monitoring in the context of this study. Table 4-3 describes the touch patterns as they were applied to the data from this study. The quality thresholds in Table 4-3 were calculated from the results of the previous study. Episodes of touch data that fell between the thresholds were labeled “medium-quality.”

Table 4-3. Touch patterns associated with quality of content monitoring. Quality thresholds are derived from the results of the lab study.

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

Based on the results of this analysis, I 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, 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 two key characteristics: (1) the group maintains shared focus on the learning task, and (2) 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, I decided to check the patterns against all video codes 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. It is also possible that additional touch patterns exist reflecting other characteristics of social regulation. However, further exploration of this sort was left to future work to allow for focus on the goal of the current study—to validate the previously discovered touch patterns.

This decision—to check if the known touch patterns applied beyond just content monitoring—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, and 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, I checked for the patterns given in Table 4-3 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 two 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 two-minute intervals contained at least one complete episode. To reduce the potential for incorrect classifications caused by the artificial interval length, two-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 4-3. 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).

4.2.3 *Results*

The analysis showed that, compared to the previous study, the high school students in a classroom setting used a greater range of social regulation processes with more variation in quality, enabling me to investigate the relationship between the touch patterns and quality of collaboration more broadly. I 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 and quality of collaboration. Unlike the adults in the lab setting, 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 4-4. The results are given as percentages to facilitate inter-group comparison as the number of episodes of social regulation in each group differed.

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

| Social regulation process | Group 1 | | Group 2 | | Group 3 | | Group 4 | |
|---------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | <i>HQ</i> | <i>LQ</i> | <i>HQ</i> | <i>LQ</i> | <i>HQ</i> | <i>LQ</i> | <i>HQ</i> | <i>LQ</i> |
| Planning - task | 0 | 36.8 | 0 | 2.9 | 0 | 0 | 0 | 6.7 |
| Planning - content | 0 | 5.3 | 14.7 | 5.9 | 7.7 | 15.4 | 10.0 | 10.0 |
| Monitoring - content | 31.6 | 5.3 | 32.4 | 17.6 | 23.1 | 30.8 | 16.7 | 33.3 |
| Monitoring - plan | 0 | 10.5 | 2.9 | 5.9 | 0 | 0 | 0 | 13.3 |
| Monitoring - progress | 0 | 5.3 | 2.9 | 2.9 | 7.7 | 0 | 0 | 0 |
| Behavioral engagement | 0 | 5.3 | 8.8 | 2.9 | 0 | 15.4 | 10.0 | 0 |
| Total | 31.6 | 68.4 | 61.8 | 38.2 | 38.5 | 61.5 | 36.7 | 63.3 |

Table 4-4 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.

The high school students also differed from the adults in the lab in terms of the behaviors described by the additional codes given in Table 4-2: 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 4-5 shows the average number of episodes of each additional code per activity for each group.

Table 4-5. Mean number of episodes of each additional code per activity by group.

| Code | Mean number of episodes | | | | |
|--------------------------------|-------------------------|---------|---------|---------|------------|
| | Group 1 | Group 2 | Group 3 | Group 4 | All Groups |
| Non-collaborative interactions | 1.7 | 1.3 | 1.3 | 0.3 | 1.2 |
| Off-task interactions | 2.0 | 2.0 | 3.0 | 0.7 | 1.9 |
| Task work | 0 | 2.0 | 0.7 | 1.3 | 1.1 |
| Software conflicts | 1.0 | 1.8 | 2.0 | 1.3 | 1.5 |

Table 4-6 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 4-6. 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 the lab study and the revised pattern found in this study.

| Pattern | High Quality | | Low Quality | | Overall |
|---------------------------------|--------------|-----------|-------------|-----------|-----------|
| | # | % correct | # | % correct | % correct |
| Unrelated Touches | 16 | 100.0 | 10 | 80.0 | 92.3 |
| Touch Time | 25 | 76.0 | 18 | 0 | 44.2 |
| Overlapping Sequences | 25 | 16.3 | 18 | 72.2 | 46.5 |
| Overlapping Unrelated Sequences | 16 | 62.5 | 10 | 60 | 61.5 |

As described above, 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, 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, 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, I 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 4-6. Overlapping Unrelated Sequences replaced Overlapping Sequences in subsequent analyses. 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 and quality of collaboration. The video analysis showed a much greater range of collaboration processes than reported in the lab study with adults. Therefore, all episodes were included in order to extend the analysis beyond content monitoring and investigate the relationship between the touch patterns and quality of collaboration more broadly. The results of this investigation are shown in Table 4-7.

Table 4-7. 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.

| (a) Pattern quality label | | (b) # episodes | (c) Most common episode code(s) | (d) Distribution of positive / negative codes (episodes) | (e) Intervals | | |
|------------------------------|---------------------------------------|-------------------|--|---|------------------|-------------|-------------------|
| Unrelated Touches | Overlapping Unrelated Sequences | | | | Positive codes | Mixed codes | Negative codes |
| High | High | 28 | Content monitoring – HQ (10) | 75.0% positive | 8 | 3 | 2 |
| High | Med. | 0 | - | - | - | 1 | - |
| High | Low | 17 | Task planning – LQ (3), Software conflict (3), Content monitoring – HQ (3) | 64.7% negative | 1 | 5 | 8 |
| Med. | High | 1 | Task work (1) | 100% positive | 1 | - | 2 |
| Med. | Med. | 0 | - | - | - | - | 1 |
| Med. | Low | 7 | Non-collaborative (2), Off task (2) | 100% negative | - | - | 3 |
| Low | High | 8 | Content monitoring – LQ (4) | 87.7% negative | - | - | - |
| Low | Med. | 0 | - | - | - | 1 | - |
| Low | Low | 19 | Software conflict (5), Content planning – LQ (5) | 84.2% negative | 1 | 3 | 24 |

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 4-2) for each possible quality combination of the two patterns (listed in column a, Table 4-7). 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 4-7.

Column d in Table 4-7 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 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 4-7 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% to 100.0% of cases), I am 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 4-7 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.

4.3 DISCUSSION

The quality of social regulation and the other behaviors observed in the classroom study differed greatly from the lab study 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 (Järvelä & Hadwin, 2013; Rogat & Linnenbrink-Garcia, 2011). Given that the same Learning Sciences literature has also demonstrated that ineffective or poor social regulation negatively impacts learning outcomes, 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 the lab study, it was not possible to state with any certainty that the patterns established as indicators of quality of collaborative processes (Table 4-3) in a lab study with adults would be valid outside the lab, with younger users, or with different software. The results from the classroom 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. I 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 in the classroom study 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 the classroom study spent less time touching the screen than the adults in the lab study.

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. Student 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 video analysis in the classroom study 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) 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; or (3) intentional disruption—the intentional action of one student, e.g., flicking an on-screen element across the screen to derail other group members' work. It is possible that the noted hyper-sensitivity of the PixelSense's screen, which meant that it often responded to motion just above the screen as well as actual touches, exacerbated the conflicts and that this problem would be less of an issue with improved hardware.

Both the lab and the classroom study addressed RQ 1 and demonstrated that it is possible to detect the quality of collaboration using only touch data. However, it is important to note that because only touch data is used, my approach 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 chapter will need to take into account the fact that the modeling approach provides insight into groups' real-time collaborative processes but does not provide the complete picture.

A requirement for my approach to detecting collaboration quality is relationship definitions 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.

A second requirement is that 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 my approach. This has been an active area of research, and a number of approaches to this problem already exist e.g. (Ackad, Clayphan, Maldonado, & Kay, 2012; Clayphan, Martinez-Maldonado, Ackad, & Kay, 2013; R. Martinez, Kay, & Yacef, 2011; Meyer & Schmidt, 2010). However, as described in Chapter 3, existing approaches are not wholly suitable for classroom use. The next chapter details the development of a new approach to distinguishing between users that can make it possible to detect collaboration patterns in real-time and work within the constraints of classroom settings.

4.4 CONCLUSION

I have described small groups of students' use of social regulation processes during collaborative learning activities at a tabletop computer in an authentic classroom setting. My results confirm that high school students do not always have the skills to regulate collaborative work effectively, demonstrating the need for interventions to support the development of these skills.

I 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. I have also demonstrated an approach using these touch patterns in combination to detect the quality of collaborative learning processes in real-time. My approach to modeling collaborative learning is the first to look at metacognitive processes rather than simply quantifying participation or cognitive performance.

The empirical results show that the touch patterns I 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.

Chapter 5. GROUP TOUCH – DISTINGUISHING USERS AT A TABLETOP COMPUTER⁵

This work described in this chapter addresses research question 2: How can a tabletop computer distinguish among users?

The touch patterns described in Chapter 4 depend on being able to distinguish when multiple users are interacting with the computer. The Microsoft PixelSense used in this work does not natively have the ability distinguish between users, as is the case with almost all tabletop computers. I considered a number of existing approaches to solving the problem of distinguishing users. After ruling out the use of external sensors due to practical considerations, I searched for approaches that leveraged the existing capabilities of a tabletop computer.

I initially chose to implement Zhang et al.'s (2012) *See Me, See You* algorithm, which used hand contours captured by the vision system to detect finger orientation and handedness. These touch features could then be used to predict a user's location at the tabletop, thus distinguishing between different users. *See Me, See You* showed promise for use in a classroom setting because it had the best results of any finger orientation-based approach published to date and it did not require any additional sensors. Zhang et al. did, however, make the assumption that users would be willing to restrict their interactions with the computer to their index finger in order to enable user tracking. This artificial constraint on users is fundamentally in opposition to an Ability-Based Design (Wobbrock et al., 2011, 2018) approach to the problem of distinguishing users. Therefore, it was important first to test if *See Me, See You* would be able to handle unrestricted multi-touch interaction.

⁵ Portions of this chapter reproduce text from Evans, Davis, Fogarty, & Wobbrock. (2017). Group Touch : Distinguishing Tabletop Users in Group Settings via Statistical Modeling of Touch Pairs. *CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 35–47.

Following Zhang et al.'s procedure, I trained a support vector machine (SVM) model using touch data from four adults in a lab setting. With 10-fold cross validation, the resulting SVM model correctly classified the user's position around the table for 93.42% of the touches. I expected that accuracy would be lower in the field with the introduction of multi-touch interactions, but that accuracy would still be high enough to make the approach viable for use in my future work given that the index finger tends to be the most commonly used finger even in multi-touch interaction (Kin, Agrawala, & DeRose, 2009).

When tested in the field with 11 high school students in an after-school science program, See Me, See You performed very poorly, correctly predicting the user's location for only 30.87% of touches. I reviewed the videos to understand why the algorithm performed so poorly outside of the lab and identified two primary factors: the wide range of finger and hand positions students used, and their tendency to huddle together around the computer. An entirely different approach was needed.

Most approaches to distinguishing among users have tackled the problem at the individual level, e.g., matching a touch to the location of its owner (Zhang et al., 2012), the chair they are sitting in (Dietz & Leigh, 2001), or the wristband they are wearing (Meyer & Schmidt, 2010). I reframed the problem at the group level because the need to distinguish among individuals only arises when there are multiple users. This led me to begin comparing touches to each other instead of looking for specific individuals' touches. Rather than attempting to track an individual user for the duration of a tabletop activity, I sought to identify when the users interacting with the computer changed—when additional users begin touching the tabletop—and to differentiate among multiple users simultaneously interacting with the computer. This reframing of the problem led to the development of a new approach, called Group Touch.

5.1 DESIGN AND EVALUATION OF GROUP TOUCH

In order to support unconstrained multi-touch interaction, I forego the goal of *identifying* and tracking users (knowing to whom each touch belongs), which has been the focus of many of the aforementioned approaches. Instead, our aim is *distinguishing* among users (knowing one user's touches are different than another's). This means that our approach cannot enable sophisticated personalization of interfaces or authentication, but it can be used to address core usability problems that arise when multiple people interact with a single shared interface, such as determining whether simultaneous touches are by a single person performing a single multi-touch gesture or multiple people performing separate gestures (e.g., to resolve conflicts). Our approach could also be used for modeling collaboration in order to understand group working practices and inform the design of collaborative tabletop applications.

Most approaches to distinguishing among users have tackled the problem at the individual level by trying to match touches to users. By relaxing the goal of tracking and identifying, we reframe the problem at the group level; after all, the need to distinguish among individuals only arises when there are multiple users in a group. This reframing leads us to compare touches to each other instead of attempting to match touches to specific individuals.

Group Touch has two components. The first is a multilayer perceptron (MLP) model trained on touch data collected entirely “in the wild.” The model predicts, given a pair of touches, whether they were carried out by the *same* person or *different* people. The second component is an algorithm that uses the predictions of the MLP model to group touches that were likely to have been carried out by the same person. The following sections describe the dataset that was used to develop and evaluate Group Touch and the design of each component in turn.

5.1.1 *Creating the Touch Dataset*

Group Touch was developed and evaluated using touch data collected “in the wild” during a classroom study of collaborative learning at tabletop computers in classroom settings. Data were collected from high-school students using five distinct multi-touch applications (Figure 5-1 and Figure 5-2) on a Microsoft PixelSense in two different educational programs, one of which is the setting of the classroom study described in Chapter 4.

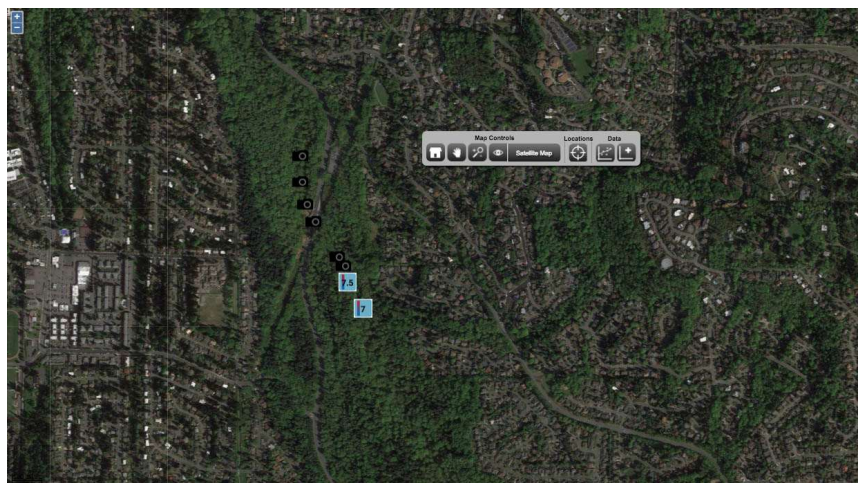


Figure 5-1. The mapping application used in the classroom. The two blue squares represent locations with available water quality data. Touching the squares opened up graphs and charts of the data. The shown toolbar provided access to additional navigation and settings.

The participants in the first study setting—an after-school science program—were eleven 9th to 12th grade students (6 male, 5 female). Figure 5-1 shows the custom-built mapping application used in this setting. The teachers of the science program had requested an application that would enable students to see stream-monitoring data they had collected laid out on a map of their study area that could be panned and zoomed to different levels of detail. Students could use the map to navigate to various data collection locations and to interact with tables, graphs, and images of data. Students interacted with the application using standard multi-touch gestures such as pinch-to-zoom, swipe-to-pan, and multiple-finger rotation. Four sessions were recorded across two class

periods with the students using the computer in groups of five to six for around 30 minutes a session. A total of 1,072 touches were logged in this setting.

Participants in the second study setting—the user-centered design course described in Chapter 4—were sixteen 10th and 11th grade students (6 male, 10 female). Figure 5-2 shows screenshots of the applications that were custom-built for this course. The students used each application for 10-15 minutes in groups of three to four over four class sessions. One application (Figure 5-2, top left) enabled the students to draw on wireframes for their design project. A second application (Figure 5-2, top right) helped groups brainstorm questions for a usability test. A third application helped students improve their search skills (Figure 5-2, bottom left) by finding and comparing resources for their projects. These applications were built using the PixelSense’s native interface components, which fully support multi-touch. The final application (Figure 5-2, bottom right) was a multi-touch-enabled browser plugin for finding and annotating, via drag and drop, real-world examples of usability heuristics. Thirteen group sessions were logged in this setting, amounting to 9,255 touches.



Figure 5-2. The applications used in the second study setting. From left to right, top to bottom, the first three applications used the computer’s native SDK. The fourth (bottom right) application was browser-based. All applications supported multi-touch.

In both classroom settings, the teacher determined the activities that the students carried out, and students were not given any instructions on how they should interact with the tabletop computer. The students elected to stand while using the computer and chose how to arrange themselves around the screen. They touched freely and rampantly, without any rules imposed on them to take turns or “be polite” to their collaborators. In this way, the environment paralleled many authentic classroom situations.

All 17 sessions across both study sites were video recorded with a single camera and every touch was logged by software on the tabletop. The camera was mounted on a tripod to the side of the table, angled down so that the whole tabletop screen was visible. The purpose of the camera was to enable us to manually label the author of every single touch using the video to obtain ground truth.

The touch data collected in both settings proved to be as “messy” as might be expected from unconstrained field settings with high school students. We observed considerable variation in how students carried out standard multi-touch gestures, such as drag and rotation, echoing Hinrichs and Carpendale’s (2011) findings from their study of multi-touch gestures on tabletops “in the wild.” Students were free to move around the tabletop and often did, occasionally pushing and shoving each other. Figure 5-3 shows multiple students simultaneously interacting with the table, which was characteristic of much of the students’ time with the table, even though taking turns and adhering to the social protocols of effective collaborative work would have been beneficial for the learning activities. Additionally, a notable proportion of student interactions appeared to be carried out without an obvious task-related purpose. For example, students regularly spun or flicked on-screen objects while engaging in off-task conversation. These behaviors have also been noted in other reports of tabletop use in classroom settings (Kharrufa et al., 2013).



Figure 5-3. High school students interacting with the PixelSense.

5.1.2 *The MLP Model and Feature Selection*

The first component of Group Touch, the MLP model, predicts whether a pair of touches was carried out by the same person or different people based on three features (Figure 5-4): (1) the difference in touch orientations; (2) the pixel distance between the two touches; and (3) the time difference between touches, in milliseconds.

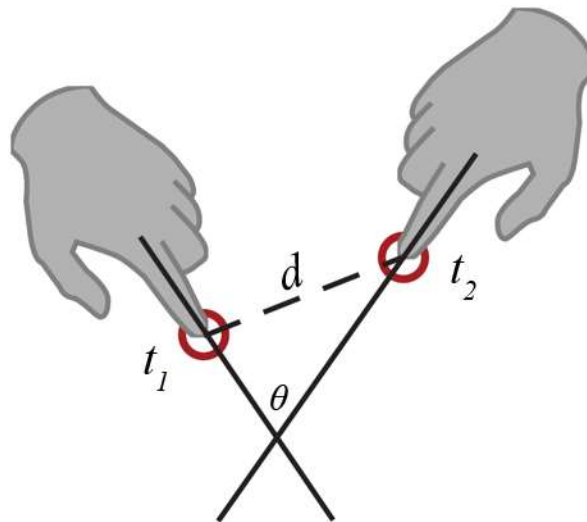


Figure 5-4. The touch features used to predict whether a pair of touches was carried out by the same user are: (1) the difference θ between the touches' orientations, (2) the distance d between the touch points, and (3) the time between t_1 and t_2 .

The three features were selected based on the following reasoning. Consider Figure 5-5, a raw image captured by the PixelSense's on-board cameras. Even without any context, a human viewer can determine that the two touch points detected by the computer (the brightest points in the image) are being carried out by two different people—we can clearly see two hands in a configuration which would be very difficult for a single person to achieve given the size of the screen (40"). The primary features of the two touch points that enable us to infer that we are seeing the hands of two separate people are the difference in their orientations and the distance between the touch points. Touch orientation refers to the orientation of the finger's contact area detected by the computer's vision system.



Figure 5-5. A raw image captured by Microsoft PixelSense on board cameras showing two different people touching the screen.

In Figure 5-5, the touch points are concurrent, which helps us determine that we are seeing two different people. If the touch points were not concurrent—for example, if the touch on the left of the screen took place 30 seconds after the touch at the top—we would be less confident that we

were seeing two people rather than one person who has taken a few steps to their right in order to reach a different area of the screen. Therefore, a third feature, the time that has passed between the touches, is also necessary to make a prediction that the two touches were carried out by the same person or two different people.

Based on the above examples, the three features included in the model are all needed to make a prediction of “same” or “different,” but the exact nature of the relationship between them is not clear and is likely non-linear. Although including just three features makes for a simple model, the non-linearity adds complexity. The MLP classifier has been demonstrated to be well-suited for complex problems where the relative importance of each feature is unknown (Gardner & Dorling, 1998), making it a good choice for this work. My main requirement for classifier selection was that the model output a probability with each classification, as this was necessary for the grouping algorithm, described below. I selected MLP as the classifier after using leave-one-out nested cross validation in Weka (Hall et al., 2009) to test a number of classifiers that output a probability. The Weka classifiers tested included Logistic Regression, LibSVM, and BayesNet among many others. My use of a classifier learned from data, instead of a heuristic approach, also means that future versions of Group Touch could incorporate additional touch features found to be predictive.

5.1.3 *The MLP Model, Pre-processing, and Training*

The purpose of the MLP model is to predict, given a pair of touches, whether or not those touches are carried out by the same or different people. To prepare a touch log to train the model, each touch was processed sequentially.

Pre-processing of the raw touch logs began by converting individual touches into pairs of touches. Each touch pair consisted of a given touch and the previous touch by each user in the

session. The values of each of the features—difference in orientation, distance, and time between touches (see Figure 4)—were calculated to form an instance with a class label of either “same” or “different.” A single instance comprised the new touch and the last touch carried out by the same user (with the class label, “same”). $N-1$ instances comprised the new touch and the last touch carried out by each of the other users in the group (with the class label, “different”). So, in a group of N users, each new touch resulted in N touch pairs, or instances.

Consider, for example, that several of the user groups comprised four students. In these groups, each new touch resulted in four instances (or touch pairs)—one instance formed from the differences between the new touch and the last touch by the same person, and three instances describing the differences between the new touch and the last touch by each of the other three people.

Creating touch pairs in the aforementioned manner helped to account for differences in how groups of users interact with any given application, as it always produced a similar distribution of instances with the “same” versus “different” class label for a group of N users regardless of the activity or the users’ working style. For example, a group of four users will always produce three instances with the class label “different” for every instance with the class label “same.” This consistency is important to prevent over-fitting the MLP model to a particular application or group working style (e.g., turn-taking versus parallel interactions).

Before training the model, each feature was normalized to the range $[0, 1]$. For two of the three features, normalization was straightforward as the possible range of values is inherently limited—the difference in orientation can only fall between 0° and 180° , and the maximum distance between two touch points is determined by the length of the screen diagonal. The elapsed time between touches, however, does not have a clear upper limit. I set the upper limit for time between

touches to be the 90th percentile for our full dataset, or about 307 seconds, in order to remove outliers. Instances with a greater time between touches were dropped from the training data. Although this upper limit may seem high at just over five minutes, it was fairly common in my dataset for groups to take breaks from interacting with the screen to discuss something, or for some individual users to refrain from interacting for extended periods of time, therefore leading to lengthy gaps between touches by particular users.

I trained and optimized Group Touch’s MLP models in Weka using leave-one-out nested cross validation to prevent overfitting. This method resulted in 17 models—each one trained and optimized on the data from 16 of the study sessions and evaluated on the unseen data from the remaining study session.

For every touch logged, as described above, there were more instances generated with the class label “different” than “same,” so the data were heavily weighted towards the “different” label (77.40% of instances across all sessions). To avoid overfitting the models to the majority class label, I applied the SMOTE filter (Chawla, Bowyer, Hall, & Kegelmeyer, 2011) to balance the training data, resulting in an almost 50%-50% split between the two labels. MLP models can potentially be sensitive to the order of training instances, so I also applied the Randomize filter to shuffle instances before training.

5.1.4 *Touch Grouping Algorithm*

After a touch pair is categorized with the “same” or “different” label, the next step is to establish groups of touches that are likely to have been carried out by the same user. The goal of Group Touch is to detect when additional users begin touching the screen and distinguish among multiple users as they are concurrently interacting with the computer; it is not intended to track or identify individual users for the duration of an activity. Key to our approach is that a “group” is

not synonymous with a “user”—it is a group of touches that belong to a single user, representing a period of sustained interaction. Figure 5-6 illustrates the algorithm used to place touches into groups.

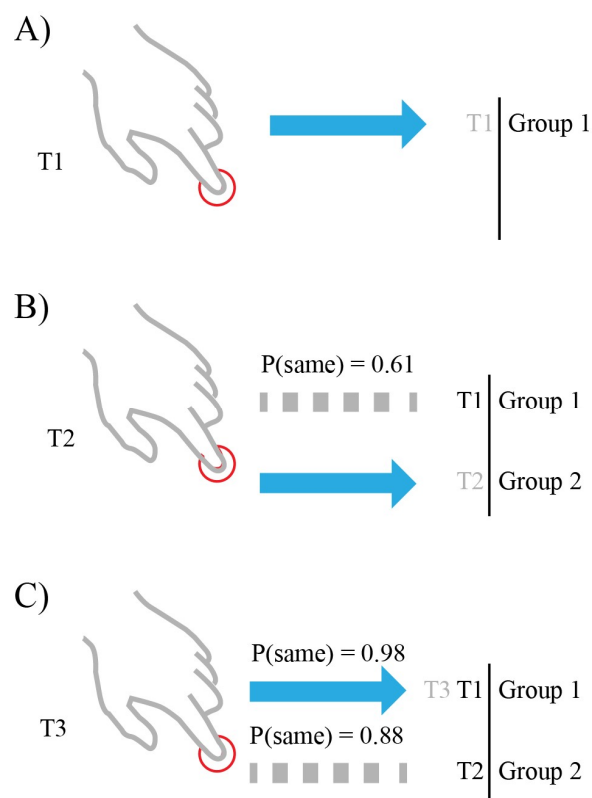


Figure 5-6. An example of the algorithm to place touches into groups. A) The first touch creates the first group. B) For subsequent touches, we create touch pairs with the last touch of each group, then query the model for each pair. We create a new group if no pairs have a same-author with probability $\geq p$, set to 0.80 in this example. C) If multiple touch pairs have a same-author probability $\geq p$, we add the touch to the group with the highest probability.

The first touch creates the first group (Figure 5-6A). For every subsequent touch, a touch pair is created with the last touch in each group established so far. When the MLP model is queried, it returns the probability that the last touch belonged to the same user. If $P(\text{same})$ is less than a specified threshold (p), a new group is started (Figure 5-6B); otherwise, the touch is assigned to the group with the highest $P(\text{same})$ (Figure 5-6C).

The value of threshold p can vary between 0.5 and 1.0. Higher thresholds mean that it is more likely a touch group will be made up of touches carried out by only one user because a more confident “same” prediction is required for a touch to be added to an existing group. However, higher thresholds also lead to new groups being created more often, with each group containing fewer touches and lasting a shorter period of time. In contrast, lower thresholds lead to touch groups that last for longer and contain more touches, but also increase the likelihood that touches by different users will be mistakenly grouped together because a weaker “same” prediction is sufficient to add a touch to an existing group.

To evaluate the grouping algorithm and establish a suitable value for threshold p , I tested it on each session of touch data using a leave-one-out procedure. This testing involved using the algorithm to group the touches in a session according to predictions from the MLP model trained on touch data from the other 16 sessions. As well as being dependent on the performance of the MLP model, the outcome of the grouping algorithm would also be affected by the value of the probability threshold, p . Therefore, I tested the algorithm with a range of probability threshold values from 0.5 to 0.9, in increments of 0.1. This test was repeated with each of the 17 sessions of touch data in order to determine an optimal value of p . I determined that a threshold of 0.80 was suitable for most purposes, a decision that I discuss in more detail in the next section, together with implications of this decision.

Approaches that use touch orientation to match touch points to hands (Dang, Straub, & André, 2009; Zhang, Zhang, & Chen, 2014) have shown thumb touches to be a source of error as the orientation of the thumb can be different from other fingers on the same hand. I anticipated this would also be an issue for Group Touch. Thumb touches are relatively infrequent in general tabletop interactions and do not feature in most standard gestures, with the exception of pinch-to-

zoom. Thumb touches, however, are heavily used when typing on a virtual keyboard—typically when pressing the spacebar. To remedy this potential problem, groups of touches on a virtual keyboard were merged when the same keyboard instance was touched and the touches occurred within 384 ms of each other. This threshold was chosen based on the typing speeds of participants in a study of typing on virtual keyboards (Findlater, Wobbrock, & Wigdor, 2011); it represents the time between keystrokes for the slowest participant. Touches to the same keyboard within this time could be assumed to have been carried out by the same user.

My evaluation of Group Touch was carried out offline but was set up as if it were grouping touches interactively at run time—touches were processed in the order they were received and logged. Accuracy was calculated as the percentage of touches added to a group where the preceding touch correctly had the same user, using the manual labels assigned from the videos as ground truth.

5.1.5 *Results*

Group Touch's overall accuracy for a session ranged from 83.43% to 99.44%, with a mean of 92.92% (SD=3.94%). However, only 2.66% of all touches logged were incorrectly added to a group of touches by a different user. This discrepancy between the overall accuracy and the number of touches added to incorrect groups occurs because my selection of a p threshold of 0.80 means the algorithm favors creating a new group of touches in the absence of a strong prediction of “same user” for any existing group. Therefore, groups of touches associated with a particular user persist for short periods of time but are highly accurate. Table 5-1 shows the complete results by session.

Table 5-1. Group Touch results by session. Columns (d) and (e) show how just the MLP model performed on the test data for each session. Column (f) shows the performance of the grouping algorithm with threshold $p=0.80$. Column (f) is therefore the performance of Group Touch.

| (a) Session - Application | (b) # of Touches | (c) # of Users | MLP Model | | (f) Grouping Accuracy (%) |
|---------------------------|------------------|----------------|-------------------------------|--------------------------|---------------------------|
| | | | (d) Accuracy on Test Data (%) | (e) Area under ROC Curve | |
| 1 - Brainstorm | 601 | 5 | 96.87 | 0.99 | 99.44 |
| 2 - Heuristics | 105 | 4 | 89.32 | 0.96 | 96.30 |
| 3 - Heuristics | 328 | 4 | 91.49 | 0.96 | 95.74 |
| 4- Heuristics | 241 | 5 | 94.29 | 0.97 | 98.40 |
| 5- Heuristics | 240 | 5 | 87.81 | 0.91 | 89.89 |
| 6 - Map | 104 | 5 | 92.19 | 0.93 | 94.74 |
| 7 - Map | 262 | 6 | 91.85 | 0.94 | 91.84 |
| 8 - Map | 519 | 5 | 93.85 | 0.95 | 91.76 |
| 9 - Map | 187 | 5 | 84.96 | 0.89 | 88.97 |
| 10 - Resources | 285 | 4 | 90.50 | 0.95 | 94.04 |
| 11 - Resources | 448 | 4 | 89.27 | 0.95 | 95.78 |
| 12 - Resources | 271 | 5 | 89.66 | 0.92 | 90.10 |
| 13 - Resources | 296 | 5 | 88.28 | 0.92 | 89.50 |
| 14 - Wireframes | 465 | 5 | 89.53 | 0.96 | 93.07 |
| 15 - Wireframes | 435 | 5 | 89.45 | 0.96 | 95.16 |
| 16 - Wireframes | 241 | 5 | 87.54 | 0.92 | 83.43 |
| 17 - Wireframes | 295 | 4 | 87.80 | 0.95 | 91.49 |
| Mean: | 313.12 | 4.76 | 90.27 | 0.94 | 92.92 |
| SD: | 138.32 | 0.56 | 2.91 | 0.03 | 3.94 |

To test the first component of Group Touch in isolation, the MLP model that predicts whether a pair of touches was authored by the same person or different people, I used leave-one-out cross-validation to evaluate the model, withholding a different session's touch data each time. Within the training data for each fold in the leave-one-out cross validation, we conducted a 10-fold cross validation of the MLP model. These scores were highly consistent across sessions—the mean cross validation score was 89.22% (SD=0.29%). When the trained models were then tested on the withheld touch data, the mean accuracy was 90.27% (SD=2.91%). Leave-one-out test results for the MLP model are shown in Table 5-1 (d). Figure 5-7 shows the ROC curves.

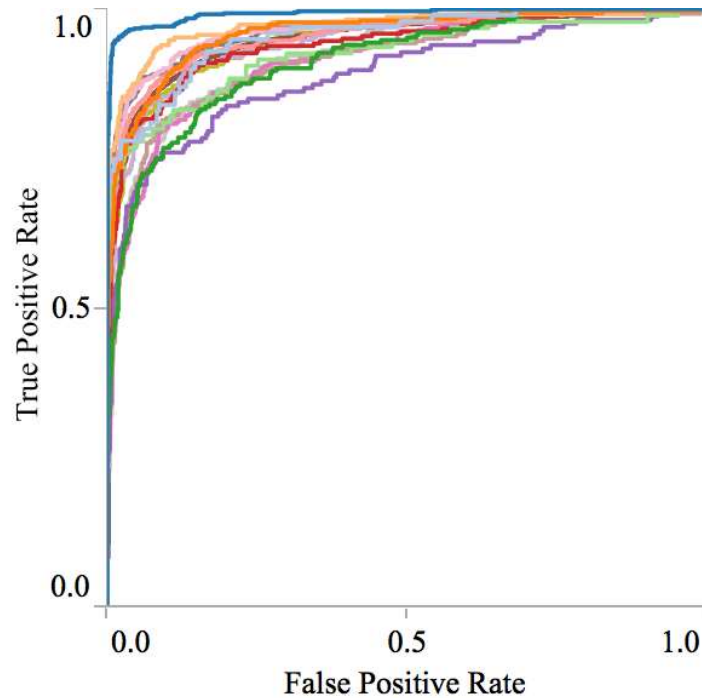


Figure 5-7. The ROC curves for each MLP model. The area under the curve, listed in Table 5-1, ranged from 0.89 to 0.99.

The second component of Group Touch, its grouping algorithm, adds a touch to the group for which the MLP model has returned a prediction of “same” person with the highest probability above a threshold p (see Figure 5-6). I tested threshold values between 0.5 and 0.9 in 0.1 increments and found that mean accuracy increased from 88.83% to 94.23% (Figure 5-8). At the same time, the median duration of a group—the time elapsed from the first touch in a group to the last—decreased from 7,207 ms to 2,383 ms. Therefore, Group Touch can distinguish among simultaneous users for longer periods of time when the value of p is lower but with a tradeoff of lower accuracy within touch groups.

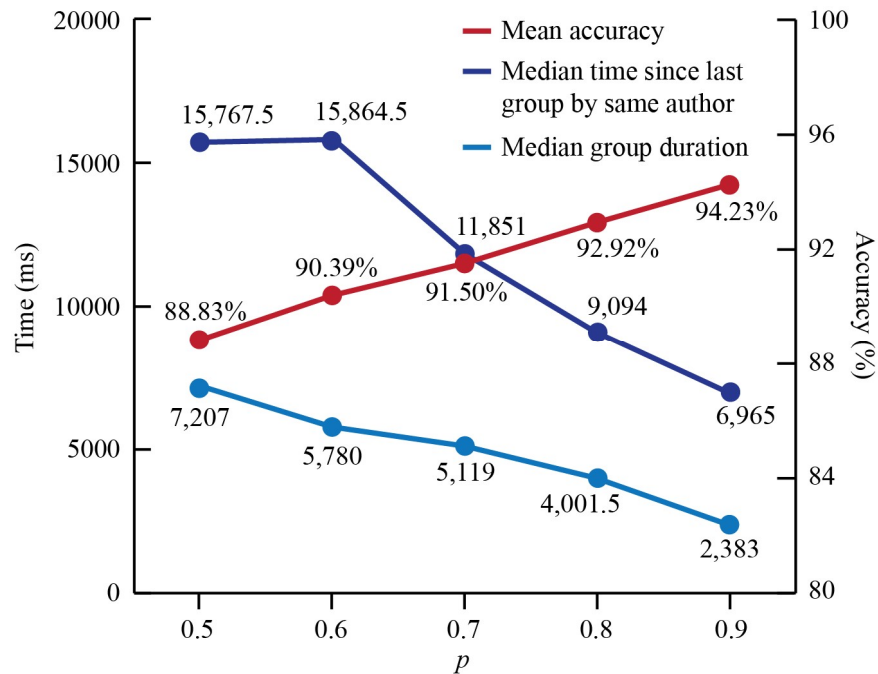


Figure 5-8. As the value of p increases, the accuracy of Group Touch increases. However, the median time since the last group of touches with the same user and the median duration of groups decreases.

To evaluate the threshold values, I also tracked another measure: the time between groups of touches belonging to the same person (i.e., the time that elapsed since that person last touched the screen). For example, Figure 5-8 shows that when $p=0.8$, the median time since the last group by the same user was 9,094 ms. This means that when a new group is started with a touch by user X, a median of 9,094 ms have passed since X last touched the screen. When the time between groups of touches belonging to the same person is brief, the system is being conservative in assigning new touches to existing groups. It therefore becomes more likely that sustained input by a single person will be ascribed to multiple people. Figure 5-8 shows that, as the value of p increased, the median time between groups of touches belonging to the same person decreased. Therefore, although accuracy within groups will be highest when $p=0.9$, longer sequences of

related gestures by a single user are more likely to be split into multiple groups than when p is lower.

Based on Figure 5-8, I chose 0.80 as the optimal value of p . Although the average accuracy increases to 94.23% when $p=0.90$, both time metrics decrease to a point where Group Touch would only be reliably able to distinguish among users for brief interactions. At $p=0.90$, Group Touch still has valuable use cases (see the Discussion, Section 5.2) but at $p=0.80$, Group Touch retains high accuracy and is able to distinguish among simultaneous users over longer periods of time, making it useful for a wider range of applications. When p was set to 0.80, the midspread (middle 50%) of groups lasted between 0.65 and 12.02 seconds and contained 2 to 6 touches. The longest group recorded lasted 77.10 seconds and contained 80 touches. Table 5-1(d) shows the results of the evaluation of the grouping algorithm with the threshold, p , set to 0.80.

To understand why Group Touch creates new groups for touches by an existing user, I investigated the differences between touches that resulted in a new group and those that were added to an existing group. Figure 5-9 shows the average features of touch pairs by the MLP model's probability that a pair was carried out by the same user, $P(\text{same})$. The "New groups created" region shows the average features of false negatives—touch pairs that were actually by the same user (based on the labels assigned manually) but were incorrectly labeled by Group Touch as being by different users, leading to the creation of a new group. For these touch pairs, the MLP model returned a probability, $P(\text{same})$, below our threshold of 0.80 (see Figure 5-6B). Moreover, 62.41% of false negatives occurred when the MLP model returned $P(\text{same}) < 0.1$, a strong prediction that the touches were carried out by different people. Clear trends emerged for each feature as $P(\text{same})$ increased. Typically, touch pairs with the lowest values of $P(\text{same})$ had large values for time between touches ($M=162.54$ s, $SD=76.70$ s). Difference in orientation ($M=69.86^\circ$, $SD=47.14^\circ$) and

distance between touch points ($M=601.37$ px, $SD=347.47$ px) also tended to be higher but there was more variation for these features. These results suggest that extended time between touches results in the creation of a new group when an existing group of touches by the same user is available.

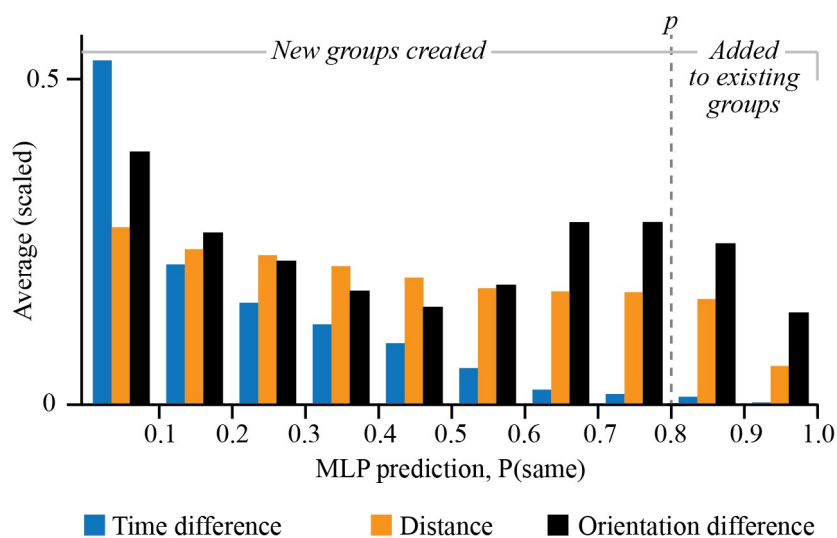


Figure 5-9. Average features of touch pairs by the probability they were carried out by the same user according to the MLP model.

For comparison, Figure 5-9, “Added to existing groups,” shows the average features of touch pairs that resulted in a touch being added to an existing group. These are pairs where the MLP model returned the highest value of P(same) above the 0.80 threshold (see Figure 5-6B). 93.58% of these touch pairs were true positives that were correctly assigned—both touches were carried out by the same user. 91.16% of these touch pairs had a P(same) of at least 0.90. The time between touches for pairs with a P(same) of at least 0.90 was brief ($M=1.03$ s, $SD=1.54$ s). Difference in orientation ($M=25.25^\circ$, $SD=24.75^\circ$) and distance between touches ($M=130.35$ px, $SD=130.79$ px) were also at their lowest when P(same) was at least 0.90. This means that a touch was more likely to be

correctly added to an existing group when it was close to the last touch in the group in terms of all three features.

5.2 DISCUSSION: GROUP TOUCH AS AN APPROACH FOR DISTINGUISHING AMONG USERS

Although the primary motivation for developing Group Touch was simply to make it possible to detect the collaboration patterns described in Chapter 4, the approach has other, less specific uses. Group Touch's overall accuracy (Table 5-1(d)) shows that it was successful at grouping touches that belonged to the same user. There was little variability in the accuracy of both the MLP model (Table 5-1(c)) and the grouping algorithm (Table 5-1(d)) across the 17 sessions of touch data collected in five different applications. These results indicate that Group Touch is user-independent and can perform consistently across a range of applications.

Group Touch favors creating a new group of touches when the MLP model does not return a strong prediction of "same user" for any touch pair. This means that group assignments are highly accurate, with only 2.66% of all touches in our dataset being added incorrectly to a group by another user. However, this also means that it is important to understand when and why the model is unable to return a strong prediction of "same user," leading to the creation of new groups. The trends shown in Figure 5-9 suggest that extended time between touches is a strong reason that new groups are created when an existing group of touches by the same user is available. Therefore, Group Touch is likely to perform better and create groups that persist for longer when users are regularly interacting with the screen, so time between touches is brief. Given that Figure 5-9 also shows that touches are correctly added to groups when the time and distance between touches are low, Group Touch is probably not suited for applications where users frequently jump around different areas of the screen. This might explain why session 16 had a much lower accuracy than

all other sessions at 83.46%, 6.46 percentage points lower than the next lowest result. The students in this session spent much of the time flicking the on-screen objects at other objects to make them bounce around the edges, then reaching across the full length of the screen to retrieve the object.

The value of the threshold p used to assign touches to groups of touches by the same user impacted the group duration and time between groups by the same user, as well as the accuracy. I selected 0.80 as the value of p because it resulted in a high grouping accuracy, and the decrease in group duration was much steeper with a p above 0.80. However, the value of p and the resulting tradeoff between group duration and accuracy affects the potential use cases for Group Touch and therefore should be adjusted depending on how Group Touch is to be used.

The simplest and most generalizable use case for Group Touch is resolving conflicting input that occurs when multiple people are simultaneously interacting with the same interface. For example, without the ability to distinguish among users, the touch input of multiple people attempting to drag an object in different directions will look very much like that of a single user carrying out a pinch-to-zoom action. Another example would be a user attempting to close a window that another user is actively working in (Morris et al., 2004). Group Touch makes it possible to build applications that can detect these types of conflict and determine how the application should respond, because it is not necessary to know the identity of the users, just whether or not multiple users are interacting with the application concurrently. In this scenario, Group Touch would affect the application's behavior in ways that are visible to users. Therefore, the high accuracy of the top value of p , 0.90, would likely be favored over groups that persist for longer, particularly as conflicting gestures will occur with a small number of overlapping touches.

In addition to resolving conflicting gestures, there are numerous application scenarios where the ability to detect when multiple people are attempting to interact with the same object

would be useful. For example, in a collaborative document editing application, it would be useful to prevent concurrent interaction in some situations (e.g., one user scrolling a page while another is making edits). Online collaboration tools, such as Google Docs, handle this situation by allowing each user to control navigation of their own view of the document—they can see others' edits as they occur but they can move around their own view of the document independently. This approach would not transfer to a similar application on a shared screen. Group Touch could help in this scenario by making it possible to temporarily block interaction from users other than the person actively working in the document. Additionally, tracking concurrent versus individual edits to a document could make document revision history more useful.

The ability to temporarily block other users from interacting with an object that a user is currently holding could also be used to enforce turn-taking in a collaborative learning environment, or as a game mechanic in a competitive game where players have to steal objects from their opponents. Taking the opposite approach to the same incident—only responding to input if multiple people are simultaneously interacting with an object—would make it possible to support cooperative gestures. Cooperative gestures have been shown to be useful for increasing awareness of important functionality and increasing participation in applications for collaborative work and play (Morris et al., 2004, 2006).

Group Touch could also be used to model the collaborative processes of small groups working at tabletop computers. For example, the ability to distinguish when multiple people are interacting with an application makes it possible to detect when users are taking turns, interfering with the actions of others, or working in parallel, all of which have been shown to impact collaborative learning (Evans, Wobbrock, & Davis, 2016; Fleck et al., 2009; Kharrufa et al., 2013; Pontual Falcão & Price, 2010; Rick et al., 2009). For this use case, it would be important to capture

longer interactions that contain more than just a small number of isolated gestures in order to identify sustained interaction patterns. Therefore, a lower p threshold, which creates groups that persist for longer, would be preferable to increasing accuracy.

Application designers could also use information about when and where conflicting gestures occur, and how groups use their applications, to evaluate and refine their designs to enable better collaborative use. For example, Group Touch would make it possible to automatically identify areas of the screen where there is typically only one user interacting at a time and those where there tend to be multiple users interacting. This information could be used to determine placement of shared interface elements and how to divide the screen into individual and shared workspaces.

Group Touch offers an enhancement to existing touch-detection on large, multi-user touchscreens that adds a new capability without breaking existing touch functionality when errors occur. For example, Group Touch false positives may cause an application to interpret the touches of multiple users as a single gesture. This outcome is typical of most current commercial tabletops. Therefore, when Group Touch produces a false positive (7.08% of touches assigned to existing groups in this study with the threshold set to 0.8), behavior will be the same as in current systems.

In the case of false negatives, the primary causes are extended time between touches, or touches close together in time that have very different orientations or locations. How false negatives impact a user will largely depend on what an application does with information provided by Group Touch. For example, when detecting and resolving conflicting gestures, false negatives due to touches that are far apart in time are unlikely to produce a conflict in the interface and therefore unlikely to affect the user. In some cases, however, a false negative could cause a zoom

gesture to look like two separate pan gestures. These cases occur infrequently and application-level design could help to mitigate any impact.

The main limitation of Group Touch is that it does not track or identify users for the duration of an activity, and therefore cannot enable personalization, such as color-coding each user's touches (Zhang et al., 2012), or tracking individual contributions to the group effort (Martinez, Collins, et al., 2011). It also cannot be used for authentication (Blazica, Vladusic, & Mladenic, 2013; Holz & Baudisch, 2013). However, existing approaches to distinguishing users that are able to provide these features either rely on external sensors or artificially constraining interaction. As stated, Group Touch is intended for situations where such methods are impractical or undesirable.

Additionally, the applications I used to design and evaluate Group Touch all featured highly collaborative tasks in which group members worked with shared objects. I expect accuracy would be about the same for other tasks of this nature. However, I expect that Group Touch may have lower accuracy in situations where individual users frequently interact with objects in very different areas of the screen (e.g., rapidly reaching from the top left corner to the bottom right). Based on the analysis of touch group characteristics (Figure 5-9), these types of interactions would look like touches carried out by different people. In applications requiring this functionality, Group Touch may not be the best option for distinguishing among users.

Finally, I have not identified the exact situations where Group Touch performs poorly. For example, I have not investigated how good Group Touch is at recognizing two-handed input by a single person versus single-handed input. This limitation is because the touch data used to develop Group Touch was collected "in the wild" in uncontrolled field settings, and extracting these types

of specific interactions would be an extremely laborious process. A controlled lab study would likely be necessary to gain this kind of insight.

5.3 CONCLUSION

I have presented Group Touch, an approach to distinguishing among multiple simultaneous users by detecting when the users interacting with a vision-based tabletop computer change. This approach achieved a mean accuracy of 92.92% ($SD=3.94\%$). Group Touch is the first approach to distinguishing users that was designed and evaluated using data collected entirely “in the wild.” It uses only the built-in capabilities of the tabletop hardware and does not require the use of extra sensors, making it more flexible and easier to deploy than many existing approaches.

Group Touch is not intended to replace existing approaches to distinguishing users, nor is it a suitable approach if individual users need to be identified and tracked for the duration of a session. Instead, it is intended to bring the capability of distinguishing among concurrent users to new settings that other approaches cannot support, namely, settings where use of external sensors is impractical and artificially constraining interaction is undesirable.

Chapter 6. DETECTING COLLABORATION QUALITY IN REAL TIME USING GROUP TOUCH

The evaluation described in the previous chapter positioned Group Touch as a general-purpose tool for distinguishing among simultaneous users at a tabletop computer. In this chapter, I describe an additional analysis to assess how Group Touch could be used for real-time detection of collaboration quality. Using only touch data to detect the quality of groups' collaborative learning process means that at least half of the picture is already lost. But, the touch patterns found in Chapter 4 show that, even with access to only a portion of the collaboration, it is possible to detect certain key indicators of the quality of collaboration. However, using Group Touch to distinguish among users and therefore enable detection of the touch patterns introduces additional noise and error. It is important to understand how well this method of detecting collaboration quality will work in practice. The goal of the analysis in this chapter was to identify the probability threshold that produced collaboration quality categories as close as possible to the categories produced using the touch data hand-labeled for each touch author. The analysis revealed that each of the touch patterns identified in Chapter 4 has a different optimal value of the probability threshold: 0.8 for Unrelated Touches and 0.5 for Overlapping Unrelated Touches. With the probability threshold set to the appropriate value for each touch pattern, intervals labeled as high- or low-quality were accurate 78.57% to 100.0% of the time depending on the pattern and quality label.

6.1 DETERMINING THE OPTIMAL VALUE OF THE PROBABILITY THRESHOLD FOR EACH TOUCH PATTERN

Group Touch uses a probability threshold to assign touches to groups of touches likely to have been carried out by the same person (see Figure 5-6). Touch groups represent a period of

sustained interaction for a single user—the groups do not persist for the entirety of the collaborative activity, and each user will produce multiple groups of touches. There is a tradeoff between the accuracy of the groups and the length of time they persist that is affected by the probability threshold (see Figure 5-8)—as accuracy increases, the length of time that groups typically persist decreases.

In Chapter 5, I identified 0.8 as the value of the probability threshold that provided the best balance between accuracy and group persistence for a broad range of applications and use cases. However, as noted in section 5.2, in practice, the optimal value of the probability threshold will vary depending on the specific needs of an application.

An application that detects quality of collaborative learning using the approach that I have proposed follows the process shown in Figure 6-1. First, Group Touch assigns incoming raw touches to groups of touches likely to have been carried out by the same user. Second, sequences of touches are extracted from the groups using the process described in Table 4-1. Finally, the sequences in two-minute intervals (see Section 4.2) are checked against the touch patterns identified in Chapter 4: Unrelated Touches (UT), which considers the proportion of touches to unrelated objects in sequences belonging to an individual; and Overlapping Unrelated Sequences (OUS), which identifies the proportion of touch sequences that are overlapping in time, a proxy for multiple people actively interacting with the screen at the same time. When there is a high degree of overlap, signaling that multiple people are actively interacting with the screen, OUS checks the proportion of those overlapping touches that involved unrelated objects—are the active students interacting with related objects, indicating a shared focus, or are they interacting with unrelated objects?

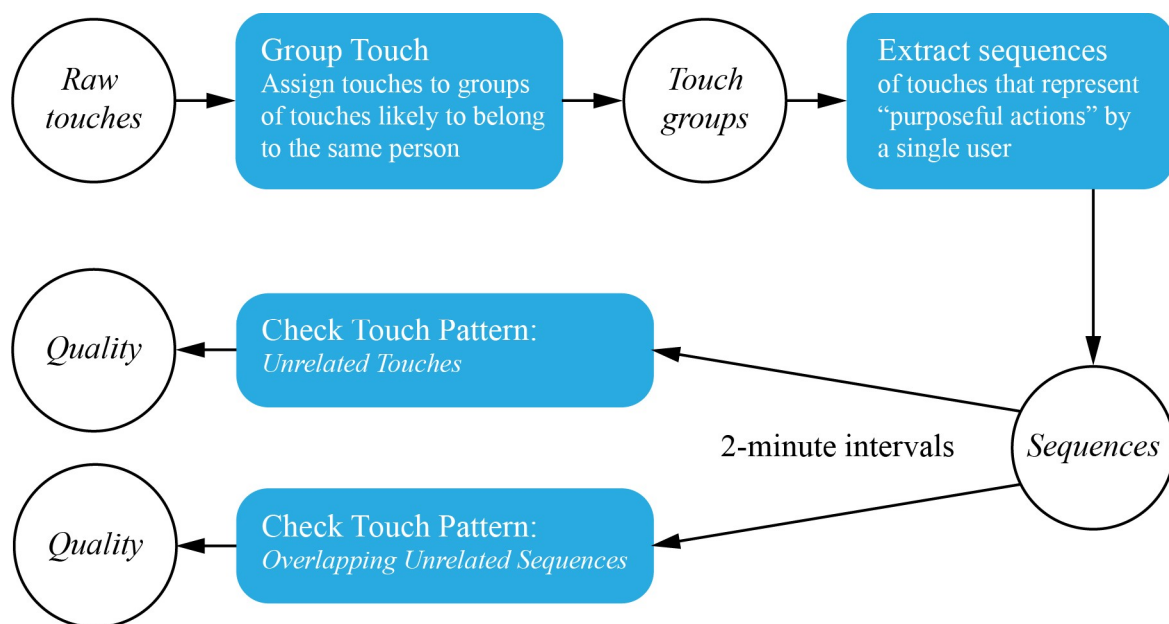


Figure 6-1. The process of automatically detecting collaboration quality: (1) Group Touch assigns raw touches to groups of touches likely to have been carried out by the same user; (2) sequences of touches are extracted from the groups using the process described in Table 4-1; and (3) sequences in 2-minute intervals are checked against the touch patterns identified in Chapter 4

My hypothesis was that different values of the probability threshold would be needed to detect each pattern, and that a higher threshold would work for UT while a low threshold would be better for OUS. Although accuracy of touch grouping is higher with a higher threshold, sequences of touches are more likely to get cut short, particularly when they involve related objects. Figure 5-8 shows that, as the threshold increases, the length of time that touch groups persist and the time since the last group by the same person both decrease. This trend means that it becomes more likely that sequential touches by the same person will be split into different groups when a high probability threshold is used. The implications are different for each of the two touch patterns.

UT is concerned with the proportion of touches to unrelated objects in sequences across the duration of an interval. The proportion of unrelated objects should remain similar no matter

the probability threshold. Therefore, a high probability threshold with high accuracy should be suitable for UT.

OUS is concerned with the proportion of overlapping sequences in an interval. When sequences are cut short by Group Touch, the proportion of overlapping sequences could be substantially lowered, causing the collaboration quality according to OUS to be labeled higher quality than it should be. Therefore, a low probability threshold, which allows touch groups to persist for longer at the cost of lower accuracy, may be more appropriate for detecting this pattern.

The analysis was carried out using the data captured in the classroom study described in Chapter 4, Section 4.2. The analysis was carried out after the data were captured, but touches were processed in the order that they occurred to simulate real-time detection of collaboration quality. Data analysis followed the process described in Figure 6-1: each touch was assigned to a group by Group Touch, sequences representing purposeful actions were extracted from the groups using the rules described in Table 4-1, then, using the interval timecodes from the classroom study described in Chapter 4, two-minute intervals of touch data were checked against the touch patterns, UT and OUS.

The procedure was repeated ten times—for each value of Group Touch's probability threshold from 0.5 to 0.9 in increments of 0.1 and for each of the two touch patterns. I then looked at the differences between the key metrics for each pattern when touches were manually labeled with the owner and when Group Touch was used. For UT, the metric was the percentage of unrelated touches in sequences in each interval. For OUS, the metrics were the percentage of sequences in each interval that overlapped and the percentage of unrelated touches across sequences that overlapped. The results of this evaluation are shown in Figure 6-2.

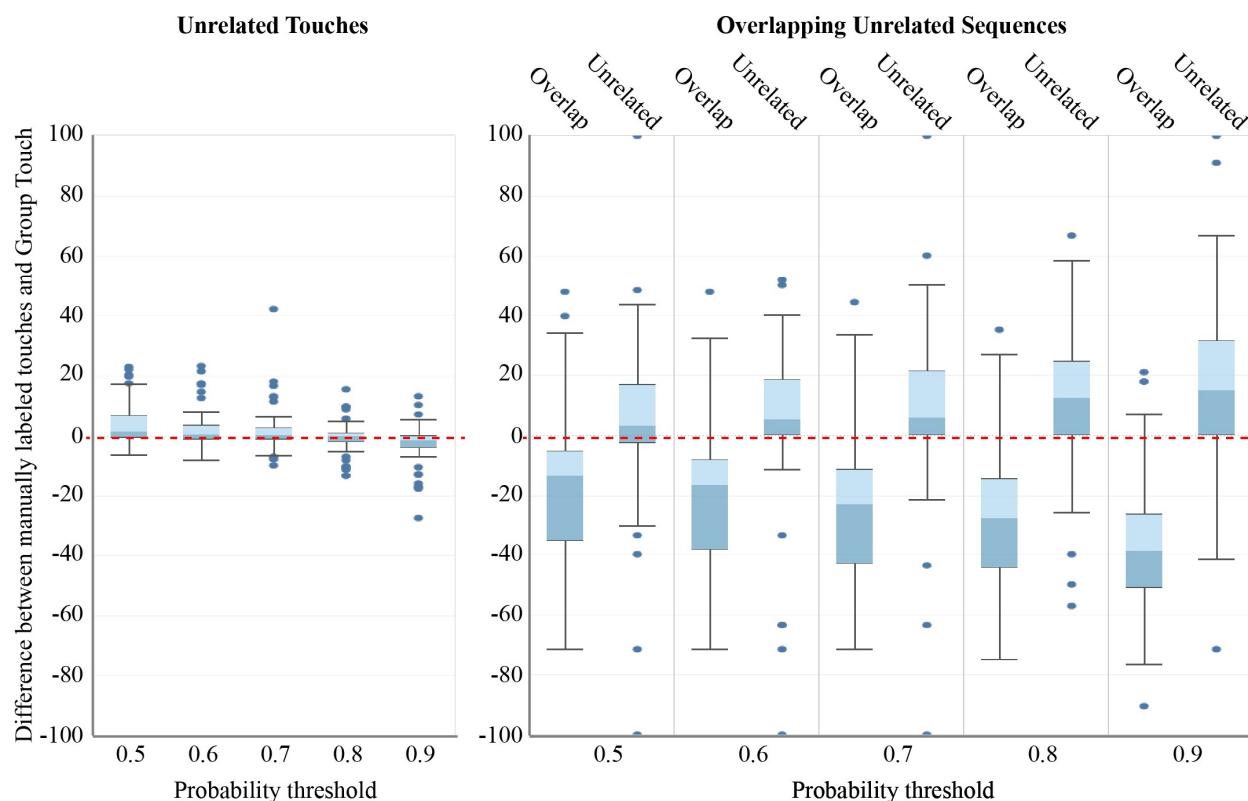


Figure 6-2. The differences between the output of each pattern’s key metrics when using touches that were labeled manually versus touches labeled by Group Touch. The box plots show the distribution of the differences for all two-minute intervals at each probability threshold. The tighter the distribution and the closer to zero on the y-axis, the closer the Group Touch-generated outputs were to those derived from manually labeled touch data.

The key to interpreting these results is that the tighter the distribution of the box plot, and the closer the median to zero on the y-axis, the better—when the distribution is tight and the median is at or close to zero, there were minimal differences between the pattern outputs when using Group Touch to label touches and when using the manually labeled touches. For UT, the distributions are tight and the medians are close to zero at each value of the probability threshold, meaning that Group Touch produced results that were reliably similar to those produced using manually labeled touches. Based on the results in Figure 6-2, I selected 0.8 as the optimal value of the probability threshold to use with the UT pattern because, at 0.8, the median difference between Group Touch and the manually labeled touches was equal to zero and the box plot had the tightest distribution.

The results for OUS, on the other hand, show that there were considerable differences between the pattern metrics derived from touches labeled by Group Touch and those derived from manually-labeled touch data. Figure 6-2 shows that a probability threshold of 0.5 resulted in the smallest differences between Group Touch and manually-labeled touch data on both metrics, although the differences, particularly for the proportion of overlapping sequences, are still quite large—the median difference for the proportion of overlapping sequences was -13.5% and the median difference for the proportion of touches to unrelated objects between overlapping sequences was 3.0%. The OUS boxplots all have much larger distributions than occurred with UT, indicating more variation in terms of how Group Touch impacted the calculation of the metrics. These results mean that the limitations of Group Touch, specifically its inability to track users for longer periods of time, may lead to unreliable results for the OUS pattern.

Before conducting this analysis, I hypothesized that each touch pattern would have a different optimal probability threshold and that the threshold for UT would be higher than the threshold for OUS. This hypothesis proved to be true, with the optimal threshold for UT being 0.8 and the optimal threshold for OUS being 0.5.

6.2 EFFECT OF USING GROUP TOUCH ON COLLABORATION QUALITY LABELS

The previous section investigated the impact of using Group Touch to distinguish among users on the metrics associated with each touch pattern. Although the results for UT showed that Group Touch produced metrics that were close to those derived using manually-labeled touches, this was not the case for OUS. However, my approach for detecting collaboration quality, described in Chapter 4, does not directly use the actual values of those metrics. Instead, the metrics are used to categorize an interval as high, medium, or low quality depending on which side of a given threshold the metric falls.

In this section, I investigate the effect that using Group Touch has on the collaboration quality labels assigned to each interval. Using the values for each metric derived in the previous section at each pattern's optimal value of the probability threshold, I assigned the appropriate quality label to each interval and then compared the labels to those generated from the manually-labeled touch data in Chapter 4, Section 4.2. In this comparison, the quality labels generated from the manually-labeled touch data are treated as the ground truth. A quality label derived from touches processed by Group Touch is considered correct when it matches the ground truth label. In total, 66 intervals of touch data were recorded in the classroom study. For UT, 84.85% of intervals were labeled correctly when using Group Touch. For OUS, 77.27% of intervals were correct. The complete results by pattern and quality label are shown in Table 6-1.

Table 6-1. The number (#) of intervals by the quality label produced using Group Touch and the percentage of intervals where that quality label matched the label generated using the manually-labeled touches.

| Pattern | High Quality | | Medium Quality | | Low Quality | | Overall |
|---------------------------------|--------------|-----------|----------------|-----------|-------------|-----------|-----------|
| | # | % correct | # | % correct | # | % correct | % correct |
| Unrelated Touches | 27 | 88.89 | 13 | 53.85 | 26 | 100 | 84.85 |
| Overlapping Unrelated Sequences | 28 | 78.57 | 3 | 0.00 | 35 | 82.86 | 77.27 |

Table 6-1 shows that, when using Group Touch, collaboration quality labels for both patterns can be considered at least fairly reliable when high or low-quality. Medium-quality labels cannot be trusted, however. As could be expected, based on the analysis in the previous section, the quality labels generated by UT using Group Touch data matched the ground truth more often than those generated by OUS.

In Chapter 4, I concluded that I could have the most confidence in the touch patterns' judgment of collaboration quality when an interval was labeled either high or low-quality by both

patterns. Table 6-2 shows that, when using Group Touch, high-high and low-low remained the most common combinations output from the touch patterns and the most clearly skewed positive or negative, respectively. Both patterns categorized intervals as high quality for all but one of the 11 intervals coded “positive” during video analysis. All other combinations of the touch pattern labels skewed negative with the exception of high-low, which was evenly split between “mixed” and “negative.”

Table 6-2. The distribution of collaboration quality codes from the video data (Chapter 4, Section 4.2.3) for each possible output of the touch patterns using Group Touch to distinguish among users. Note that intervals labeled high quality by both touch patterns (first data row in table) were mostly made up of episodes coded positively during video analysis and that intervals labeled low quality by both touch patterns (last data row in table) were mostly made up of episodes coded negatively during video analysis.

| (a) Pattern quality label (using Group Touch) | | (b) # of Intervals (ground truth video codes) | | |
|---|---------------------------------|---|-------------|----------------|
| Unrelated Touches | Overlapping Unrelated Sequences | Positive codes | Mixed codes | Negative codes |
| High | High | 10 | 3 | - |
| High | Med. | - | - | 1 |
| High | Low | - | 5 | 5 |
| Med. | High | - | - | 5 |
| Med. | Med. | - | - | - |
| Med. | Low | - | 1 | 7 |
| Low | High | 1 | 1 | 4 |
| Low | Med. | - | - | 2 |
| Low | Low | - | 2 | 15 |

The goal of this work is to detect collaboration breakdowns in real-time and adapt to support more effective collaboration. Table 6-2 shows that intervals labeled low quality by both patterns were strongly associated with negative video codes, making them the most obvious trigger for real-time adaptation. Intervals labeled high quality by both patterns were strongly associated with positive video codes, suggesting that any adaptations in place during these intervals should be removed, as they would no longer be needed. Given that all other combinations of quality labels output by the patterns skewed negative (or at least not positive) but each combination was fairly

infrequent, these intervals could trigger a “warn” state to give students an opportunity to improve on their own before triggering adaptive support.

6.3 CONCLUSION

This evaluation shows that, although the outputs of the touch patterns do vary when using Group Touch to distinguish among users, they still reflect overall quality of collaboration, demonstrating that Group Touch does make it possible to detect the quality of groups’ collaborative learning behaviors in real time. In Section 6.2, I propose an approach to triggering and releasing adaptations to encourage effective collaboration behavior based upon the quality labels output by the touch patterns. The following chapter explores what those adaptations should look like and why.

Chapter 7. DESIGN OF ADAPTATIONS TO ENCOURAGE EFFECTIVE COLLABORATION

The above chapters describe my approach to real-time detection of the quality of a group's collaborative learning process at a tabletop computer. As a group of students works on a task at the tabletop, the touch patterns described in Chapter 4 provide quality indicators every minute for the duration of the collaboration. The final stage of my dissertation work was to design ways for the computer to adapt in response to these quality indicators to scaffold more effective collaborative learning.

The work described in this chapter addressed research question 3: How can tabletop software adapt to a group's social regulation processes in order to support effective collaboration in real time? The design process involved consideration of the following:

1. What is the nature of the adaptations? This question covers exactly *what* adapts and *how* it adapts, including the extent of the constraints it imposes on how students interact with the computer. An example of a highly constraining adaptation would be restricting multi-touch interaction to a single person at a time to enforce turn-taking. An example of low-constraint adaptation would be simple prompts or on-screen messaging that provides the group with tips on how to be more effective. Also important is how students are made aware of the adaptations, if at all, and whether they have control over the adaptations, e.g., whether students are able to manually reverse an adaptation.
2. When and how are adaptations triggered? Although the general idea is that the computer will adapt in response to the quality indicators provided by the model, the exact timing of those adaptations requires careful consideration. A basic outline for the timing of adaptations was proposed at the end of the last chapter but further details need

consideration. For example, should the computer adapt immediately upon detecting low-quality collaboration, or should there be a waiting period in case the group is able to self-correct and render adaptation unnecessary? A related question is, once triggered, how long should an adaptation last—should it persist for the duration of the activity, or should it be reversed if quality indicators suggest that the group is collaborating effectively?

This project had two phases: (1) design and development of the software adaptations that would be triggered when poor quality collaboration was detected, described below; and (2) evaluation of the adaptations in a classroom setting, described in Chapter 8.

7.1 DESIGN PROCESS

The design process began with initial brainstorming of a variety of potential adaptations covering a range of approaches to scaffolding collaboration. Brainstorming was carried out in collaboration with another researcher and using the following dimensions as a guide: (1) *compliance* – ranging from voluntary to enforced; (2) *location* – from a single control or element to global interactions, e.g., affecting all touch input; (3) *desired outcome* – the specific characteristics of socially shared regulation or behaviors that the adaptation is intended to encourage; and (4) *risks* – any potential undesirable behavior that could result from the adaptation. These dimensions were chosen to ensure variety in terms of what would adapt within the software and the aspects of social regulation targeted, as well as to identify any potential unintended consequences that may make an otherwise promising adaptation ill-advised.

In the introduction to this chapter, I highlighted two key questions that needed to be addressed in this phase of the project. Designing with the aforementioned dimensions in mind helped me to address the first key question: what is the nature of the adaptations? As much as

possible, I intended to design adaptations that are, in principle, generalizable beyond the context of the software created for the study described in the next chapter. All adaptation ideas generated during initial brainstorming are described in Table 7-1.

Table 7-1. All adaptations designed during initial brainstorming. Note: this table spans multiple pages.

| | |
|--|---|
| <i>Voting</i> : A consensus must be reached among members of the group before an action is carried out. | |
| Compliance | Enforced |
| Location | Specific controls, e.g., buttons |
| Desired outcomes | Increased interdependence, equitable participation. |
| Risks | Could become tedious or possibly halt the activity if consensus can't be reached. |
| <i>Single-user touch interaction</i> : Ignoring touch input from “new” users until the student currently interacting with the screen passes control to another person—a virtual “talking stick.” | |
| Compliance | Enforced |
| Location | Global |
| Desired outcomes | Inclusion, active listening – could help ensure that each student has time to speak without being interrupted or shut down. |
| Risks | Control passes to a student with a propensity for other-regulation (one student directs and dominates the others). |
| <i>Prompts</i> : Messaging that temporarily halts all interaction to provide feedback on the collaboration and how to improve. | |
| Compliance | Voluntary |
| Location | Global |
| Desired outcomes | Increased awareness of effective collaboration. |
| Risks | Easily ignored. |

| | |
|--|-----------------------|
| <i>Group Awareness Icon</i> : An icon that changes color depending on how the group is collaborating. Similar in principle to several mirroring and metacognitive tools described in Section 3.3.3 (Järvelä & Hadwin, 2013). | |
| Compliance | Voluntary |
| Location | Global |
| Desired outcomes | Increased reflection. |
| Risks | Easily ignored. |

| | |
|--|--|
| <i>Heatmap of Touches</i> : A visualization showing how active students are in different areas of the screen. Similar in principle to several mirroring and metacognitive tools described in Section 3.3.3 (Järvelä & Hadwin, 2013). | |
| Compliance | Voluntary |
| Location | Global |
| Desired outcomes | Increased reflection, potential identification of dominant students or forgotten aspects of a task. |
| Risks | Active areas may be primarily determined by the design of the application and learning task rather than how students are collaborating, which could prove misleading for students. |

| | |
|--|--|
| <i>View-Only Regions</i> : Reduce the interactive areas of the screen to a small area. Inactive areas of the screen stay visible but cannot be manipulated. This draws upon the <i>Duplicate</i> tabletop coordination policy described by Morris et al. (2004), in which an on-screen element creates read-only duplicates of itself in response to a conflict between users. | |
| Compliance | Enforced |
| Location | Global and specific controls |
| Desired outcomes | Shared focus and turn-taking. |
| Risks | Confining interaction to a small area removes some of the advantages of the tabletop's large screen. It becomes easier for individual students to block other students or make it hard for them to see what is happening in the active area. |

| | |
|--|--|
| <i>Control Lockout:</i> Block or disable controls, reducing the number of active controls that students can interact with. | |
| Compliance | Enforced |
| Location | Specific controls e.g., buttons |
| Desired outcomes | Shared focus and increased interdependence. |
| Risks | Blocking the wrong controls could slow progress or even prevent the group from moving forward. This is particularly risky in truly open-ended collaborations where there is no defined “right” path to a solution. |

| | |
|--|--|
| <i>Focused Region:</i> This is a variation on View-Only Regions. In this version, a small region of the screen remains in focus while other areas are dimmed. The focused region acts as a lens through which students can act with on-screen objects. | |
| Compliance | Enforced |
| Location | Global |
| Desired outcomes | Shared focus and increased interdependence. |
| Risks | Care would have to be taken not to obscure needed content or objects otherwise progress on the task could be slowed or even halted. As with <i>view-only regions</i> , it becomes easier for individual students to block other students or make it hard for them to see what is happening in the active area. |

| | |
|---|---|
| <i>Directed Messages:</i> Messaging directed at particular students to provide feedback on their engagement in the collaboration. | |
| Compliance | Voluntary |
| Location | Global, though aimed at individual students |
| Desired outcomes | Increased self-reflection and awareness of individual students, more equitable participation, and behavioral engagement. |
| Risks | This may be beyond what is currently possible with Group Touch. Inaccurate feedback could cause more harm than good. There is also a risk that messaging could reward quantity of physical participation rather than quality of participation in the learning activity. |

| | |
|--|---|
| <u><i>Exaggerated Audio/Visual Feedback</i></u> : Exaggerate touch feedback provided by the interface, e.g., touching the screen causes a loud sound effect. | |
| Compliance | Voluntary |
| Location | Global |
| Desired outcomes | Increased awareness of physical actions and increased discussion before action. |
| Risks | Could quickly become tedious or prove distracting for students who struggle to maintain attention and engagement with an assigned task. |

| | |
|---|---|
| <u><i>Escalate to Authority</i></u> : Automatically call the teacher via messaging on his or her mobile device. | |
| Compliance | Enforced |
| Location | Global |
| Desired outcomes | The group gets attention and help from the teacher. |
| Risks | Because the teacher is summoned by the computer and not the students, the group may not understand why help was called and therefore may not reflect on their behavior. |

The 11 potential adaptations described above each come with potential benefits and drawbacks. I chose to select a subset of the potential adaptations to prototype and evaluate in a classroom study. I initially rejected the following adaptations: *single user touch interaction* and *directed messages*, because these adaptations would require sustained tracking of individual students in order to know exactly which student to target, which is beyond the capability of Group Touch; and *heatmap of touches*, because it may simply show which controls are heavily used rather than how a group is collaborating. In order to select a subset from the eight remaining adaptations, I considered when and how adaptations would be triggered.

The question of *when* adaptations are triggered is important because there are tradeoffs involved that could impact the growth that is possible in students' collaboration skills. For example, adaptations that trigger immediately upon detecting indicators of low-quality

collaboration may get students back on track faster, but they will likely become tedious and frustrating if they are triggered repeatedly (Martinez, 2001). Delaying the triggering of an adaptation, on the other hand, may leave room for students to learn to recognize collaboration problems and self-correct without the computer's assistance, which would be a positive outcome for this work.

The question of *how* adaptations are triggered encompasses the Ability-Based Design principle of transparency: "Interfaces may give users awareness of adaptive behaviors and what governs them, and the means to inspect, override, discard, revert, store, retrieve, preview, alter, or test those behaviors" (Wobbrock, 2018). Transparency has implications for both usability and learning. In practical usability terms, an adaptation represents a change in how users interact with the interface. Students will need to be made aware of what has changed simply so they can continue operating the software. From a learning point of view, students are more likely to become better collaborators if they have some awareness of what aspects of the collaboration triggered the adaptation, why the adaptation functions as it does in terms of the learning goals, and how they can be more effective (Wise, 2014). The students need to understand why the adaptation was triggered, how they can change their behavior to remove the intervention, and why changing their behavior in that way will help them to meet the learning goals. Therefore, I decided that the adaptations should be triggered as a sequence, where each new adaptation is added to those already in use but is only triggered if the collaboration quality does not improve. When collaboration quality does improve, all active adaptations are removed at once. Removing adaptations in response to improved collaboration is an example of "fading", which, in the scaffolding literature, refers to the process of removing supports that are no longer needed (Pea, 2004).

Layering new adaptations on top of already active adaptations, instead of triggering them in isolation, better supports Wise's (2014) principle of integration and process of grounding. A goal of the adaptations is to scaffold effective collaborative interactions at the tabletop by making it difficult for students to interact with the screen in ways that are deemed to be ineffective. However, no single adaptation covers the full range of desirable collaborative behaviors, all of which need to be adopted by the group if they are to be successful. For example, *control lockout* encourages shared focus by limiting the interactive controls to only those that are relevant in the current stage of the activity. However, this adaptation does nothing to promote interdependence among group members. *Voting*, on the other hand, promotes interdependence, but does not necessarily encourage shared focus. Combining the two adaptations would make it possible to encourage both shared focus and interdependence. Therefore, layering adaptations together can make it easier for students to ground themselves to the activity's expectations, because together the adaptations give a more complete picture of how to collaborate effectively than any isolated adaptation.

Adaptations that are removed when collaboration quality improves, whether they are triggered in isolation or as a sequence, can be said to support Wise's (2014) principle of agency as well as the scaffolding concept of fading (Pea, 2004). The sequence approach may help groups to adopt the behaviors that the adaptations are intended to promote and encourage students to recognize and correct problematic behaviors more quickly than if adaptations were triggered in isolation. After a group has experienced a sequence of adaptations, they will know what is coming next when the sequence is begun again. Because the students can anticipate the next step of the sequence, they also have the opportunity to preemptively adopt the behaviors that the next adaptation will enforce, removing the need for that adaptation and thereby preventing it from being

triggered. For example, if students know that the next adaptation to be triggered if their collaboration doesn't improve will be *control lockout*, they can aim to improve their collaboration by interacting only with objects that are directly relevant to the aspect of the task they are working on.

7.2 IMPLEMENTATION OF THE ADAPTATIONS

I chose to implement and pilot test the selected adaptations as a sequence with four stages. The sequence begins with the *group awareness icon*, a mirroring tool. The icon remains visible on screen for the duration of the activity and changes color according to the detected collaboration quality. The second stage, or second adaptation, is triggered when the icon turns red, representing poor collaboration quality, and no other adaptations are active. I chose to trigger *control lockout* as the first restrictive adaptation based on my observations in the previous classroom study, described in Chapter 4, that groups often jumped into an activity without taking time to plan or check their understanding of the instructions. This meant that they either failed to focus on anything at all or focused on aspects of the task that were not relevant or counterproductive. If collaboration doesn't improve after a two-minute interval of *control lockout*, the third adaptation, *voting*, is added. This adaptation prevents actions that affect the global state of the application, such as deleting work or marking an activity complete and moving on, from being carried out unless the majority of group members vote for it. The goal of this adaptation is to encourage discussion and to prevent one or two students from dominating the group. *Voting* can slow down progress considerably and become frustrating, which is why it is only triggered after students have already had the opportunity to reflect on and improve their collaboration. Finally, if collaboration still does not improve after a period of voting, the application will *escalate to authority* and call the teacher to intervene. There are many other ways to combine, order, and layer the four selected

adaptations in a sequence but this particular sequence was chosen initially as the adaptations progress in order from least to most restrictive or interventionist, giving students time to reflect on and change their collaboration behavior before increasing the level of restrictions imposed on the group.

In addition to deciding which adaptations to build and the order in which they should be triggered, I had to make a number of choices regarding exactly when an adaptation should be triggered, how it should be presented and explained to students, how long each adaptation should remain active, and when adaptations should be removed. Each of these choices could impact the outcome of the intervention.

The criteria I chose for triggering adaptations are shown in Figure 7-1. Based on the analysis (Section 6.2) of how collaboration quality labels produced by the touch patterns using Group Touch compared to video codes for collaboration quality, new adaptations were triggered when both touch patterns labeled a two-minute interval as low quality. Adaptations remained active until both touch patterns produced high quality labels for three consecutive intervals, or four minutes. I initially chose to release any active adaptations after three rather than one high quality interval to have extra confidence that the group had improved and could sustain its collaboration behavior.

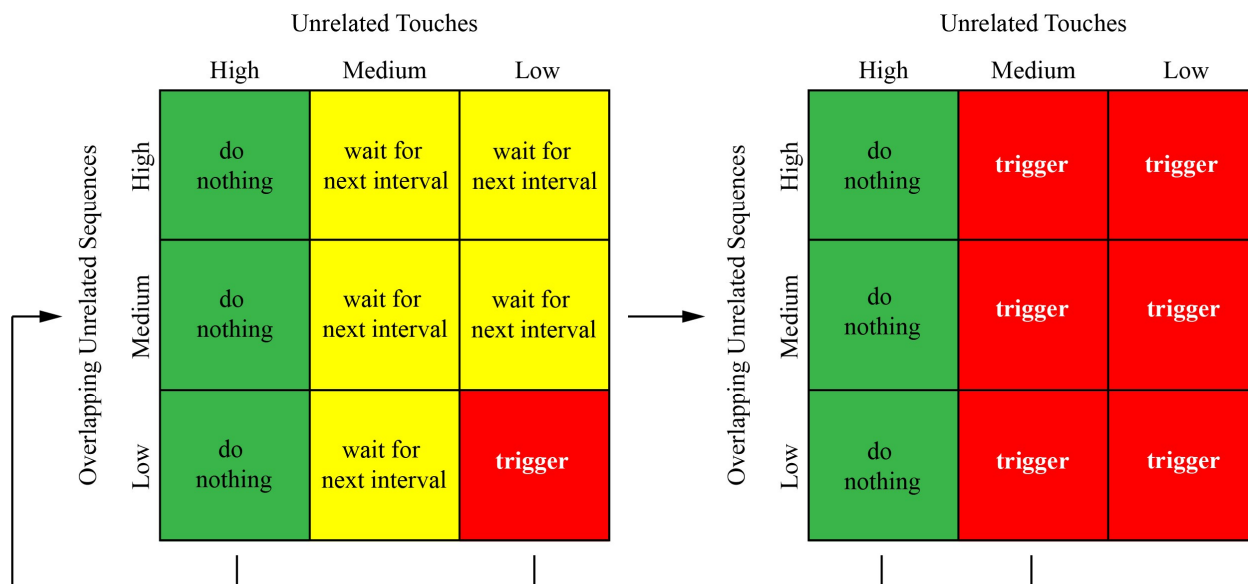


Figure 7-1. The criteria for triggering adaptations. Adaptations are triggered immediately if both touch patterns label a single interval low quality (left-red) or two consecutive intervals are labeled low or medium quality by Unrelated Touches (left-yellow to right-red). The colors in this figure indicate the color of the *group awareness icons* for each possible combination of quality labels produced by the touch patterns.

All other combinations of touch pattern labels (e.g., one pattern labeled an interval high quality and the other labeled it medium quality) maintained the status quo in the application for at least one interval, continuing any active adaptations and not triggering any new adaptations. Exceptions to this were intervals where Unrelated Touches labeled two consecutive intervals as low or medium-quality. These intervals were treated as if both patterns had labeled the interval as low quality and the next adaptation in the sequence was triggered. The reasoning behind this decision was that the analysis in Section 6.2 showed that intervals of Unrelated Touches labeled low or medium-quality but Overlapping Unrelated Sequences labeled medium or high-quality were strongly skewed negative (see Table 6-2), and Unrelated Touches labels were typically more accurate than those of Overlapping Unrelated Sequences (see Table 6-1). I opted to wait for two consecutive intervals that met these criteria before triggering a new adaptation because they had been relatively few in number in the previous study. Consecutive intervals labeled low- or

medium-quality by Unrelated Touches increased my confidence that sustained poor quality collaboration was occurring.

Chapter 8. EVALUATION OF SOFTWARE ADAPTATIONS TO SUPPORT COLLABORATIVE LEARNING

The selected adaptations—*group awareness icon*, *control lockout*, *voting*, and *escalate to authority*—were prototyped and evaluated in a classroom setting with high school students. The purpose of this evaluation was to answer research question 4: Do the adaptations encourage effective collaboration? Determining effectiveness involved answering the following sub-questions:

1. Are the adaptations triggered appropriately? This served as an additional check on the modeling work already completed. Although the touch patterns were refined over the course of two separate studies, it was important to validate them further. The success of an adaptation is judged based on how a group of students responds to it. If adaptations are triggered falsely, there is a risk that they could create the problem they are trying to solve by frustrating or annoying students to the point that they lose motivation to engage with the learning activity. Conversely, identifying missed opportunities for adaptation—cases when collaboration issues go undetected by the computer—helps to determine the limitations of this work.
2. How do students respond when an adaptation is triggered? At a minimum, short-term change in the group's behavior in response to an adaptation would be desirable, such that they show more characteristics of socially shared regulation. Also desirable would be for students to take away an understanding of *why* the adaptations were triggered and how they might approach collaboration differently in the future. The most positive outcome would be a long-term decrease in the frequency at which adaptations are

triggered, signaling improved collaboration skills. However, tracking long-term changes would require a longer study than is the scope of this dissertation.

The evaluation showed that the adaptations were successful at deterring disruptive behavior, enabling those students who were engaged to make progress on the task and reducing the length of periods of low-quality collaboration. The adaptations could not, however, re-engage disengaged students.

8.1.1 *Participants*

Eleven high school students (9 male, 2 female) participated in this study. The students were 10th and 11th graders enrolled in a six-week course on beginner video game development offered as part of a college preparation program serving low income students who would be the first generation in their family to earn a four-year college degree. The students in the program attended classes full time for six weeks during the summer. All students in the program had to take several required courses depending on their grade but they could also choose from several electives. The course used in this study was one of the elective options. The focus of the course was educational games, and students were asked to design a game to raise awareness of an environmental issue: snow leopard conservation. In the first part of the course, the students learned about snow leopards and ongoing efforts to protect the species. In the second part of the course, the students learned the basics of video game development using Unity. All but one of the students were new to writing code and all were new to video game development and Unity. In the third and final part of the course, the students worked on their game projects.

My role in the course was co-teacher as well as researcher. I taught the aspects of the course related to game development and introduced the tabletop sessions. Another graduate student taught

the parts of the course focused on environmental issues and, while I was engaged with a group at the tabletop, led class activities for the other students.

The students were randomly assigned to three small groups for several collaborative learning activities in the first and third parts of the course—two groups of four and one group of three. The students stayed in the same groups for all collaborative activities.

8.1.2 *Apparatus*

The groups completed three activities across four sessions at a Microsoft Surface Hub, formerly known as the PPI. The Surface Hub differs from the Microsoft PixelSense used in the previous studies in several ways. It has a 55” screen, considerably larger than the PixelSense’s 40” screen; it runs any Windows 10 application rather than requiring a specific API on Windows 7; and it has much improved touch detection.

I built three distinct applications for the students to use at the Surface Hub. The applications were designed to be used alongside other classroom activities and resources, and each application addressed the specific learning objectives for the class session in which it was used. All applications were created using Unity with the TouchScript⁶ plugin to enable the full range of standard multi-touch gestures.

Snow Leopards 101, shown in Figure 8-1, was used during the first phase of the course. It is an introduction to snow leopards and their ecosystem adapted from an existing non-tabletop curriculum⁷ (Facing the Future & Snow Leopard Trust, 2009). The activity begins with a jigsaw puzzle (Figure 8-1, left)—this is primarily a warm-up exercise to get students used to using the tabletop. Next, students are given information about four big cat species and their ecosystems

⁶ Information about the plugin is available here: <https://www.assetstore.unity3d.com/en/#!/content/7394>

⁷ The curriculum can be downloaded here: <https://www.snowleopard.org/snow-leopard-facts/resources/>

(Figure 8-1, center), which they use to complete graphic organizers focused on how each species is uniquely adapted to its ecosystem (Figure 8-1, right).



Figure 8-1. Screenshots of the Snow Leopards 101 application used in the baseline data collection settings. The activity began with a warm-up puzzle (left), then students explored how the four different species of big cat were uniquely adapted to their habitat (center). Finally, students completed a graphic organizer describing each cat’s physical characteristics, habitat characteristics, and unique adaptations (right).

Help a Scientist, shown in Figure 8-2, was also used during the first phase of the course. In this application, students take on the role of scientists studying wild snow leopards in Mongolia. The application home screen (Figure 8-2, left) is an aerial view of the scientists’ real-world study area in Mongolia, which shows the locations of 32 research cameras that students can visit to view and label the species photographed (Figure 8-2, right). Students also explore a 3D simulation of the study area (Figure 8-2, center), using videogame-style controls to walk from camera to camera to find additional “hidden” content, such as short video clips taken by the scientists.



Figure 8-2. Screenshots of the Help a Scientist application. From the home screen (left), students can see an aerial view the scientists' study area with overlays of snow leopard home ranges. Upon selecting a camera, students are taken to a 3D simulation of the study area (center). Selecting a camera from the 3D view gives students access to the photos taken by the camera (right). The goal of the activity is to label the species captured in the photos taken by each camera.

Game Challenge, shown in Figure 8-3, was used during the third part of the course. This application was used to extend the students' knowledge of game development basics covered in the second part of the course. The application featured a partially completed game world (Figure 8-3, left) that students could build upon (Figure 8-3, center) and adjust to create different outcomes while learning important and often complex concepts used in Unity, such as the use of its built-in physics engine to apply realistic forces to game objects. Students could test their edits in "play mode" (Figure 8-3, right), which simulated game play.

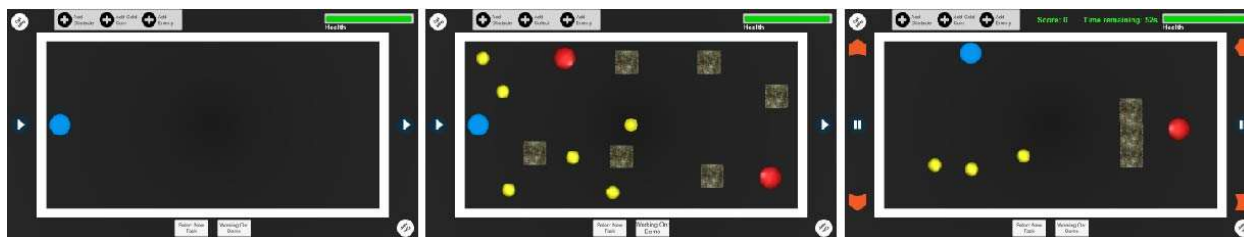


Figure 8-3. Screenshots of the Game Challenge application, a partially-built game world that students can edit to bring about particular outcomes. Beginning with a mostly empty play field with a Player (the blue circle) and four containing walls (left), students can add collectable objects, obstacles, and enemies, all of which have editable settings (center). Students test their edits in play mode (right).

All three applications implemented the process of detecting collaboration quality described in the previous chapter (see Figure 6-1), and Group Touch was used to assign touches to groups.

In Chapter 6, each of the touch patterns used to detect quality of collaboration was found to work best with a different value of Group Touch's probability threshold. Therefore, two parallel sets of groups were maintained—one set of groups created by Group Touch using a probability threshold of 0.8 for the *Unrelated Touches* pattern, and another set created using a probability threshold of 0.5 for the *Overlapping Unrelated Sequences* pattern. Next, sequences of touches representing “purposeful” actions were extracted from each set of touch groups using the rules described in Chapter 4 (see Table 4-1). The touch patterns, also described in Chapter 4 (see Table 4-3), then used the sequences to derive the collaboration quality for the preceding two-minute interval. The collaboration quality was measured every minute, so the intervals overlapped.

The Surface Hub is primarily intended to be used in a vertical orientation, e.g., hung on a wall, but for this study, it was placed flat on a table in the middle of a small breakout room across a hallway from the main classroom (Figure 8-4). A wide-angle video camera was mounted on a wall so that students could freely move around the tabletop. The camera was angled toward the screen so that it could capture every touch and the interactions among the group members. Ten group sessions at the tabletop were video recorded for this study and the computer logged every touch. Two to three log files were produced for each session: a log of all raw touches, a log of touch pattern labels for each interval, and, for group sessions in which adaptations were used, a log of the adaptations present in each interval.



Figure 8-4. Group 3 uses the tabletop. The computer was set up in a small breakout room across a hallway from the main classroom. Students had enough room to freely move around the tabletop. A single wide-angle camera was pointed at the screen. The full shot included students faces—the view is cut off in this screenshot to obscure students' identity.

8.1.3 Study Design

The study was split into three phases: (1) baseline data collection, which enabled me to uncover differences in each group's working style and ability to collaborate effectively; (2) a pilot test of an initial implementation of adaptations, which resulted in a revised implementation; and (3) an experimental evaluation of the final implementation.

To collect baseline data on differences between the three groups of students in terms of how they collaborate, all groups worked on the *Snow Leopards 101* activity (Figure 8-1) without

any intervention from the adaptations. Students also individually completed a survey of their attitudes toward group work (Appendix B). The survey was adapted from an existing instrument, the Student Attitudes toward Group Environments (SAGE) (Kouros & Abrami, 2006). SAGE is an empirically validated questionnaire aimed at high school and junior college students. SAGE's 54 Likert-type questions cover the quality of products and processes of group work, peer support, student interdependence, and frustrations with group members. The majority of SAGE's questions were relevant to this work, but those that were less relevant were omitted to shorten the survey. See Appendix B for the modified list of questions. The survey directions were also modified from asking students to reflect on their experiences in a class they are taking to reflecting on their experience of being in groups in general.

In phase 2, the adaptations were piloted with Group 1 using the *Help a Scientist* application (Figure 8-2), in which students take on the role of snow leopard scientists and analyze photographic data. Groups 2 and 3 completed the same activity with the same application but the adaptations were not made available. Group 1 were selected for the pilot because they had already been assigned to the experimental condition and would have adaptations available in their remaining tabletop sessions. There was, of course, a risk that their encounters with the adaptations in the pilot session would influence their interactions with adaptations in the experimental sessions but, with two and a half weeks between the pilot and the first experimental sessions, I expected that their memories of the adaptations would fade. Additionally, the first adaptation triggered in the sequence was different in later sessions—a change made due to my observations in the pilot session, described in full below—reducing the risks associated with familiarity.

In the introduction to the activity in the pilot session, I explained to Group 1 that the computer was tracking how they interacted with it in order to help them work together effectively,

if needed. I explained that the color of the *group awareness icons* (Figure 8-5) updated every minute to give them feedback on how they were collaborating. I told them that, if the icons turned red or stayed amber for several minutes, the computer would block certain controls or ask them to vote before carrying out certain actions. I also told them that the adaptations would go away if the icons turned green for four minutes and that they could keep the icons green by maintaining a shared focus on the assigned task, discussing the content, and listening to each other's ideas. Groups 2 and 3 also used the *Help a Scientist* application but without the adaptations. The video recorded in Group 1's pilot session allowed me to get a quick sense of how the adaptations were received by the students and to make adjustments as needed. The adjustments made as a result of the pilot are discussed in the next section.

In phase 3, about two and a half weeks after the pilot session, the groups worked on the *Game Challenge* activity (Figure 8-3) across two sessions. For this phase, each group was assigned to a different condition: Group 1 was assigned to the experimental condition—this group used the tabletop with the adaptations for both sessions with the third application; Group 2 was assigned to the control condition—this group used the tabletop without the adaptations for both sessions; and Group 3 was assigned to a split condition—this group used the tabletop without the adaptations for the first session, then with the adaptations for the second session. The control and experimental conditions allowed for a between-groups comparison, and the split condition allowed for within-group comparison.

Ideally, had there been enough students in the class to make a fourth group, a second split group could have experienced the adaptations in the reverse order—with adaptations in the first experimental session and without adaptations in the second. Without a fourth group, however, I had to pick one within-group ordering to test—either (1) using the tabletop *without* adaptations in

the first session and *with* adaptations in the second session or (2) *with* adaptations in the first session and *without* adaptations in the second. The first option would allow me to assess if the presence of the adaptations could lead to any immediate, short-term improvements in collaboration and the second option would allow me to assess the possibility of longer term improvements. Ultimately, I chose the first option because of the short duration of the study and the fact that the experimental sessions only included two tabletop sessions for each group. It seemed unrealistic to expect that a single exposure to the adaptations would lead to lasting improvements in a group's ability to work together effectively. Therefore, I chose the option that would support analysis of any shorter-term effects on a group's collaboration skills.

Randomly assigning students to groups and randomly assigning groups to conditions allowed me to isolate the effects of the adaptations to some degree. However, because there was only one group per condition, differences between the groups in terms of their ability to collaborate effectively as well as the individual students' motivation (or lack thereof) to engage with the activities were confounding factors that could not be accounted for in a study of this size. The data collected in the baseline phase of the study helped to illuminate those differences.

Group 1, the experimental group, was given a brief reminder about the adaptations that reflected changes made after the pilot (described in the next section below). At the start of Group 3's (split group) session with the adaptations, they were given the same introduction that the Group 1 had received in the pilot session, with a slight adjustment to reflect the changes made to the adaptations between Phase 2 & 3 of the study.

All but one of the tabletop sessions lasted around 25 minutes. The one exception was Group 3's last session, which ended after 15 minutes because they completed the assigned task.

8.1.4 *Pilot Test*

The adaptations selected for a pilot were *group awareness icon* (Figure 8-5), *control lockout* (Figure 8-6), *voting* (Figure 8-7), and *escalate to authority* (Figure 8-8). The adaptations were triggered in a sequence based on the quality labels output by the touch patterns identified in Chapter 4—Unrelated Touches and Overlapping Unrelated Sequences (Figure 7-1). Each of the decisions made as part of the design process—which adaptations to build, the order in which they should be triggered, when they should be triggered, and how they should be presented to students—could impact the outcome of the intervention. The pilot session allowed me to test out my initial decisions and quickly make changes as needed before the experimental sessions.

The first adaptation in the sequence was the *group awareness icon* (Figure 8-5). Unlike the subsequent adaptations, this adaptation was permanently displayed on screen. The icon was displayed in two of the corners of the screen so that it would be visible to all group members without obstructing the work area. The background of the icon changed color to reflect the detected quality of the most recent interval. At the start of a group session, the background color was gray until the first interval was recorded. The background turned gray again if the computer was not touched for an entire interval, for example, when a group left the tabletop with the application running. The icon turned green when both touch patterns labeled an interval high quality and red when both touch patterns labeled an interval low quality. For all other combinations of collaboration quality output by the touch patterns, the icon turned yellow, unless *Unrelated Touches* had been low- or medium-quality for two or more consecutive intervals, which would trigger the next adaptation as described above. In these cases, the icon turned red.



Figure 8-5 The *group awareness icon*. The background changes color (left) to reflect the detected quality of collaboration. Green represents high quality collaboration, yellow represents medium quality collaboration, and red represents low quality collaboration. The icon turns gray when no touches are detected. The screenshot on the right shows the placement of the icons in the software used in this study—top left and bottom right of the screen, next to the “Call help” button that students could use to summon the teacher.

The second adaptation in the sequence was *control lockout* (Figure 8-6). When *control lockout* is active, certain controls are disabled and covered with an icon to alert students that the control is locked. The purpose of *control lockout* is to encourage focus on the assigned task by reducing the number of active controls to only those that are necessary for the specific part of the task that students are working on. Before *control lockout* is triggered, a warning pops up on screen to let students know what is about to change in the application and why. The warning states, “Some buttons and controls will be temporarily disabled to help you to stay focused on the task.” In the pilot, the warning remained on screen for 30 seconds and all interaction was blocked for the duration that it was on screen.

Although in principle, this adaptation could be application-independent, its implementation was application-specific. In this work, controls were blocked based on the learning goals of the activity and which aspect of the activity students were working on when the adaptation was

triggered. Figure 8-6 shows *control lockout* active in the application in which students take on the role of snow leopard scientists and analyze photographic data. The white rectangle partially obscured by the photo is a calendar that shows all of the photos available for students to view. Clicking on a colored circle in the calendar brings up the photos taken at that time—the blue rectangular object. The goal of the activity is for students to identify the species they see in the photos taken in the field. By default, students are able to open up and interact with as many photos as they want. However, for the purpose of the activity, the group is supposed to work through a single set of photos at a time and reach consensus on the identification of the species they see. The ability to open more than one photo set can be helpful for comparison purposes when dealing with hard-to-identify species. For example, wolves and foxes often look very similar in the nighttime photos collected by the scientists. The ability to open more than one set could, however, also prove to be a distraction, allowing individual students to work independently from the rest of the group or be disruptive. Therefore, if a photo set is open when *control lockout* is activated, the calendar is blocked to prevent the group from opening additional, unnecessary photo sets until they are finished with the current set. If multiple photo sets are open when *control lockout* is activated, all sets but the topmost photo set are blocked. The field guide is never blocked, as it is always relevant and necessary for the task.



Figure 8-6. The Calendar control is blocked by control lockout.

The next adaptation in the sequence was *voting* (Figure 8-7). When *voting* is active, students are required to vote on any action that changes the global state of the application or that could impact their progress on the activity, e.g., moving to a new section of the application without saving changes. The action is carried out if the majority of students agree to it, or stopped if the majority disagrees. If there is a stalemate, the group is asked to vote again. As with *control lockout*, a warning pops up on screen before *voting* is activated to inform students of what is about to change. The warning text reads, “You will temporarily be asked to vote in order to carry out certain actions such as changing activity or closing windows,” and blocks all interaction for 30 seconds before activating *voting*.

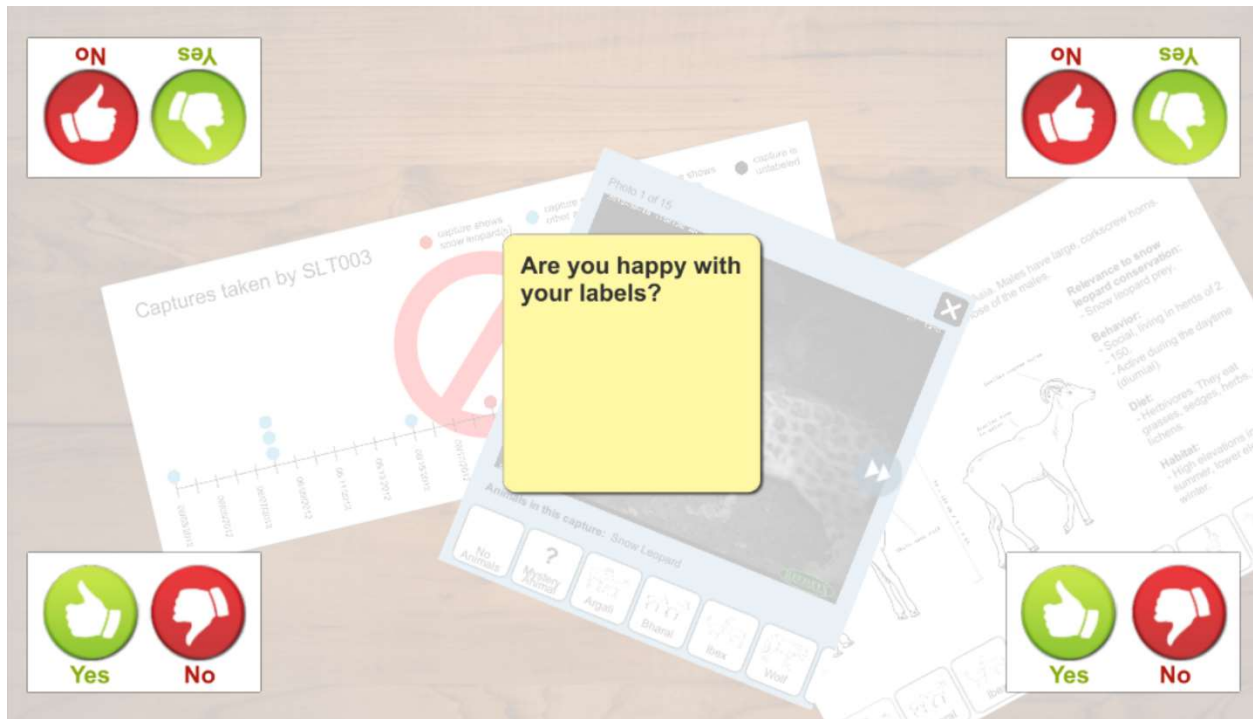


Figure 8-7. The voting adaptation requires all students to vote on global changes to the application.

The purpose of voting is to encourage students to discuss the action to ensure that all group members agree to it. When students attempt an action that requires a vote, a message pops up on screen and remains until the outcome has been decided. The message shown varies depending on the action that has been attempted. In Figure 8-7, a student has opted to close a photo set, which will save any species labels applied to the photo set as the group's answer. The vote message asks, "Are you happy with your labels?"

The final step in the sequence is *escalate to authority* (Figure 8-8). Unlike the previous adaptations, this adaptation is silent from the students' perspective. When triggered, the tabletop sends an alert that the group may be struggling to collaborate effectively to the teacher's mobile device or computer. The group is not notified that the teacher is about to be called. I implemented

the alert using the Slack messaging service's API⁸, which allows for free and easy-to-use instant messaging over Wi-Fi. The study setting meant that groups using the tabletop were in a separate room from the rest of the class. To enable the group to easily ask the teacher for help from the other room, I also added a "Call Help" button to the application, positioned next to the group awareness icon. Like the *escalate to authority* adaptation, pressing the button sent a Slack message to the teacher. The message text noted whether the group had requested help or an intervention was needed.



Figure 8-8. *Escalate to authority* sends an alert to the teacher's mobile device via the Slack API. This intervention is silent from the students' point of view—they are not notified that the teacher has been summoned and there are not visible changes to the UI.

The adaptations were piloted with Group 1, a group of four boys randomly assigned to the experimental condition, as they worked on the *Help a Scientist* application in the first phase of the course. The objectives of the activity were to become familiar with the species found in the snow leopard's Mongolian habitat and to learn about the day-to-day field work of conservation biologists. In the larger course context, this activity was part of the preparation for the final project, which was to build a videogame to raise awareness of snow leopard conservation issues.

⁸ <https://api.slack.com/>

To evaluate the piloted adaptations, I reviewed the video of the session and the log files. Given that the purpose of the pilot was to get feedback on a number of design choices quickly enough to make changes in time for the experimental portion of the study, I did not do a full in-depth analysis of the video at this point. Instead, I observed what the group was doing when adaptations were triggered and how they responded. This informal analysis was verified by a formal analysis of the pilot session once the study was over (see Section 7.2.6). Key questions that I hoped to answer with the pilot sessions, at least informally, were:

- 1) Were triggered adaptations justified—were students actually struggling to collaborate effectively when an adaptation was triggered?
- 2) Did the group read the warning messages that popped up on screen?
- 3) Were warning messages on-screen for too much, too little, or the right amount of time?
- 4) Did the group appear to understand what the adaptations were doing and why?
- 5) Did the adaptations cause frustration or prove too distracting?
- 6) Was positive collaboration recognized?
- 7) Were students able to get back on track and, if so, did the length of time that the group had to maintain high quality collaboration seem reasonable?

After reviewing the video, specifically the verbal and physical interactions between the students, I determined that the adaptations did appear to be triggered appropriately and that they did appear to get the group to collaborate more effectively. However, the positive behavior change that occurred seemed to be a result of coercion—the adaptations proved so annoying that they forced the students to improve their working style without encouraging reflection. Although it could be said that the adaptations were successful because collaboration did in fact improve, I believed that the improvement was too superficial to be desirable. Additionally, annoying the

students into working together effectively may have yielded positive changes on the first exposure, but with repeated exposure the students' frustration would likely turn to resentment and the adaptations may end up worsening collaboration over time.

The video did, however, suggest adjustments to the implementation that would improve the students' experience while encouraging positive collaboration behaviors. The main lesson learned was that it was too difficult to get the adaptations removed—the students had to sustain high quality collaboration for four minutes before restrictions were removed. Additionally, the students did read the initial warning that appeared with each new adaptation, but they felt that they were already doing as the instructions asked and were not able to see how they should change their behavior. As a result of these observations, I reduced the length of time a group has to sustain high-quality collaboration in order to remove the adaptations, and I added an additional adaptation, *prompt*, before *control lockout* in the sequence. This additional adaptation provides students with more information on what is expected of them and gives them the opportunity to self-correct before the more restrictive adaptations are triggered.

The sequence of adaptations was triggered twice during the pilot: at 8 minutes into the activity and again at 23 minutes. In both cases, the sequence progressed from *control lockout* to *voting* and was canceled before *escalate to authority* was triggered. This result means that low-quality collaboration was still detected after at least one interval of *control lockout*, causing *voting* to be triggered. The group was able to sustain good quality collaboration for at least three intervals while *voting* was active, so the adaptations were removed without messaging the teacher. The restrictive adaptations (*control lockout* and *voting*) were on screen for a total of 13 minutes—9 minutes in the first sequence and 4 minutes in the second sequence.

The first instance of *control lockout* was triggered after two intervals labeled poor-quality by the Unrelated Touches pattern and high-quality by the Overlapping Unrelated Sequences pattern. During these intervals, the students were mostly on task but had taken a divide-and-conquer approach, with multiple students working independently in silence with few exceptions. Therefore, I felt that *control lockout* had triggered appropriately. However, the group had just begun to engage collaboratively when the adaptation was activated, interrupting their discussion. Indeed, after reading the warning associated with *control lockout*, one student exclaimed, “ ‘...to help us stay focused,’ we *are* focused!” One student pressed the “Call help” button to ask the teacher, “Why is this red?”, referring to the *group awareness icon*. As soon as he asked, the icon turned green, reflecting their improved collaboration. The student seemed to be satisfied by this, although the perceived delay may have caused initial frustration.

The next two intervals assessed after *control lockout* was activated were labeled high quality by both touch patterns, which appeared consistent with what was happening in the video—a relatively deep, whole-group discussion in which they reasoned about the species they were seeing in the photos. However, the collaboration began to degrade as students expressed frustration that they were still locked out of various controls after several minutes—in this instance, the implementation of the adaptations potentially caused collaboration problems. Because of the degraded collaboration, the next adaptation, *voting*, was activated, prompting much swearing and sarcasm from the students. It took the students approximately one interval to adjust to the new adaptation and they were then able to maintain a shared focus for long enough to get the adaptations removed. However, there was relatively little discussion, and the group appeared to be just going through the motions to complete the activity. Once the adaptations were removed, content-related discussion picked up again although one student’s attention wandered to his phone.

As he was not interacting with the tabletop, however, this would not have been picked up by the touch patterns.

The second sequence was triggered after an interval labeled poor quality by both patterns, in which the students decided to tap through the photos on screen as quickly as possible and stop discussing the content. This behavior was contrary to the goals of the activity; therefore, I believed *control lockout* to have triggered appropriately. The group's behavior did not change immediately after *control lockout* was triggered, and the next interval assessed was also labeled poor quality by both patterns, causing *voting* to be activated just one minute after *control lockout*. This time, the students were able to immediately manifest and sustain the type of interactions that meet both touch patterns' standards for high quality collaboration, causing the adaptations to be removed in the minimum amount of time, 4 minutes. As with the first sequence, although the group maintained a shared focus during the "high quality" intervals, there was little discussion or deep engagement with the content.

After my informal analysis of the pilot session, I felt that the adaptations showed some promise, but there was much room for improvement. The most obvious drawback to the approach taken in the pilot was that it took too long for adaptations to be removed once the group got back on track. Although the students did receive some feedback on their improved collaboration when the *group awareness icon* turned green, this did not seem satisfactory while the adaptations remained active. Additionally, the pilot approach appeared unbalanced in that adaptations could be triggered after as little as two minutes of poor quality collaboration but double that amount of high quality collaboration was required to get them removed. This imbalance may have caused the adaptations to appear excessively punitive to the students. Therefore, for the next phase of the

study, I reduced the duration of high quality collaboration required to remove active adaptations from four minutes to two.

The time that warning messages remained on-screen before activating *control lockout* and *voting* also appeared to be too long at 30 seconds. The group read and digested the messages in a much shorter time so the warnings caused unnecessary delay to the group's progress. Therefore, I decided to reduce the time that messages were displayed from 30 to 20 seconds. It is worth noting, however, that although the shorter time might be appropriate for students with strong reading skills, such as the students in this study, extended time may be necessary for students with reading difficulties.

Finally, the first time the sequence was triggered, the students appeared to lack sufficient understanding of what was happening and why, despite the fact that these points were covered in the verbal introduction to the activity. Wise's (2014) process of *grounding*, which is important to the overall success of a learning analytics intervention, had not taken place and the group did not have a clear idea of what was expected from them in terms of collaboration. To remedy this issue, I decided to reinforce the information given in the introduction as a *prompt*, which would become the first intervention to be triggered after a collaboration breakdown.

The prompt, pictured in Figure 8-9, would appear on screen when the touch patterns detected poor quality collaboration and no other adaptations were on screen—the start of a new sequence. The text was longer than the warnings used with the *control lockout* and *voting* adaptations, so the prompt stayed on screen blocking interaction for 30 seconds. The text of the prompt focused on actionable behaviors, with the intention of helping students connect the concept of effective collaboration discussed verbally at several points in the course, with how they were actually interacting with each other and the tabletop in the moment. Tips 1 and 3 and the button

linking to the activity instructions referenced what I had observed to be a common blocker to collaboration for the students in the previous classroom study (Chapter 4), as well as the students in this study—a tendency to dive into a task without much care for the specific instructions or learning goals. My hope was that if this prompt appeared just as collaboration was starting to break down, likely due at least in part to a lack of task planning, it may help to encourage groups to be a little more reflective about how their interactions with the tabletop aligned with the assigned task.

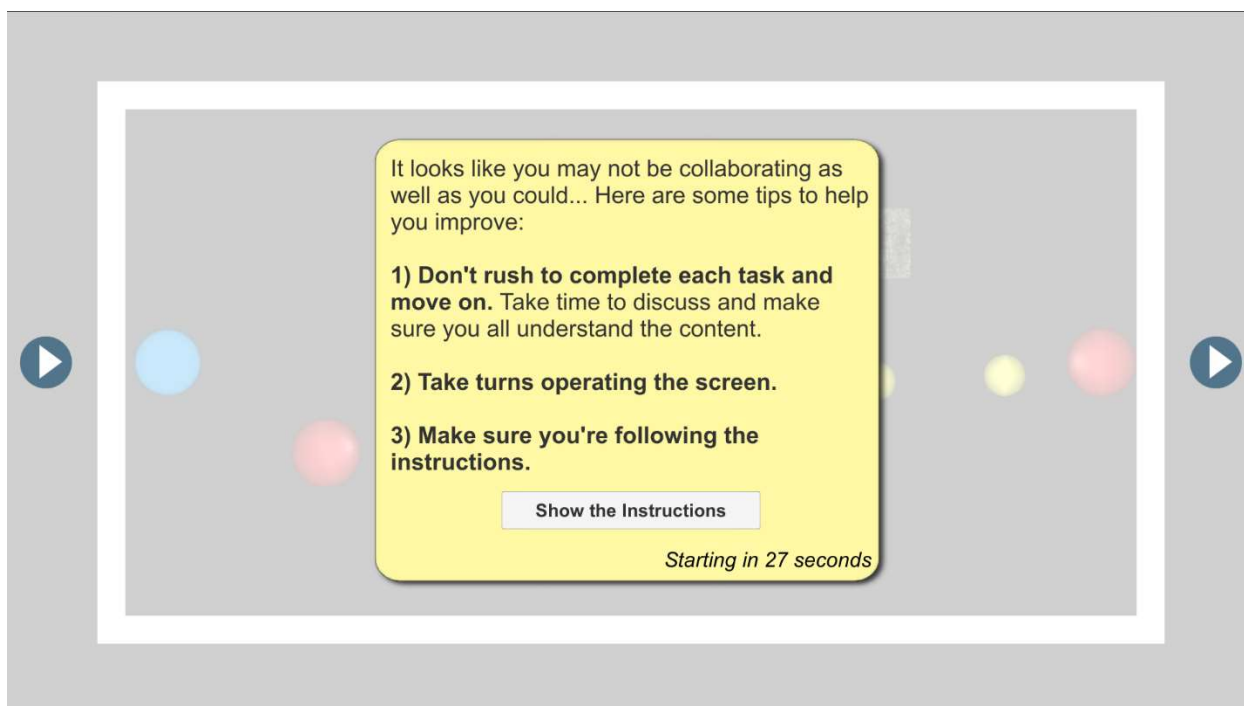


Figure 8-9. A prompt encourages group members to revisit task instructions, listen to each other, and coordinate their actions.

The prompt also triggered a temporary amnesty, meaning that the quality label of the first interval to be assessed after the prompt appeared (one minute later) would be ignored to give the group a fresh start and a full two minutes to attempt to regulate their interactions before *control lockout*, now the second intervention in the sequence, could be triggered. I felt that it was important to give the students a chance to self-correct before proceeding with the sequence of adaptations. The only way for a group to break out of the sequence completely, however, would be for both

touch patterns to label a complete two-minute interval as high quality. Otherwise, the next step in the sequence would be triggered.

I decided to leave the rest of the sequence the same as it had been for the pilot, both in terms of the order and the timing of the adaptations. *Voting* remained the last adaptation before summoning a teacher because, although it seemed to temporarily stop discussion in the pilot, it also appeared to force a reset of the collaboration, coercing students to engage with the assigned task and focus on what they were actually supposed to be doing. For both sequences of adaptations that occurred in the pilot, collaboration substantially improved after students were required to vote on collective actions for a period of time. I hoped that reducing the time that students had to sustain “high quality” collaboration in order to remove the adaptations would mitigate the negative effects of *voting* while retaining the benefits of the reset that it seemed to encourage.

8.1.5 *Data Analysis*

Formal data analysis of all study sessions, including the pilot session, began by coding the videos for social regulation, using the same codes that I used in the classroom study described in Chapter 4 (Table 4-2). The purpose of the video analysis was (1) to check that the touch patterns detected by the computer matched what was happening in the videos; and (2) to learn how the adaptations affected collaboration. For each of the videos recorded in this study, I used the touch pattern log files to produce an edited version with annotations indicating the start of each interval recorded by the computer. This made aligning the video coding with the touch log files straightforward.

The bulk of the coding was carried out by another person to reduce the risk of bias, an Education doctoral student who was unfamiliar with the adaptations being evaluated in this study. We met prior to the video analysis to discuss the concepts covered by the codes and their definitions. He was also briefed on the learning goals of each tabletop activity as well as the various

aspects of each application. To establish inter-rater reliability, we each independently coded a session from the first phase of the study, the baseline data collection. Codes were applied to episodes, as in the previous classroom study, and each episode could contain multiple codes. We then met to discuss disagreements episode by episode. For the social regulation codes (see Table 4-2), the quality of the process is extremely important. Therefore, for those processes, high- and low-quality instances were treated as separate codes. We agreed not to code any of the social regulation processes as medium quality unless absolutely necessary because occurrences of these codes were extremely rare in the previous study. Additionally, where multiple episodes were happening at once, such as when two students were engaged in task work while the other group members were having a side conversation away from the computer, we would code the interactions taking place at the computer and separately note the codes for verbal interactions taking place away from the computer. The reasoning behind this decision is that the computer is only able to detect interactions in which it is involved. However, interactions that take place away from the computer are important aspects of the whole collaboration that should be noted even if they are undetectable. Once we reached consensus for all codes in the first session, we independently coded the remaining two sessions from the first phase of the study.

Inter-rater reliability was to be calculated at the interval level from the second and third session of phase 1 of data collection. For touch pattern detection, the computer assesses overlapping two-minute intervals, meaning every minute of collaboration is assessed twice. As this double-counting could distort the reliability calculation, intervals were split in two and reliability was calculated using the codes for each minute of video data. Both coders recorded the occurrence of each code by interval, using 1 to indicate the presence of a code and 0 to indicate

that code did not occur during the interval. Inter-rater reliability was therefore calculated per code and is shown in Table 8-1.

Table 8-1. Inter-rater reliability statistics by code.

| Code | Occurrences | % of intervals | % agreement | Cohen's kappa |
|--------------------------------------|-------------|----------------|-------------|---------------|
| Task planning – high quality | 6 | 17.65 | 100.00 | 1.00 |
| Task planning – low quality | 0 | 0.00 | 100.00 | - |
| Content planning – high quality | 1 | 2.94 | 100.00 | 1.00 |
| Content planning – low quality | 1 | 2.94 | 97.06 | 0.00 |
| Content monitoring – high quality | 16 | 47.06 | 85.29 | 0.70 |
| Content monitoring – low quality | 23 | 67.65 | 91.18 | 0.80 |
| Monitoring the plan – high quality | 7 | 20.56 | 97.06 | 0.91 |
| Monitoring the plan – low quality | 4 | 11.76 | 100.00 | 1.00 |
| Monitoring progress – high quality | 17 | 50.00 | 91.18 | 0.82 |
| Monitoring progress – low quality | 10 | 29.41 | 91.20 | 0.77 |
| Behavioral engagement – high quality | 7 | 20.56 | 94.12 | 0.80 |
| Behavioral engagement – low quality | 0 | 0.00 | 100.00 | - |
| Off task | 7 | 20.56 | 97.06 | 0.91 |
| Non-collaborative interactions | 20 | 58.82 | 82.40 | 0.65 |
| Task work | 7 | 20.56 | 97.06 | 0.91 |
| Software conflicts | 0 | 0.00 | 100.00 | - |

The majority of codes had a Cohen's kappa above 0.61, typically considered “substantial” agreement, with several codes above 0.81, or “almost perfect” agreement (Landis & Koch, 1977). One exception was *content planning – low quality*, which had a kappa of 0, indicating no agreement at all. However, as this code had only been perceived one time by one coder, this issue was not concerning but we agreed to discuss any instances of this code going forward.

Once all the video coding was complete, I inspected the computer logs collected during the experimental sessions. I focused on the logs of triggered adaptations and touch pattern labels to address the questions listed in the introduction to this chapter: (1) Are adaptations triggered appropriately? (2) How do students respond when the adaptations are triggered?

To determine if the adaptations were triggered appropriately, I looked at the video codes in the intervals leading up to the triggering of each adaptation. If the video codes were primarily negative, I considered the adaptations appropriate. Additionally, I looked for video intervals that

showed improvement while adaptations were active to determine if these improvements were recognized and rewarded by removing the adaptations. Finally, I looked for video intervals that showed collaboration problems that were undetected by the computer.

To understand how students responded to the adaptations, I first looked at the video codes in intervals immediately following the triggering of an adaptation. For an adaptation to be considered successful, the video codes should show improved collaboration. Due to the level of frustration experienced by students in the pilot session, I also reviewed the videos from the experimental sessions to understand the groups' emotional responses to the adaptations. It was my hope that the changes made after the pilot would have greatly improved the students' experiences of the adaptations.

The study design allowed for comparison between groups that received the adaptations and those that didn't. Although this comparison would help me to assess the impact of the adaptations, it is important to recognize that differences between the groups in terms of the students' motivation, willingness to engage in the assigned activity, and overall group dynamics would also have a great deal of impact on whether or not the adaptations were needed and how they were received by the students. Students were randomly assigned to groups and groups were randomly assigned to conditions in an attempt to reduce these differences between groups. However, as anyone who has spent time in a high school classroom may expect, the likelihood of the groups being truly equal, especially in a study of this scale, was very small. The purpose of the baseline data collection in phase 1 of the study was to understand these differences.

Baseline data analysis involved review of the survey of students' experiences and attitudes about collaborative work as well as review of the coded videos of each group's first session at the tabletop, in which no adaptations occurred. This analysis of the baseline data allowed me to

determine the most appropriate way to approach the analysis of the comparative, experimental sessions in phase 3 of the study. A purely quantitative comparison (e.g., of the occurrence of positive and negative collaboration codes in the experimental sessions versus those of the control sessions) would only be appropriate if the baseline data showed the groups to be equivalent in terms of their ability to collaborate effectively without intervention. Should the groups be functionally different in their ability to collaborate effectively, a more qualitative and less generalizable comparison would be appropriate.

8.1.6 *Results*

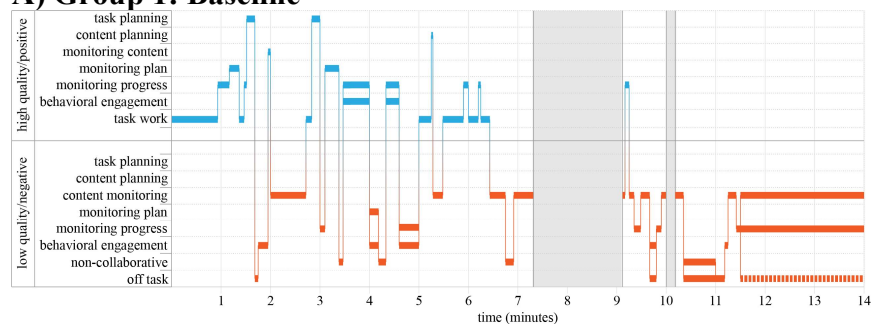
The results from the experimental phase of the study suggest that the revised approach to triggering adaptations showed promise. Collaboration improved, on some dimensions, immediately following the first interventionist adaptation in the sequence every time it was triggered, meaning that later adaptations were never triggered. This is a positive result because the adaptation appeared to lead to more effective collaboration very quickly and consistently, but a side effect is that most of the adaptations were not tested in this phase of the study.

Additionally, the adaptations could not address all collaboration problems—disengaged students who showed no inclination to participate in the group work remained disengaged whether or not the adaptations were present, and off-task interactions taking place away from the computer could not be detected. The biggest impacts were reduced disruption caused by individual students and reduced periods of low-quality collaboration.

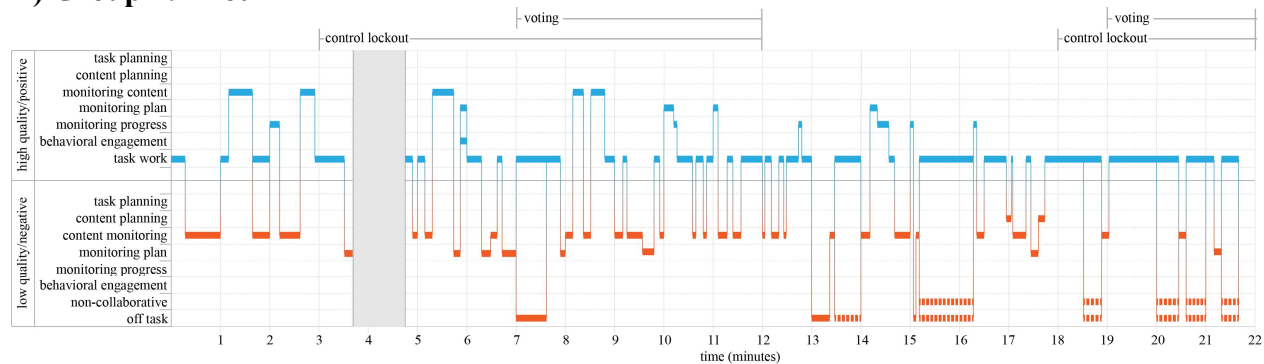
Both the survey of students' attitudes toward group work and the analysis of the baseline session videos showed differences among the groups in terms of their ability to work collaboratively. Therefore, the focus of this section is more qualitative than quantitative. Figure 8-10, Figure 8-11, and Figure 8-12 illustrate how each group's collaboration progressed in each

session, including when adaptations were triggered. These figures were created using the video analysis codes. Blue bars above the horizontal center line represent episodes of high-quality social regulation or otherwise positive codes, such as collaborative task work. Orange bars below the horizontal center line represent episodes of low-quality social regulation or other negative codes, such as off task interactions. Dashed lines indicate that the code was applied to interactions or behavior that was taking place away from the computer, and therefore not visible to the touch patterns used to trigger adaptations. Interactions occurring away from the computer usually coincided with off-task behavior. For example, when group members were turned away from the tabletop and looking at their phones or doodling on the whiteboard.

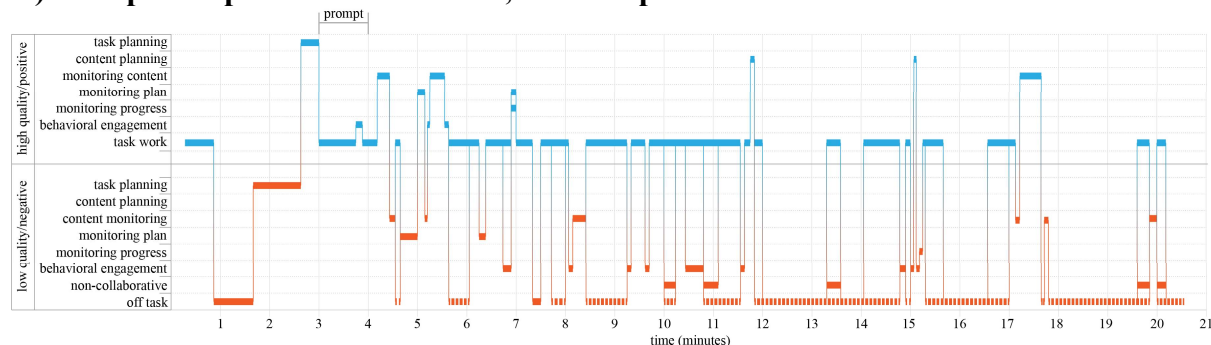
A) Group 1: Baseline



B) Group 1: Pilot



C) Group 1: Experimental Session 1, with adaptations



D) Group 1: Experimental Session 2, with adaptations

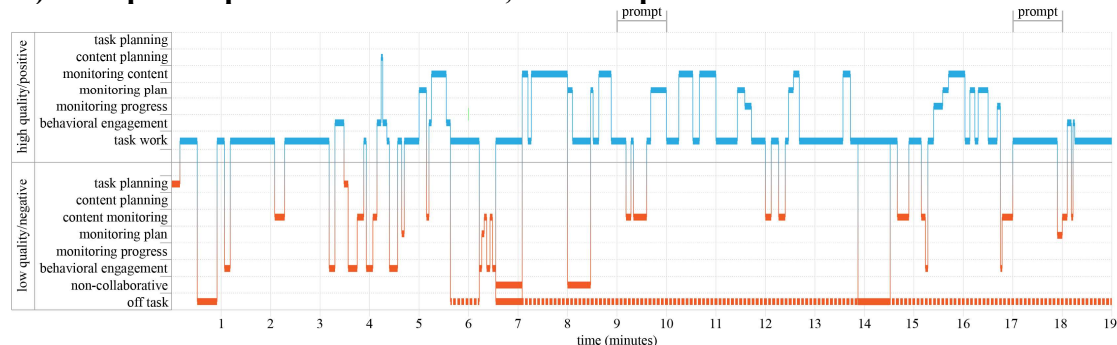
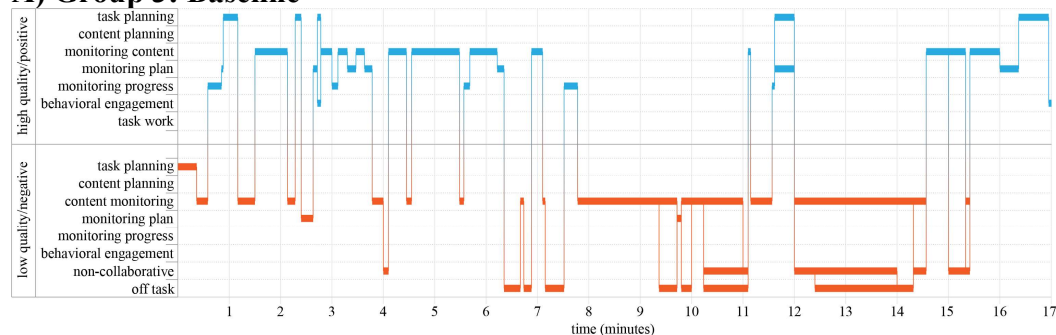
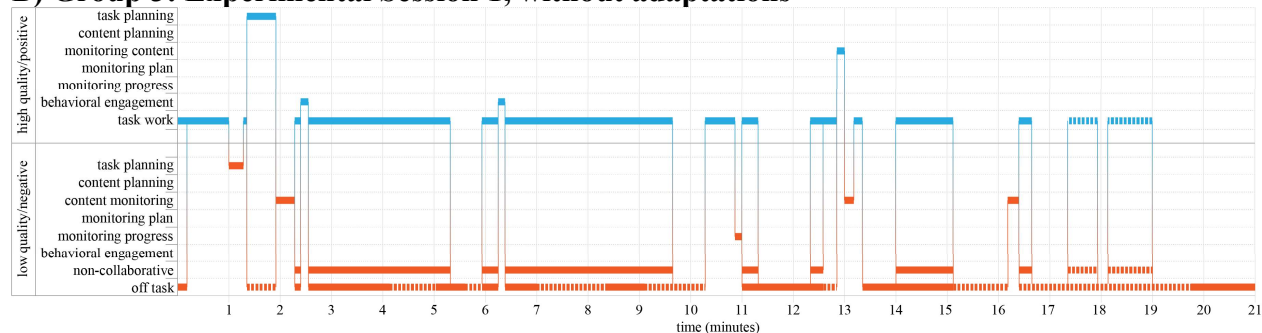


Figure 8-10. Social regulation and collaboration codes over time for Group 1. Blue bars represent episodes of high quality social regulation/positive group work codes. Orange bars represent low quality social regulation/negative codes. Dashed bars represent interactions taking place away from the computer. Gray blocks represent interruptions by program staff, which were not coded.

A) Group 3: Baseline



B) Group 3: Experimental Session 1, without adaptations



C) Group 3: Experimental Session 2, with adaptations

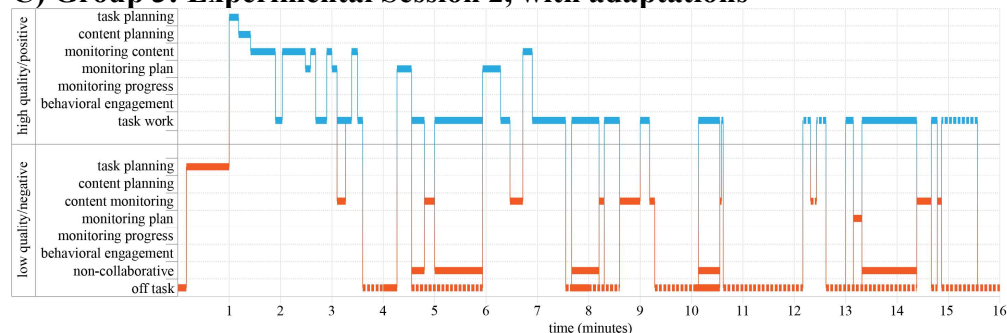


Figure 8-12. Social regulation and collaboration codes over time for Group 3. Blue bars represent episodes of high quality social regulation or positive group work codes. Orange bars represent low quality social regulation or negative codes. Dashed bars represent interactions taking place away from the computer.

8.1.6.1 Survey of student attitudes toward group work

The survey of students' attitudes toward group work, conducted on the first day of class, asked students to indicate their agreement or disagreement with 36 statements about their experiences of group work in their classes in general. The results showed that the class as a

whole was comfortable in small group settings. The students reported feeling confident in their ability to contribute to group work and generally supported by group mates. All but one student agreed or strongly agreed that it was important to them that the group get their work done on time. The students began the summer program the day after the last day of the regular school year. Therefore, it seems likely that their responses on the survey were heavily influenced by their experiences in the classes they had most recently taken in school.

The population size was too small to determine statistically significant differences between groups in terms of attitudes toward group work. However, the survey did provide some insight into student attitudes that might affect their group dynamic. Survey statements that showed differences between the groups are highlighted below.

Six of the eleven students in the class believed that they typically do most of the work in group settings compared to other group members. The majority of students in groups 1 and 3 believed that they do most of the work while the only two students who did not think they did most of the work in group settings were both in Group 2.

Overall, students indicated that group members respected their opinions. Two students felt that this was not the case, however. Both of these students were in Group 1.

Three students in Group 1 agreed that working in groups made course material easier to understand and reduced the workload. The attitudes of students in groups 2 and 3 were more varied.

All students in Group 2 felt that they were a part of what was going on in group settings. Two students in Group 1 and one student in Group 3 were neutral on this statement.

All students in Group 1 agreed that, in group settings, one student typically makes all the decisions. This was not the case for groups 2 and 3.

Only one student in Group 1 disagreed with the statement that they found it hard to express their thoughts in group settings. This is in contrast to the other groups, especially Group 2.

This survey was administered on the first day of class, before many of the students got to know each other and before they began working together for this study. Therefore, it may not necessarily reflect their experiences in the tabletop sessions described below. However, it does suggest some key differences between the groups. Overall, more students in Group 1 indicated negative opinions about group work while Group 2 had more positive opinions. The students in Group 3 tended to have differing opinions on these statements.

8.1.6.2 Results of the video analysis of baseline tabletop sessions

In addition to the differences in students' attitudes toward and experiences of group work revealed by the survey, the video analysis of the baseline sessions also showed differences in each group's working dynamic. Two of the students in Group 1 were motivated to complete the activity, but the group as a whole struggled to manage the process, causing them to get stuck on superficial tasks. In Group 2, a single student—the only girl in the group—took on the role of group leader, attempting to direct the rest of the group through the activity and keep them on track, although her efforts were often resisted or ignored. One member of Group 2 was absent the day of baseline data collection, which appeared to have a major impact on the group dynamic as he was a highly collaborative student, responsive to the female student's suggestions and ideas. In Group 3, two of the students spent most of the time chatting and gossiping, leaving another student to do almost all the work while he joined in with the conversation.

The baseline data was collected while students worked on the *Snow Leopards 101* introductory activity. All groups completed the activity, though all groups also figured out a

shortcut to finish the activity without really engaging with the content. In the final part of the activity, the students were asked to complete graphic organizers describing the physical characteristics and adaptations of four species of big cat. Each group worked out that by dropping words and phrases on a graphic organizer and pressing a feedback button, they could see which answers were correct and repeatedly retry incorrect answers without penalty. The effects of this unintentional loophole in the software served to demonstrate that task design itself can have an impact on collaboration quality.

Table 8-2 shows the occurrence of each video code by group in the baseline data collection phase. For all groups, content monitoring was the most common code. Group 1's episodes of content monitoring were almost entirely low quality, while Group 2 had equal numbers of high- and low-quality episodes of content monitoring. Group 3 had more low-quality episodes than any other group but still managed a considerable number of high-quality episodes and proportionally had fewer low-quality episodes than Group 1.

Table 8-2. The occurrence of each video code by group for the baseline data collection sessions, in which the groups used the *Snow Leopards 101* application.

| Video code | Group 1 | | Group 2 | | Group 3 | |
|--------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| | HQ/ + | LQ/ - | HQ/ + | LQ/ - | HQ/ + | LQ/ - |
| Planning - task | 2 | 0 | 4 | 0 | 5 | 2 |
| Planning - content | 1 | 0 | 0 | 0 | 0 | 0 |
| Monitoring - content | 1 | 14 | 19 | 19 | 17 | 24 |
| Monitoring - plan | 2 | 0 | 5 | 3 | 7 | 2 |
| Monitoring - progress | 6 | 8 | 8 | 5 | 5 | 0 |
| Behavioral engagement | 2 | 4 | 0 | 0 | 2 | 0 |
| Non-collaborative interactions | - | 7 | - | 14 | - | 9 |
| Off-task interactions | - | 5 | - | 2 | - | 11 |
| Task work | 6 | - | 2 | - | 0 | - |
| Total | 20 (33.90%) | 39 (66.10%) | 38 (46.91%) | 43 (53.09%) | 36 (42.86%) | 48 (57.14%) |

Group 1 struggled to get started, spending around 6 minutes on a superficial warm-up activity (a 16-piece puzzle) that the other groups completed in under a minute. The group made

some attempts to verbally coordinate their efforts at the start of the activity. However, their physical coordination was lacking and they managed to “lose” the puzzle multiple times, which meant they had to start over more than once. There was almost no deep content talk in this session because the group focused on getting the correct answers by guessing and pressing the feedback button as often as possible. The video analysis revealed a single minute of sustained high-quality collaboration as they coordinated how best to complete the puzzle without losing it for a fourth time (see Figure 8-10A, around 5 and a half minutes into the activity). However, because the students were concentrating on rectifying problems caused by the initial lack of collaboration, there was no meaningful progress on the task.

After completing the puzzle, Group 1 dismissed the on-screen instructions for the next part of the activity without reading them, so they utilized the ‘Call Help’ button to summon the teacher via Slack message. The period of time during which the teacher was present was not coded—the first grayed out portion of Figure 8-10A. The second grayed-out portion of Figure 8-10A marks a brief interruption caused by an accidental press of the Call Help button, which summoned the teacher a second time.

Nearly two-thirds of the episodes coded in the video analysis of Group 1’s first session showed low-quality or negative collaboration behaviors, more than any other group. The second half of the session was almost entirely low quality, as can be clearly seen in Figure 8-10A. Group 1 also had fewer coded episodes than the other groups because they guessed answers so quickly that they finished the task ten minutes early, though little meaningful work took place. The group then decided to hang out in the breakout room instead of going back to the main classroom, even devising a plan for how best to appear to be working in the event that a teacher made a surprise visit to the room before the allotted time was up.

Group 2 was missing one group member for the baseline session, so their first session was not a good representation of their later collaborations. Of the three students who were present, one student made numerous attempts at content monitoring and planning. Her efforts were often ignored by the other two group members. Despite this, Group 2 had frequent episodes of high quality collaboration, particularly content monitoring. Comparing the diagram of Group 2's baseline session (Figure 8-11A) to those of the other groups (Figure 8-10A and Figure 8-12A) shows that, unlike the other groups, Group 2 had consistent though intermittent episodes of high-quality/positive collaboration up until the end of the session when one of the boys in the group figured out the unintentional shortcut for the graphic organizer task (see Figure 8-11A, beginning at around 17 minutes). At this point, the group had already made substantial progress by engaging with the learning material. The girl attempted to encourage the boys to keep working on the task but eventually gave up, exclaiming "You're terrible!" One of the boys replied, "No, we're smart!" and the other boy agreed that they had "outsmarted the program." Even though the collaboration devolved toward the end of the session, this group showed more effective collaboration than either of the other groups.

Like Group 2, Group 3 had frequent episodes of high quality content monitoring, especially during the first half of the session. However, they grew bored with the activity and discussion turned to other topics, ranging from former classmates to what they liked to eat for lunch. Although they continued working on the task while the conversation was off-task, they were working mostly independently. Group 3 figured out the shortcut faster than Group 2 but, unlike Group 2, they attempted to make educated guesses based on the activity content before dropping answers on graphic organizers and getting feedback (see Figure 8-12A, minutes 8 through 15). Additionally, although their collaboration had been primarily low quality during the

second half of the session, they returned to high quality collaboration at the very end as they reflected on the completed graphic organizers and used their phones to record the information they thought would be important in later class activities.

The baseline sessions showed that each group differed greatly in approach and interpersonal dynamics. Group 1's working style could best be described as physical and even aggressive at times, with a fair amount of colorful language and insults. The four boys were on good terms, however, and made genuine efforts to get to know each other. Their collaboration woes could be attributed to the two group members who were determined not to engage with the activity as intended, refusing to read any text on screen or reason about the learning content. One of the disengaged students made repeated attempts to gain control of the screen, causing substantial disruption and eventually derailing the entire collaboration.

In stark contrast, Group 2's interactions were cold and slightly hostile as the two boys in the group repeatedly responded to the girl's contributions with silence. Unlike the engaged students in Group 1, who eventually gave up in the face of repeated disruption, the girl in Group 2 persisted until the very end of the session. In later sessions, she found an ally in the student who was absent the day of the baseline session.

Where Group 1 was lively and Group 2 was cold, Group 3 was very laid back and congenial. All students in Group 3 appeared motivated to engage with the activity at the start, but two of the students had a propensity for off-task chatting, causing them to frequently lose focus. These differences between the groups seemed likely to impact both the need for adaptations and how each group would respond when adaptations were triggered. For example, although Group 1's physical interaction style led to conflicts in the software, much of the disruption could be attributed to one student. I expected that, as the experimental group, Group 1 would be likely to

trigger the adaptations because of his behavior and hoped that the adaptations could help to reduce it. In contrast, the students in Group 3 alternated between working in parallel (Figure 8-4), which could cause the adaptations to trigger, or leaving one student to work while the other two chatted, which would not trigger the adaptations as Group Touch and the touch patterns are unable to detect when students are *not* interacting with the computer.

8.1.6.3 Results – Experimental Sessions

In the experimental phase of the study, groups worked on the *Game Challenge* activity across two class periods. Group 1, the experimental group, used the application with adaptations both times. The sequence of adaptations was triggered three times, always at appropriate moments. The adaptations encouraged on-task collaboration from two students while deterring off-task group members from blocking progress. Group 2, the control, used the application without adaptations in both sessions. The log files showed that, had the adaptations been present, the sequence of adaptations would not have triggered in either session although the *group awareness icon* would have turned orange on several occasions. Group 3 used the application without adaptations in the first class period, then with adaptations in the second class period. The log file shows that the adaptations would have been useful in their first session. Their collaboration was much improved in the second session and no adaptations were triggered because no sustained low quality collaboration was detected, although the *group awareness icon* did turn orange on a couple of occasions. Table 8-3 shows the summary of video codes for each group as they used the *Game Challenge* application during the experimental phase of the study. Figure 8-10C & D, Figure 8-11B & C, and Figure 8-12B & C show how each session progressed.

Table 8-3. Summary of video codes from the experimental sessions. Group 1 received the adaptations in both sessions. Group 2 was the control. Group 3 used the application without adaptations in the first session, then with the adaptations in the second session. *Group 3's episodes of *non-collaborative interactions* were coded as *task work*-only by the video coder. I added the *non-collaborative interactions* code for task work episodes during which one student was working alone and the other two students were chatting.

| | | Group 1 - adaptations | | Group 2 – no adaptations | | Group 3 – no adaptations | |
|--------------|--------------------------------|-------------------------------|------------------------------|-------------------------------|------------------------------|-------------------------------|------------------------------|
| Session 1 | Video code | <i>High quality/ positive</i> | <i>Low quality/ negative</i> | <i>High quality/ positive</i> | <i>Low quality/ negative</i> | <i>High quality/ positive</i> | <i>Low quality/ negative</i> |
| | Planning - task | 2 | 2 | 0 | 1 | 1 | 1 |
| | Planning - content | 2 | 0 | 3 | 5 | 0 | 0 |
| | Monitoring - content | 4 | 6 | 13 | 9 | 1 | 4 |
| | Monitoring - plan | 2 | 3 | 9 | 1 | 0 | 0 |
| | Monitoring - progress | 0 | 2 | 1 | 2 | 0 | 1 |
| | Behavioral engagement | 3 | 8 | 2 | 2 | 2 | 0 |
| | Non-collaborative interactions | - | 5 | - | 0 | - | 23* |
| | Off-task interactions | - | 13 | - | 2 | - | 12 |
| | Task work | 31 | - | 29 | - | 23 | - |
| Total | 44 (53.01%) | 39 (46.99%) | 57 (72.15%) | 22 (27.85%) | 27 (39.71%) | 41 (60.29%) | |

| | | Group 1 - adaptations | | Group 2 – no adaptations | | Group 3 – adaptations | |
|--------------|--------------------------------|-------------------------------|------------------------------|-------------------------------|------------------------------|-------------------------------|------------------------------|
| Session 2 | Video code | <i>High quality/ positive</i> | <i>Low quality/ negative</i> | <i>High quality/ positive</i> | <i>Low quality/ negative</i> | <i>High quality/ positive</i> | <i>Low quality/ negative</i> |
| | Planning - task | 0 | 1 | 2 | 2 | 1 | 1 |
| | Planning - content | 1 | 2 | 1 | 3 | 1 | 0 |
| | Monitoring - content | 10 | 15 | 2 | 15 | 6 | 11 |
| | Monitoring - plan | 8 | 2 | 4 | 5 | 6 | 0 |
| | Monitoring - progress | 3 | 0 | 1 | 2 | 0 | 0 |
| | Behavioral engagement | 7 | 10 | 0 | 1 | 0 | 0 |
| | Non-collaborative interactions | - | 2 | - | 0 | - | 5* |
| | Off-task interactions | - | 2 | - | 6 | - | 13 |
| | Task work | 41 | - | 15 | - | 20 | - |
| Total | 70 (67.31%) | 34 (32.69%) | 25 (42.37%) | 34 (57.63%) | 34 (53.12%) | 30 (46.88%) | |

In total, adaptations were triggered three times. All three occurrences happened with Group 1, the group of four boys in the experimental condition, and the sequence never

progressed past the *prompt* stage. The planned sequence progression was as follows: (1) *group awareness icons*, which changed color based on the quality of collaboration detected by the touch patterns, were always visible on screen; (2) when poor quality collaboration was initially detected after a period of high quality collaboration, a *prompt* appeared, providing some tips for effective collaboration and a link to the task instructions; (3) if collaboration did not improve, *control lockout* would be activated; (4) *voting* would be the final interface-based intervention; and (5) *escalate to authority* would summon the teacher if all else failed. The fact that the adaptations never progressed past the *prompt* means that the detected collaboration quality immediately improved once the prompt was displayed.

Group 1 had been part of the pilot session, in which the original sequence of adaptations was triggered twice—once at 3 minutes and again at 18 minutes into the session, with adaptations active on-screen for a total of 13 minutes (see Figure 8-10B). The formal video analysis of the pilot session confirmed the initial informal analysis described in Section 7.2.4—adaptations were triggered after periods of low-quality collaboration, and high-quality collaboration increased following the triggering of an adaptation. Although low-quality codes still occurred when adaptations were triggered, Figure 8-10B shows that episodes of low-quality collaboration were briefer when adaptations were present. Additionally, high-quality collaboration occurred throughout the session, unlike Group 1’s baseline session, which derailed about halfway through the activity.

In the first experimental session with the finalized sequence of adaptations, Group 1 only triggered the *prompt* once, at 4 minutes into the session (see Figure 8-10 C). The *group awareness icon* turned from green to orange three times—in the two early intervals leading to the prompt, again at 17 minutes into the activity, and finally at 19 minutes in. The icon also turned

grey on two occasions as the logs and video showed the group stopped interacting with the screen (see Figure 8-10 C, at around 12 and 18 minutes). However, there was no evidence in the video that the students noticed these color changes.

The *prompt* was triggered after an interval labeled low-quality by *Unrelated Touches* and medium-quality by *Overlapping Unrelated Sequences* (see Figure 8-10C, 0-2 minutes) followed by an interval labeled medium-quality and high-quality by each pattern respectively (see Figure 8-10C, 1-3 minutes). The video of Group 1's first experimental session showed that the *prompt* was triggered at an appropriate time, after several minutes of the group being off task and pressing buttons on-screen without any explicit coordination. The prompt included a reminder that students should make sure they understood the task goal and that they were working toward it. This reminder appeared to be appropriate content for the prompt when triggered early in the activity. Immediately before the prompt appeared, one of the students expressed confusion about what the group was supposed to be doing but didn't want to read the instructions, preferring to "just figure it out." Refusing to read the instructions was consistent with this student's behavior in both of Group 1's previous sessions. The activity instructions had been given verbally in the introduction and the group had a printed copy of the instructions on a worksheet in addition to the on-screen version. Another student was in the process of explaining the activity to the confused student when the prompt appeared, leading the group to quickly revisit the instructions. The sequence of adaptations did not progress to the next stage because, immediately after reading the prompt and the instructions, the group began to engage in task work, which was by far the most common video code in this session. It is worth noting, however, that only some students in the group were engaged during most of the task work episodes. The disengaged, off task students typically refrained from touching the computer in this session and were therefore

undetectable. Figure 8-10C shows that the majority of negative codes (primarily *off-task*) in this session occurred away from the computer after the prompt was triggered. This behavior was noticeably different from Group 1's baseline and pilot sessions, in which these students would attempt to interact with the screen without fully engaging with the activity, disrupting the progress of students who were engaged.

When the *group awareness icon* turned gray due to lack of interaction, the video analysis showed the group to be completely off task, talking about clothing and not engaging with the activity at all. In the intervals that caused the icon to turn orange, some students were off task while others were engaged in intermittent task work and low-quality content monitoring.

In Group 1's second experimental session, they triggered the *prompt* adaptation twice, at 9 and 17 minutes into the activity (see Figure 8-10D). The sequence did not progress beyond the *prompt* in either case. In this session, the group focused on a single piece of the *Game Challenge* activity, in which they were tasked with editing a conditional statement (in C# code) to make a game object, the Player, change color and size depending on where it was on screen. In "play mode," the students could test out their code by moving the Player using the orange buttons shown in Figure 8-3, right.

The first *prompt* in Group 1's second experimental session was triggered after two intervals labeled medium-quality by *Unrelated Touches* and high-quality by *Overlapping Unrelated Sequences* (see Figure 8-10D, 6-8 minutes & 7-9 minutes). The video analysis showed that, in the intervals leading up to the prompt, the whole group was engaging in primarily low-quality collaboration for around a minute, followed by a period of high-quality collaboration between two students with the other two students completely off task, their attention focused on their smartphones as they swapped social media details. One of the off-task students also

occasionally fiddled with objects on-screen that were not needed for the activity that the two on-task students were working on, likely contributing to the prompt being triggered.

Review of the video suggested that the prompt was justified due to the behavior of the off-task students plus the whole-group low-quality collaboration. However, collaboration between the two engaged students had improved during the second interval and the interruption appeared to be disheartening for them—when the prompt appeared, one of the engaged students let out a big sigh and the other dismissed it. The engaged students continued to collaborate on the task for the next five minutes. One of the off-task students briefly attempted to catch up with what his teammates were doing before declaring, “I have no idea what’s going on” and returning to his phone. These intervals were labeled high quality by both touch patterns. This is likely because the off-task students were not interacting with the computer so only the actions of the engaged students factored into the touch pattern calculations.

The prompt was triggered a second time after an interval labeled medium quality by *Unrelated Touches* and high quality by *Overlapping Unrelated Sequences* (see Figure 8-10D, 14-16 minutes) followed by an interval labeled low quality and high quality by each pattern respectively (see Figure 8-10D, 15-17 minutes). In these intervals, one of the off-task students became interested in the screen, wanting to control the Player (Figure 8-13). He took over the orange controller buttons (see Figure 8-3, right) and repeatedly hammered on the play mode button. Although Figure 8-10D shows that this student’s low-quality interactions were brief compared to the high-quality collaboration of the other students, the video showed his interruptions to be highly disruptive. The engaged students, who were very close to solving the puzzle, tried to continue working on the task despite the chaos caused by their teammate rapidly toggling “play mode” on and off. The engaged students were, once again, annoyed by the

appearance of the prompt, but they were able to dismiss it quickly and it did appear to deter the other student from hammering on-screen. He sat back from the computer but made some verbal contributions to the collaboration—encouraging his teammates as they solved the challenge.

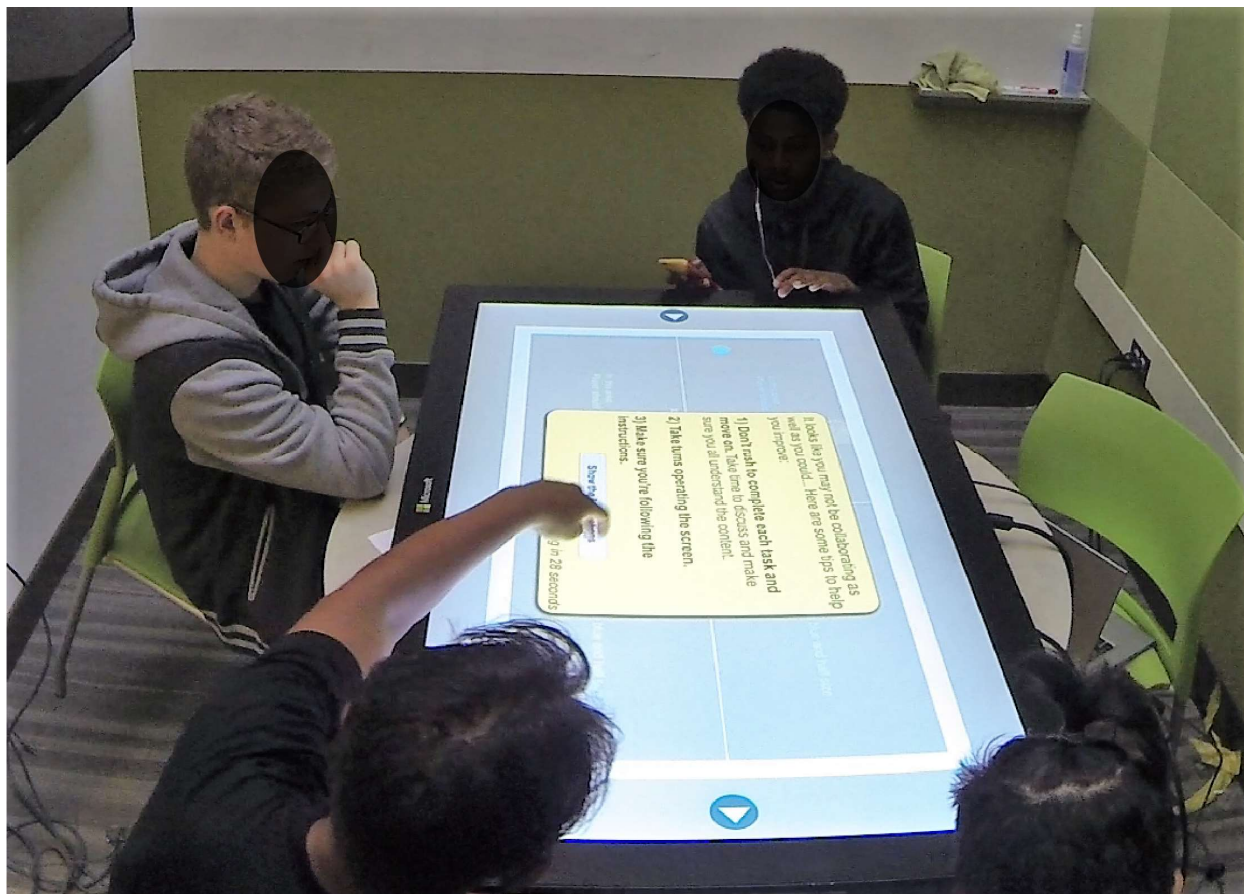


Figure 8-13. Group 1 triggers the prompt for the second time in their second experimental session. The two students on the left were engaged in the activity. The student at the top right disrupted the activity by hammering on the screen. The student in the bottom right was sitting back from the tabletop and using his phone.

Table 8-3 shows many more instances of positive collaboration behaviors for Group 1 in the experimental sessions than in the baseline session (Table 8-2). This result should be approached with caution, however, as this increase in positive collaboration codes may be due to the differing characteristics of each activity and the students' increased familiarity with each other after five weeks in class together. Table 8-3 shows many more occurrences of the *task*

work code across groups in the experimental sessions compared to the baseline sessions in Table 8-2, for example. This difference suggests that the different activities demanded different approaches, affecting how the groups managed their collaboration. Additionally, the video recordings also suggested another explanation—one student who was quite often a driving force behind problems in the earlier sessions, largely due to his refusal to read or follow task directions, had mostly withdrawn from participation in the experimental sessions. It is unclear whether this was due to the *prompt*. His withdrawal allowed his more engaged team members to complete the activity with minimal disruption, but he would have gained nothing from the activity used in the experimental sessions.

As the main impact of the adaptations on Group 1's collaboration appeared to be reduced disruption rather than an increase in high-quality collaboration, I compared the length of periods of sustained low-quality collaboration taking place at the computer in three types of intervals: (1) "pre-adaptation"—intervals that caused an adaptation to trigger; (2) "adaptation present"—intervals in which an adaptation is present; and (3) "no adaptation"—intervals during which no adaptations are present. I use the phrase "periods of sustained low-quality collaboration" to mean periods of time with one or more continuous episodes of low-quality collaboration. For example, Figure 8-10A shows that, after the last interruption in that session, collaboration was coded as negative for the remainder of the session. In this analysis, that period of time was counted as a single period of sustained low-quality collaboration that contained multiple episodes. Although Figure 8-10 shows that the adaptations did not eliminate low-quality collaboration altogether, a reduction in the length of periods of low-quality collaboration after the introduction of adaptations would indicate improvement. Table 8-4 shows the result of this comparison.

Table 8-4. The median length of sustained periods of low-quality collaboration at the tabletop in each of Group 1's sessions. The “pre-adaptation” row shows the median length of periods of low-quality collaboration during intervals that caused an adaptation to be triggered. “Adaptation present” refers to periods of low quality collaboration that occurred while a triggered adaptation was active. “No adaptation” covers all other intervals. Episodes of negative collaboration that took place away from the computer are not included.

| Interval Type | Session | | | | | | | |
|--------------------|-------------------|------------------|-------------------|------------------|-------------------|------------------|-------------------|------------------|
| | Baseline | | Pilot | | Experimental 1 | | Experimental 2 | |
| | Median length (s) | # of occurrences | Median length (s) | # of occurrences | Median length (s) | # of occurrences | Median length (s) | # of occurrences |
| Pre-adaptation | | | 16 | 7 | 106 | 1 | 15 | 5 |
| Adaptation present | | | 9.5 | 19 | 0 | 0 | 11 | 2 |
| No adaptation | 20 | 11 | 9 | 6 | 8 | 19 | 10 | 15 |

Table 8-4 shows that the adaptations did appear to reduce the length of periods of low-quality collaboration involving the computer. The median length of periods of low-quality collaboration was consistently longer in intervals that caused an adaptation to trigger than in other intervals in all three sessions where the adaptations were present. In the baseline session, when the adaptations were not in use, the median length of sustained low-quality collaboration was at least twice that of the sessions where the adaptations were available. In both experimental sessions, there were considerably more occurrences of low-quality collaboration when no adaptations were present than in pre-adaptation intervals or intervals where adaptations were active. This effect occurred because, when there were no adaptations present, occurrences of low-quality collaboration were brief and punctuated by high-quality collaboration, causing the number of occurrences to increase and the length of the occurrences to decrease.

Group 2, the control group, had all four group members present for the *Game Challenge* sessions—one student had been out sick during baseline data collection. The presence of the fourth student had a noticeable effect on the group dynamic. In the baseline session, the girl in

the group had attempted social regulation on many occasions but was typically ignored by her teammates. However, the boy who had been out sick for the baseline session was a lot more responsive to her efforts and together, these two students led the collaboration in both *Game Challenge* sessions.

As the control group, Group 2 used the tabletop without any interventions. The logs of touch pattern outputs showed that the sequence of adaptations, beginning with *prompt*, would not have been triggered in either session, even if they had been available. This outcome suggests considerable differences between Groups 1 and 2 in terms of their ability to collaborate effectively, at least in terms of those behaviors that are detectable by the touch patterns. Table 8-3, which shows the count of positive and negative video codes, confirms that Group 2 exhibited more effective social regulation than Group 1 in the first session, with many more instances of high-quality content monitoring. Their collaboration looked very different in the second session, however, with very few instances of high-quality content monitoring compared to low-quality content monitoring and more negative collaboration codes overall. Figure 8-11B and C illustrates how collaboration progressed in these sessions, showing an increase in off-task behavior away from the tabletop.

The pattern log file for the first session showed that a *group awareness icon*, had one been present, would have turned orange to represent potentially poor-quality collaboration twice, at 9 and 11 minutes into the activity, with one minute of green to represent good-quality collaboration in between the two occurrences. In both cases, the intervals that would have turned a *group awareness icon* orange were labeled as medium-quality according to Unrelated Touches and high-quality according to Overlapping Unrelated Sequences (see Figure 8-11B, minutes 7-9 & 9-11). The video analysis for these intervals showed brief low-quality interactions interspersed

with periods of task work and occasional high-quality interactions. Overall, however, this session was characterized by high-quality *content monitoring*, high-quality *plan monitoring*, and *task work* for three group members. One student was disengaged for much of the activity.

The touch pattern log for Group 2's second session showed that all but one of the intervals were labeled high-quality by both patterns. One interval, beginning six minutes into the session, was labeled poor-quality by Unrelated Touches and high-quality by Overlapping Unrelated Sequences (see Figure 8-11C, minutes 6-8). The video analysis revealed that this interval was bookended by extended periods of task work but primarily contained low-quality social regulation episodes, particularly *content monitoring*. Almost one quarter of all negative collaboration codes for the whole session occurred in this single interval.

The touch pattern detection of almost entirely high-quality collaboration in this session did not line up with the video analysis, which showed that this session was skewed toward poor quality collaboration. Inspection of the video analysis alongside the pattern logs suggested that much of the poor-quality collaboration was not detected because low-quality episodes of social regulation were brief and fairly evenly distributed across intervals (with the exception of the interval labeled low-high quality by the touch patterns as described above). Low-quality episodes that involved the computer were interspersed with longer periods of task work and high-quality episodes of social regulation. Additionally, most of the occurrences of *off-task* took place away from the tabletop.

Group 3, the group of three students in the "split" condition, used the *Game Challenge* application without adaptations in their first session of the experimental phase and with the adaptations in their final session. Unlike Group 2, Group 3 would have triggered the sequence of

adaptations in their first session, had the adaptations been present, because 11 of the 21 intervals logged were labeled low-quality by Unrelated Touches.

The video showed that the group spent the first couple of minutes chatting, messing around with objects on-screen without any clear purpose, and carrying out administrative activities such as writing their names on a worksheet. Then, after a few brief attempts at planning and engaging as a group were shut down by one student, another student took on the responsibility of completing the activity by himself and the majority of codes for this session were *non-collaborative task-work*. The other two students in the group were quite happy to spend the rest of the session sitting back and chatting, as they had done in their previous sessions at the tabletop. One of the off-task students continued to press buttons and touch objects on screen absentmindedly while chatting, which is likely to have contributed to the “low-quality” labels generated by Unrelated Touches. This purposeless fiddling with the interface was the type of interaction that appeared to be deterred by the adaptations in Group 1. The student that took over the activity was the highest achieving student in the class and, by the end of the session, Group 3 had completed more of the activity than any other group thanks to his solo work. One of the other group members acknowledged his efforts, saying, “Thanks so much! You’re so nice!”

For the second session of the *Game Challenge* activity, Group 3 used the application with the adaptations—the first time the adaptations were made available to this group. The pattern logs looked very different for this session than Group 3’s first session—apart from two intervals in which the *group awareness icon* turned orange, all intervals in which the group were active on the tabletop were labeled high-quality by both touch patterns. The sequence of adaptations was not triggered.

The video recording showed that Group 3 behaved very differently in this session—both of the students who spent the previous session chatting were more engaged in the second session, helping to plan the group’s approach and contributing their ideas to solve the challenge, especially toward the beginning of the session. It is unclear what prompted this change. The group was introduced to the adaptations at the start of the activity so it is possible that simply being aware that the computer was looking for signs of poor collaboration was enough to bring about improvement. However, there was no evidence of such an effect in the communication among the group members. The summary of video codes in Table 8-3 shows that, despite increased engagement, there were still more episodes of low quality social regulation than high quality, especially for content monitoring, and there were still many more off-task episodes. The visualization of the collaboration in Figure 8-12C provides a more complete picture, however. Figure 8-12C shows that, low-quality content monitoring episodes were typically brief and the majority of time spent off task did not involve the computer and therefore could not be detected, once again highlighting the major limitation of detecting collaboration quality using only touch data.

The video codes for this session did not initially line up with the touch pattern logs, which showed almost entirely high-quality intervals. The video explained this discrepancy, however. As in this group’s previous session, off-task episodes took place away from the computer and were therefore undetectable—a limitation of detecting collaboration quality using only touch data. Episodes of low-quality content monitoring were closely coupled with off-task episodes—one student would make an argument or a contribution to the task but was ignored by the other students who had reverted to chatting and not interacting with the computer. Therefore, these episodes were also undetectable. Also problematic were the episodes where a single

student did all of the work, signified in Figure 8-12C by simultaneous bars for both *task work* and *non-collaborative interactions*. To the touch patterns, these single-student episodes looked very much like episodes of high-quality collaboration where multiple students made positive verbal contributions while one student operated the computer.

8.2 DISCUSSION

The results of this study show that my approach to detecting collaboration quality at tabletop computers was able to detect certain collaboration problems—primarily disruption caused by individual students and poor coordination among group members—and trigger adaptations appropriately. Collaboration problems that happened away from the computer, namely disengaged students sitting back and interacting with their phones instead of the tabletop, were not detectable, however. Episodes where a single student did all of the work were also undetectable.

How the adaptations were received by students differed by adaptation and how motivated students were to engage with the activity—motivated students tended to adjust their behavior in a positive direction but already disengaged students found the adaptations to require too much effort and were put off completely. Although this effect may have led to positive outcomes for the motivated students, who were able to make progress where they had previously been completely blocked by disruptive students, it is obviously problematic for the disengaged students. Additionally, if one of the goals of collaborative learning is for the whole group to benefit from the contributions of all group members, then the group suffers when individual students are disengaged. This problem of disengaged students is out of the scope of this dissertation, however.

The adaptations tested in this study (including the pilot session) were the always-on *group awareness icon*, *prompt*, *control lockout*, and *voting*. *Escalate to authority* was also implemented as the final stage of the sequence of adaptations but it was never deployed because the group that triggered the sequence due to low-quality behavior returned to high-quality collaboration after earlier adaptations each time the sequence was triggered, meaning *escalate to authority* was not needed.

Both groups that used the tabletop with adaptations saw the *group awareness icons* change color in response to the collaboration patterns detected by the computer. However, with the exception of the pilot session, there was no evidence that students made use of the icon. No verbal or gestural references were made to the icon in three out of four sessions with the adaptations. This could be due to their small size and placement in the corners. However, numerous references were made to the even less visually noticeable “Call help” buttons, which were the same size and located right next to *group awareness icon*. Another reason that groups did not make use of the *group awareness icons* could be that they needed more explicit training on when and how to use them—it is possible that they simply forgot what they were and why they were there.

The *prompt*, which was introduced after the pilot test, appeared to be effective at encouraging students to think about how they were interacting and to make sure they were, in fact, working on the assigned task. The first time the prompt appeared, the students took time to read it and follow its advice—to revisit the task instructions and make sure that what they were doing aligned with the activity goals. The second and third time it appeared, the students were quicker to dismiss it, possibly due to familiarity, but both times, it caused an off-task student (a different student each time) to stop disruptive behavior. One of the goals of this study was to

develop evaluations that were as task-independent as possible. However, in the case of prompts, it would be worthwhile exploring if prompts that are closely tied to the specific part of the task that a group is working on could be more effective. In this study, the second and third time the prompt appeared it had the effect of stopping a disruptive student but it may have been more useful to the collaboration if it had also provided advice that was useful to the two students who were engaged and making good progress on the task.

The second adaptation in the sequence, *control lockout*, was only triggered in the pilot session with Group 1. This adaptation blocks students from interacting with on-screen objects that aren't related to the aspect of the task the group is working on. By preventing students from interacting with objects on screen, *control lockout* dictates what the group is able to do thereby forcing them to work only on what is relevant to the specific assigned task. Therefore, although the intervals that followed *control lockout*'s appearance in the pilot showed improved collaboration, it was unclear whether that improvement was due to students reflecting on how they were collaborating and making conscious adjustments or simply because any other ways of interacting were blocked by the adaptation. The former outcome would be the most desirable but the latter seems more realistic.

In the video of the pilot session it appeared that, although the group read the warning that precedes activation of *control lockout*, once the adaptation was triggered, there was no indication that the group put any thought into *why* particular controls were blocked or why the number of on-screen objects they could interact with were reduced. In terms of Wise's (2014) framework for the design of learning analytics interventions, the process of grounding, during which students come to understand how interventions are tied to the goals of an activity, was not entirely successful. The simple design of *control lockout* supported one of the goals of the

study—to develop adaptations that were as application- and task-independent as possible. However, as with the *prompt*, it would be worthwhile to explore if supplementing *control lockout* with task-specific guidance would increase its effectiveness, for example, by giving students the option to learn why a particular object was blocked in terms that are closely tied to the activity goals.

Voting was also only triggered in the pilot session. Like *control lockout*, *voting* did lead to improved collaboration, primarily increased coordination among group members. Also like *control lockout*, the improvements following the activation of *voting* appeared to be due to the fact that the adaptation forced particular types of interactions rather than the students' own reflection and conscious effort to change their approach.

I consider the fact that, once *prompt* was added to the sequence of adaptations, no further adaptations were triggered, to be a positive outcome for this study even though this means that further investigation would be needed to verify the later adaptations. In all cases, the prompt was followed by sustained periods of high-quality collaboration, albeit only for those students who were motivated to work on the task. Beyond this study, the ideal outcome of using these adaptations over a longer period of time would be they render themselves unnecessary—with such an outcome it would be possible to conclude that the adaptations successfully scaffold effective collaboration, fading once a group has adopted the principles the adaptations support (Pea, 2004).

The results of this study point to a possible trade-off between mostly task-independent adaptations, which appear to force behavior change, and highly task-specific adaptation, which may lead to more conscious reflection among the students. Task-independent adaptations are easy to implement in any application and allow for flexibility in how an application is used in a

classroom alongside other activities. However, providing fairly general, high-level prompts and advice, which was the case for the adaptations in this study, requires that students do the work of figuring out how to apply that advice to their specific situation. Unfortunately, in this study, two of the students in the group that saw the adaptations were unwilling to read or engage with the assigned tasks. Therefore, it is not surprising that they did not want to do the work of engaging with the additional prompts or advice provided by the adaptations.

Following from Wise's (2014) framework, better grounding of the adaptations in terms of the course goals may increase their effectiveness. Although each group that experienced the adaptations were introduced to them as part of each activity and were given the opportunity to ask questions, these introductions appeared to promote awareness but stop short of understanding and grounding.

The problem of unmotivated, disengaged students is unlikely to be fixed by prompts or advice embedded in a tabletop application, however, no matter how well-grounded. The following anecdote may help to illustrate the motivational challenges of some of the students in the class. Toward the end of the course, the most disruptive, disengaged student from Group 1 was scrambling to make some progress on his videogame project, which was far behind that of most of his classmates. A disengaged student from Group 2 was overheard chastising him, "Why are you doing work? It's only a summer program. It doesn't count." Classroom-level or even program-level changes, such as changes to the design of learning activities and better communication of expectations, would be needed to combat such extreme lack of motivation.

Occurrences of off-task behavior taking place away from the computer were considerably more prevalent in this study than in the previous classroom study. I wonder if this was due, at least in part, to the fact that students were left in a separate room by themselves while they

worked on the tabletop activities, whereas the tabletop had been in the same room as the teacher and the rest of the class in the previous classroom study. Off-task students in the previous study typically continued to fiddle with the screen, possibly to appear to be working as the teacher was always nearby. In this study, however, there were two closed doors between the tabletop group and the teacher and therefore no pressure to appear to be working for the benefit of the teacher. Additionally, students in the breakout room could easily use their smartphones, which were banned by the educational program, without fear of getting caught. The tabletop was put in a separate breakout room because there was no room for it in the main classroom. This breakout room was initially seen as a positive because it meant that the audio recordings of the group would be a lot clearer than they had been in the earlier study, where the substantial background noise that goes hand in hand with small group collaboration in classroom settings made the audio very difficult to work with. However, an unintentional side effect of the breakout room may have been that it was a lot easier for students to be completely off task.

Finally, the main limitation of this study was its small size. With only three groups, each of which was quite different in collaboration ability and approach, it is difficult to make meaningful comparisons between groups in the intervention and non-intervention conditions. Related to this was the need to run a pilot with the same students that would be part of the experimental evaluation, simply because there were not enough students to have an entirely separate pilot group. Therefore, this study should serve as an exploration of the potential of adaptations to support more effective collaboration rather than a definitive assessment of their usefulness.

8.3 CONCLUSION

In this chapter, I presented eleven design ideas for tabletop adaptations to encourage effective collaboration when problems are detected. Of these eleven ideas, I implemented five for evaluation in a classroom setting with high-school students. Presented as a sequence, which grows increasingly restrictive if collaboration does not improve, four of the five adaptations were used in this study. These adaptations showed promise as supports for groups that struggle to work together effectively.

Group 1, which experienced the adaptations in three out of four sessions at the tabletop had many more episodes of high-quality collaborative work when the adaptations were present than when they were not. The adaptations helped to reduce disruptive behavior from individual students and encouraged engaged students to revisit the task instructions and their plan for completing it. In the pilot session, in which the sequence progressed to the second-to-last adaptation, the adaptations acted as a kind of reset button, forcing the group to interact with the tabletop in a manner that was conducive to effective collaboration. In later sessions, a simple prompt was enough to deter disruptive behavior.

This approach is not without its limitations, however—namely, that problematic behavior that takes place entirely away from the computer, such as off-task students chatting or engaging with their smartphones rather than the assigned task, is undetectable using only touch information. Also undetectable were episodes where a single student was doing all of the work. Further study is needed to determine the extent of this approach's effectiveness, but overall, this study, despite its small scale, demonstrates that tabletop applications that can detect and adapt to poor-quality collaboration can encourage more effective group work.

Chapter 9. CONCLUSION

The goal of this dissertation has been to investigate how tabletop computers can better support high school students' collaborative learning by detecting collaboration problems and adapting to encourage more effective small group work. Through lab and field studies of collaborative learning at tabletop computers (Chapter 4), I have uncovered application-independent touch patterns associated with the quality of collaboration. To make it possible to use these touch patterns to detect collaboration quality in real time, I developed Group Touch (Chapter 5 & 6), an approach to distinguishing among tabletop users that has additional applications beyond collaborative learning. Finally, I designed, implemented, and evaluated a set of software adaptations that were shown to deter disruptive behavior and reduce the duration of periods of low-quality collaboration (Chapters 7 & 8).

Tabletop computers are still an emerging technology and are uncommon in workplaces, let alone classrooms. Much prior work has touted the benefits they could bring to collaborative learning (e.g., Buisine, Besacier, Aoussat, & Vernier, 2012; Dillenbourg & Evans, 2011; Higgins, Mercier, Burd, & Hatch, 2011), though studies of tabletop computers in authentic classrooms caution that many of the challenges of face-to-face collaboration still exist with tabletops (Kharrufa et al., 2013), a finding echoed by the classroom studies described in this dissertation. Investigating ways to address some of these challenges before tabletops are widely adopted will hopefully help to make sure that, when tabletops do arrive in the classroom, they are able to better support collaborative learning.

To reach this goal, I set out to develop an approach to supporting collaborative learning that would work well within the constraints of a typical classroom and not require additional effort on the part of a teacher. This meant that a primary design requirement taken on by this

dissertation work was that it should not rely on sensors external to the tabletop hardware to detect or respond to collaboration problems. This requirement sets my dissertation work apart from prior research on detecting collaboration quality (Ackad et al., 2012; Clayphan et al., 2013; Martinez-Maldonado et al., 2013; Martinez, Kay, et al., 2011; Martinez-Maldonado et al., 2012), which has utilized external sensors such as microphones and depth cameras.

In this chapter, I summarize the major findings of this dissertation, reflect on some of the challenges and limitations of this work, as well as the contributions to human-computer interaction.

9.1 SUMMARY OF FINDINGS

9.1.1 *Modeling Collaboration at Tabletop Computers*

To provide adaptive support for effective collaborative learning at a tabletop computer, the computer must first be able to detect when support is needed. Prior to my dissertation work, a small number of studies had investigated ways to distinguish high- and low-achieving groups by monitoring their task-specific interactions (Martinez-Maldonado et al., 2013; Martinez-Maldonado, Yacef, & Kay, 2013; Martinez, Kay, et al., 2011). Other studies used external sensors to quantify individual group members' verbal and physical contributions to a tabletop activity as a proxy for equitable participation and therefore quality of collaboration (Clayphan et al., 2013; R. Martinez, Collins, et al., 2011). I took a different approach to the problem of modeling collaboration, aiming to uncover task-independent touch patterns that reflected quality rather than quantity of participation.

Drawing on prior work that demonstrated that the verbal and physical aspects of collaboration at tabletop computers are often closely coupled (Fleck et al., 2009; Pontual Falcão & Price, 2010), I conducted two studies to find out if patterns in a group's physical interactions

with a tabletop could reveal the quality of their collaboration. I used the learning sciences concept of social regulation, the task-independent processes that groups use to manage their shared learning, and an existing framework that described social regulation processes in detail (Rogat & Linnenbrink-Garcia, 2011), to identify high- and low-quality collaboration in both studies. In the first study, small groups of adults in a lab setting worked on a poetry analysis task at a Microsoft PixelSense using software I built specifically for the study. The products of this lab study were (1) a simple heuristic for segmenting streams of raw touches by individual group members into sequences representing “purposeful actions” in the software, and (2) a set of touch patterns that reflected quality of collaboration.

In the second study, small groups of high school students in a classroom setting worked on four purpose-built PixelSense applications as part of a course on user-centered design. The purpose of this study was to determine if the touch patterns identified in the lab study with adults would apply with younger students using different applications. Additionally, the touch patterns from the lab study reflected quality of collaboration when the start and end of episodes of particular forms of collaboration were known based on post-hoc analysis of videos of groups collaborating. Therefore, in the second study, I also aimed to develop an approach to using the touch patterns to detect collaboration quality in real time. The outcomes of this study were (1) revised touch patterns that could detect quality of collaboration with up to 84.2% accuracy, and (2) an approach to detecting collaboration quality in real time by checking touch patterns for two-minute intervals of touch data.

The touch patterns finalized in the classroom study, tested by four groups of students using three distinct applications, are largely application-independent although they do require a simple taxonomy defining which on-screen objects are related to each other in the context of the

activity. In the classroom study, these relationships were defined in XML files stored alongside each application. An example is provided in Appendix A. The touch patterns are *Unrelated Touches*, which determines quality based on the frequency of each active user's touches to unrelated objects, and *Overlapping Unrelated Sequences*, which determines if multiple people are interacting with the screen at the same time and whether they are interacting with related or unrelated objects.

9.1.2 *Distinguishing Users at a Tabletop Computer*

Most commodity tabletops, such as the Microsoft PixelSense, are unable to distinguish among users—a major barrier to using the touch patterns described in the previous section to detect collaboration quality. Although numerous solutions to the problem of distinguishing users had been proposed prior to my dissertation work, these approaches either rely on external sensors (Ackad et al., 2012; Annett, Grossman, Wigdor, & Fitzmaurice, 2011; Clayphan et al., 2013; Paul Jermann, Zufferey, & Dillenbourg, 2008; Maekawa, Masuda, & Namioka, 2016; Marquardt, Kiemer, Ledo, Boring, & Greenberg, 2011; R. Martinez, Collins, et al., 2011; Meyer & Schmidt, 2010; Ramakers, Vanacken, Luyten, Coninx, & Schoning, 2012; Richter, Holz, & Baudisch, 2012; Roth, Schmidt, & Benjamin, 2010; Tanase, Vatavu, Pentiu, & Graur, 2008) or constrain users in some manner (Dang et al., 2009; Ewerling, Kulik, & Froelich, 2012; Garcia-sanjuan, Jaen, & Catala, 2013; Zhang et al., 2012; Zhang et al., 2014). Therefore, the existing approaches were not suitable for use in an authentic classroom environment. I set out to develop a new, lightweight approach that could distinguish among users without requiring additional sensors or constraining users.

Most existing approaches to distinguishing users aimed to track or even identify individual tabletop users for the duration of an activity. Although these capabilities have

powerful use cases, such as enabling sophisticated personalization and assessment of individual users' contributions to group work, they were not required to enable real-time use of the touch patterns for detecting collaboration quality. Therefore, I relaxed the goal of tracking users and instead focused on distinguishing multiple users as they are simultaneously interacting with the tabletop. Additionally, where previous approaches sought to match touches to individual users, I reframed the problem to compare touches to each other to determine if they were carried out by the same person or different people.

My approach, known as Group Touch, uses a multilayer perceptron machine learning model to predict whether a pair of touches was carried out by the same person based on the pixel distance between the touches, the difference in their orientations, and the time elapsed between the touches. The model was trained using leave-one-out cross validation and data collected from six groups of high school students in two different educational programs using five distinct multi-touch applications on a Microsoft PixelSense. As well as predicting whether a pair of touches was carried out by the same person, the model outputs the probability that they were by the same person. Group Touch uses this probability to assign touches to groups of touches likely to have been carried out by the same person—a new touch is assigned to the group with the highest probability that the new touch and the last touch in the group are by the same person, as long as that probability is above a minimum threshold. If no existing groups meet that minimum threshold, a new touch group is started.

I determined 0.8 to be the optimal value of the probability threshold for general-purpose use cases by testing threshold values ranging from 0.5 to 0.9 and comparing the overall accuracy (percentage of touches assigned to a group where the previous touch belonged to the same user), the median duration of touch groups, and the median time between groups belonging to the same

user at each threshold. At 0.8, Group Touch had an overall accuracy of 92.92%, a median group duration of 9,094 ms, and a median time between groups of 4001.5 ms.

9.1.3 *Using Group Touch to Detect Collaboration Quality in Real Time*

Although 0.8 was determined to be the optimal value of Group Touch's probability threshold for general purposes, I conducted a further analysis to determine the optimal threshold value for each of the two touch patterns used to detect collaboration quality. This analysis used the touch data and the touch pattern outputs from the classroom study. For each two-minute interval in the classroom study, I compared the touch pattern output using Group Touch to the pattern output using touches hand-labeled with their author based on the video recordings of each session (the pattern outputs reported in Section 4.2.3).

Each of the touch patterns uses one or two metrics to distinguish high-, medium-, and low-quality collaboration. For Unrelated Touches, that metric is the percentage of unrelated touches in sequences representing purposeful actions, and for Overlapping Unrelated Sequences, the metrics are the percentage of touches carried out by multiple people interacting with the screen simultaneously and the percentage of those touches that involved unrelated objects. To determine the optimal Group Touch probability threshold for each pattern, I calculated the differences between the pattern metrics generated using the hand-labeled touch data and those generated using Group Touch with probability thresholds ranging from 0.5 to 0.9. The probability thresholds that produced the smallest differences in the metrics were selected—0.5 for Unrelated Touches and 0.8 for Overlapping Unrelated Sequences. At these thresholds, the quality labels assessed by the patterns using Group Touch matched those generated using the hand-labeled touch data for 84.85% of intervals in the case of Unrelated Touches and 77.27% of intervals for Overlapping Unrelated Sequences.

The final step in this analysis was to compare the touch pattern outputs of high, medium, or low-quality to the collaboration quality in the videos of each group session. When both touch patterns assessed an interval as high-quality, the video analysis showed that the collaboration was skewed toward positive social regulation codes, and when both touch patterns assessed an interval as low-quality, the collaboration was skewed toward negative codes. Intervals that produced all other combinations of touch pattern quality labels (e.g., high-quality according to one pattern and low-quality according to the other) were also skewed toward negative codes, though intervals with each possible combination were few in number. Overall, this analysis showed that I could be confident that, when both touch patterns gave an interval the same label of high or low-quality using Group Touch, the label was a reasonable reflection of what was occurring among the students.

9.1.4 *Adaptations to Encourage Effective Collaboration*

My research up to this point had focused on inferring the quality of a group's collaboration processes in real time using only touch data. The culminating stage of my dissertation research was to use this new ability to detect collaboration quality, specifically breakdowns in collaboration, to enable the computer to adapt to provide just-in-time support for groups struggling to work together effectively.

I designed, implemented, and evaluated a set of software adaptations in a classroom deployment with high school students. I proposed 11 potential adaptations to encourage and selected four to implement: (1) a *group awareness icon*, an always-present icon that changed color to reflect the collaboration quality detected by the touch patterns; (2) *control lockout*, which disabled certain controls; (3) *voting*, which required group members to vote on any actions that affected the global state of the application; and (4) *escalate to authority*, which would

quietly summon the teacher via Slack message. Adaptations 2 - 4 were designed to trigger in sequence, with each successive adaptation layering on top of existing adaptations, adding further restrictions to how the group could interact with the computer. These adaptations could be reversed if the touch patterns detected consistently high-quality collaboration.

The evaluation of the adaptations began with baseline data collection, in which a survey of students' attitudes toward group work was conducted and each group used the tabletop without intervention. Analysis of the baseline data showed each group to be quite different in working style.

The adaptations were then piloted with one group of students. In the pilot session, *control* *lockout* and *voting* were both triggered on two occasions. Although collaboration did technically improve after the adaptations were triggered, the improvements were superficial and the students were clearly frustrated by the restrictions. To remedy these issues, I added a gentler adaptation—the *prompt*—to the start of the sequence and reduced the length of time that a group had to sustain high-quality collaboration to get adaptations removed.

Finally, in the experimental phase of the study, groups used the same tabletop application across two sessions. One group was assigned to an experimental condition and used the tabletop with adaptations in both sessions. Another group was assigned to the control condition and used the tabletop without adaptations in both sessions. The third group was assigned to a split condition, using the tabletop without adaptations in the first session and with adaptations in the second session. The experimental group triggered the *prompt* three times across the two sessions. Subsequent adaptations in the sequence were not triggered because collaboration improved immediately after the prompt in each case. The group in the split condition did not trigger the adaptations.

The improvements that followed the adaptations were (1) reduced task disruption from disengaged and/or off-task students, allowing engaged students to make progress, and (2) a shortening of periods of poor-quality collaboration involving the tabletop. However, this study also highlighted two significant limitations of my approach to detecting collaboration quality: (1) negative collaboration behavior taking place away from the tabletop is undetectable, and (2) episodes where a single student does all the work while the others do not engage with the task or the tabletop look very much like high-quality collaboration to the touch patterns.

9.2 REFLECTIONS AND DISCUSSION

This dissertation has led to the development of methods for detecting collaboration quality and distinguishing among simultaneous tabletop users with only the information provided by the hardware's built-in sensors, and a set of task-independent adaptations that can help students to work together effectively. However, the major design requirement of this work—the use of built-in sensors only—likely limited the effectiveness of the adaptations in the implementation described in this dissertation. It has already been noted that the main limitation of my method for detecting collaboration quality is that it is unable to detect problematic behavior that happens away from the computer. To detect that some students in a group are not interacting with the computer, it would be important to actually track individual students in the manner that is possible when using external sensors such as a depth camera to distinguish among users (Ackad et al., 2012; Clayphan et al., 2013; R. Martinez, Collins, et al., 2011). Despite the enhanced tracking capabilities that external sensors could bring, I maintain that foregoing cumbersome additional equipment, which requires additional setup, calibration, and maintenance, is particularly important in the case of classroom technologies.

I anticipate that in the future, the built-in sensing capabilities of tabletop computers will likely improve to the point that it is possible to track individual students without external sensors. When that happens, Group Touch, my approach to distinguishing among tabletop users, will no longer be needed, but the touch patterns I developed to detect collaboration quality will still be relevant and likely more effective. A major challenge in this work has been the multiple steps between the fundamental problem—that students do not always have the skills needed to collaborate effectively—and the goal of building software that can adapt to scaffold effective collaboration. With each step, realizing the goal becomes a little more removed from what is actually happening in a collaborative activity, and the precision with which it is possible to detect collaboration quality decreases. The first decrease in precision occurs because the tabletop computer is only able to detect the physical elements of a collaboration—verbal interactions are lost and a representation of the whole collaboration must be inferred from touch data alone. The second decrease in precision occurs because it is not possible to automatically segment the touch data into meaningful episodes in the manner that is possible after the fact with video analysis—artificial time-based segmentation (i.e., splitting the touch data into overlapping two-minute intervals) is instead used to approximate these types of episodes. A third and final decrease in precision occurs when using Group Touch to distinguish among users because it is not able to track users for extended periods of time and therefore introduces more noise into the collaboration quality detected by the touch patterns. It is my hope that by the time tabletop computers become more widely used, their built-in sensing capabilities will have improved to the point that Group Touch is no longer needed, and my approach to detecting collaboration quality will be two steps removed from the actual collaboration instead of its current three steps.

Although my approach to detecting collaboration quality has been well tested and honed across multiple studies and multiple settings, the same cannot be said for the adaptations intended to scaffold effective collaboration. Due to the small size of the study evaluating the adaptations, only one group of students had real exposure to the adaptations, albeit over multiple sessions and with positive outcomes. I believe that the results of the final stage of my dissertation work demonstrate that the adaptations have promise for encouraging certain positive collaborative behaviors, but more extensive evaluation would be needed to determine if these results are widely generalizable. This evaluation would require additional iterations of design and classroom deployments, preferably with larger class sizes to allow for more groups of students.

I also believe, however, that it would be extremely difficult to conduct further research at the scale needed to quantitatively determine the extent of the adaptations' effectiveness given the resources currently available. Conducting classroom-based research with a single tabletop computer posed substantial practical challenges for curriculum planning and classroom management, particularly in the final study. With a single tabletop computer, only one group can work with it at a time, and the teacher must come up with meaningful activities for the rest of the class, rotating groups around activities until all groups have had their session at the tabletop. In practice, this means that the non-tabletop activity (or activities) must last for the length of the tabletop session multiplied by the number of groups minus one. For example, in the first classroom study, tabletop sessions lasted around 15 minutes and there were four groups of students. For each tabletop activity, 45 minutes of additional activities were needed to occupy the students who were not at the tabletop, and it took 60 minutes in total for all groups to complete the 15-minute tabletop activity. In the final study, there were three groups, and tabletop sessions

lasted up to 25 minutes, meaning that 50 minutes of additional activities were needed per tabletop activity, and it took 75 minutes to get all groups through a single tabletop activity. As the length of the tabletop session increases, or the class size (and therefore the number of groups) increases, the amount of time needed to rotate all groups around a single activity increases rapidly to the point that it can take days of class time for each group to complete a single tabletop session. In addition to the amount of time that the non-tabletop activities need to fill, they also need to meet certain constraints, namely that none of the activities, tabletop or non-tabletop, can be pre-requisites for each other as each group will complete them in a different order. Given these challenges, any future research seeking more quantitative evidence of the adaptations' ability to scaffold effective collaboration would greatly benefit from the use of more than one tabletop computer, preferably one computer per group. Currently, the cost of acquiring enough tabletop computers to support a whole classroom would be prohibitive. It is my hope that the cost of tabletops and other large multi-touch surfaces will decrease over time; it is unlikely that tabletops will find their way into classrooms otherwise.

The high cost of tabletop computers is at least partly due to their status as an emerging technology. The fact that these devices are not yet established classroom technology is one of the reasons that I became interested in them. I believe it is important to consider how new technology can best support teaching and learning *before* it arrives in classrooms instead of waiting until after it has arrived and been found wanting. However, the newness of the technology also meant that, although the goal of my dissertation research was to develop ways to enhance collaborative educational software, that educational software didn't exist, so I had to create it myself to make it possible to address the questions I was interested in. Although this dissertation is of course focused on the contributions specifically relevant to my research

questions and thesis statement—touch patterns for detecting collaboration quality, an approach to distinguishing simultaneous users at tabletop computers, and a set of interface adaptations to encourage effective collaboration—I also want to acknowledge the tremendous amount of time and effort that went into designing and building the educational technology applications that were needed to gather the data to develop and evaluate each of those contributions.

In total, I designed and built 12 separate learning applications (see Appendix C). Of those 12 applications, 9 were used in data collection and described in this dissertation. Of the remaining 3 applications, 2 were built for use in a particular after-school software program but were never used to collect data because a change in program staff led to a radical change in curriculum and my applications were no longer useful. The third unused application was intended for use in the final study but was also dropped after a change in curriculum meant it was no longer relevant to the course objectives. Most of the applications took weeks to build but some took months and one took about a year, not including the year it took to build its original, almost fully built-out version that had to be abandoned and rebuilt in a new platform after a technical hitch was discovered. In total, I estimate that building these supporting applications took approximately three years, more than half the time spent on the dissertation as a whole.

9.3 CONTRIBUTIONS

The contributions of this dissertation are:

1. Insight into which aspects of a group's collaborative process at a tabletop computer can be inferred using only touch data.
2. A heuristic for segmenting an individual user's touches in sequences representing purposeful actions in tabletop software.

3. A set of touch patterns reflecting the quality of high school students' collaborative learning processes.
4. A field evaluation of an existing approach to distinguishing among users that highlights some of the challenges specific to field settings.
5. Group Touch—a field-tested, machine learning-based approach to distinguishing among users working at a vision-based interactive tabletop that does not restrict users' multi-touch interaction or require additional sensors.
6. A method for using the touch patterns (contribution #3) to model tabletop collaborative learning processes and detect breakdowns in collaboration in real-time.
7. A set of interface adaptations that respond to detected breakdowns to encourage more effective collaboration.
8. A field evaluation of these adaptations that demonstrates the impact they have on high school students' tabletop collaboration.

9.4 FUTURE WORK

Avenues for future work exist in all three areas of this dissertation: improving the detection of collaboration quality, improving my approach for distinguishing among users, and further evaluation and development of the adaptations.

The final field study demonstrated that, although my approach to detecting collaboration quality was successful at detecting disruptive behavior at the tabletop, the problem of disengaged students who were completely off-task and away from the computer went undetected. Related to this, episodes in which a single student did all the work while the others sat back were also undetected. Future work focused on developing ways to detect these types of behavior would be valuable. Instances where the whole group has stopped working are already detectable, but it is

not currently possible to identify when only some students are disengaged. I remain committed to using only the computer's built-in sensors to detect collaboration quality, which are currently limited to the infrared cameras used for touch detection. It is possible that future generations of tabletop hardware will include additional sensors that can support more sophisticated detection, such as directional microphones capable of capturing students' verbal interactions even in noisy classrooms. Improved capabilities of distinguishing among users would likely also be needed to address this problem.

Currently, Group Touch is only able to distinguish among simultaneous users for short periods of time and it is unable to detect when individual students stop interacting with the screen altogether. Future work should explore whether adding additional features to Group Touch's MLP model or the grouping algorithm could improve accuracy and enable it to distinguish among simultaneous users for longer periods of time. One area to explore would be the effect of combining Group Touch with a technique that uses hand geometry to map touches by different fingers to hands (Dang et al., 2009). Another area to explore would be if additional measures such as the density of touches on screen could help to identify when only some group members are interacting with the screen.

Given the small size of my final study, further research is needed to confirm the effectiveness of the adaptations intended to scaffold effective collaboration. Additionally, the evaluation of the adaptations showed that they appeared to deter disruption, allowing engaged students to collaborate more easily, but they didn't increase effective collaboration behaviors in those students causing the disruption. Future work should explore ways to not only deter disruptive behavior, but also help disruptive students to re-engage with the learning material.

Eleven adaptations were proposed but only five were built. It would be interesting to implement the remaining six adaptations and compare their effectiveness to those already tested. Future work may also wish to refine some of the details of the current implementation, such as the length of the intervals and the order in which adaptations are triggered.

Once the adaptations are refined, the next step would be to conduct a long-term study exploring the impact of using the adaptations over time on students' ability to collaborate. Ideally, the adaptations should help students to reflect on their own contributions to small group work and learn to avoid behaviors that are problematic for collaboration. Over time, the adaptations should be needed less and less.

9.5 FINAL REMARKS

The aim of this dissertation was to demonstrate the following thesis statement:

Interactive tabletop software that can automatically detect breakdowns in collaboration and adapt in real-time to scaffold effective social regulation can improve secondary school students' collaboration skills.

My approach to detecting collaboration quality using only touch patterns has been shown to be effective at detecting breakdowns in collaboration in two field studies with students using a range of applications. The impact of the software adaptations implemented in this dissertation is less definitive though still promising. The group of students that experienced the adaptations had reduced disruption and shorter periods of low-quality collaboration when the adaptations were present. Further work is needed to determine the generalizability of the adaptations and their long-term impact on students' collaboration skills.

BIBLIOGRAPHY

- Ackad, C. J., Clayphan, A., Maldonado, R. M., & Kay, J. (2012). Seamless and continuous user identification for interactive tabletops using personal device handshaking and body tracking. In *Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)* (pp. 1775–1780). New York: ACM.
- Aldunate, R., & Nussbaum, M. (2013). Teacher adoption of technology. *Computers in Human Behavior, 29*(3), 519–524.
- Annett, M., Grossman, T., Wigdor, D., & Fitzmaurice, G. (2011). Medusa: a proximity-aware multi-touch tabletop. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '11)* (pp. 337–346). New York: ACM.
- Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307–359.
- Blazica, B., Vladusic, D., & Mladenic, D. (2013). MTi: a method for user identification for multitouch displays. *International Journal of Human-Computer Studies, 71*, 691–702.
- Buisine, S., Besacier, G., Aoussat, A., & Vernier, F. (2012). How do interactive tabletop systems influence collaboration? *Computers in Human Behavior, 28*(1), 49–59.
- CAST. (2011). Universal design for learning guidelines version 2.0. Wakefield, MA: CAST.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2011). SMOTE: Synthetic minority over-sampling technique. *Journal Of Artificial Intelligence Research, 16*, 321–357.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: a practical guide. *Journal of the Learning Sciences, 6*(3), 271–315.
- Clayphan, A., Martinez-Maldonado, R., Ackad, C., & Kay, J. (2013). An approach for designing and evaluating a plug-in vision-based tabletop touch identification system. In *Proceedings of the Australian Computer-Human Interaction Conference (OzCHI '13)* (pp. 373–382).
- Creswell, J. W. (2003). *Research design: qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA.: Sage Publications.
- Dang, C. T., Straub, M., & André, E. (2009). Hand distinction for multi-touch tabletop interaction. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '09)* (pp. 101–108). New York: ACM.

- De Backer, L., Van Keer, H., & Valcke, M. (2015). Exploring evolutions in reciprocal peer tutoring groups' socially shared metacognitive regulation and identifying its metacognitive correlates. *Learning and Instruction, 38*, 63–78.
- Dietz, P., & Leigh, D. (2001). DiamondTouch: a multi-user touch technology. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '01)* (pp. 219–226). New York: ACM.
- Dillenbourg, P. (1999). *Collaborative learning: cognitive and computational approaches*. New York: Pergamon.
- Dillenbourg, P., & Evans, M. (2011). Interactive tabletops in education. *International Journal of Computer-Supported Collaborative Learning, 6*(4), 491–514.
- Dillenbourg, P., & Jermann, P. (2007). Designing Integrative Scripts. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives*. New York: Springer.
- Do-Lenh, S. (2012). *Supporting Reflection and Classroom Orchestration with Tangible Tabletops*. École Polytechnique Federale De Laussane.
- Evans, A. C., Wobbrock, J. O., & Davis, K. (2016). Modelling collaboration patterns on an interactive tabletop in a classroom setting. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '16)* (pp. 860–871). New York: ACM.
- Ewerling, P., Kulik, A., & Froelich, B. (2012). Finger and hand detection for multi-touch interfaces based on maximally stable extremal regions. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)* (pp. 173–182). New York: ACM.
- Facing the Future, & Snow Leopard Trust. (2009). *Engaging Students in Conservation: Protecting the Endangered Snow Leopard*. Seattle, WA. Retrieved from <http://www.snowleopard.org/learn/resources-for-educators>
- Findlater, L., Wobbrock, J. O., & Wigdor, D. (2011). Typing on flat glass: examining ten-finger expert typing patterns on touch surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)* (pp. 2453–2462). New York: ACM.
- Fleck, R., Rogers, Y., Yuill, N., Marshall, P., Carr, A., Rick, J., & Bonnett, V. (2009). Actions speak loudly with words: unpacking collaboration around the table. In *Proceedings of the*

- ACM International Conference on Interactive Tabletops and Surfaces (ITS '09)* (pp. 189–196). New York: ACM.
- Frost, R. (1969). Birches. In *The Poetry of Robert Frost*. New York: Henry Holt and Co.
- Garcia-sanjuan, F., Jaen, J., & Catala, A. (2013). Evaluating heuristics for tabletop user segmentation based on simultaneous interaction. *Expert Systems With Applications*, 40(14),
- Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14/15), 2627–2636.
- Hadwin, A. F., Oshige, M., Gress, C. L. Z., & Winne, P. H. (2010). Innovative ways for using gStudy to orchestrate and research social aspects of self-regulated learning. *Computers in Human Behavior*, 26(5), 794–805.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1), 10–18.
- Higgins, S. E., Mercier, E., Burd, E., & Hatch, A. (2011). Multi-touch tables and the relationship with collaborative classroom pedagogies: a synthetic review. *International Journal of Computer-Supported Collaborative Learning*, 6(4), 515–538.
- Hinrichs, U., & Carpendale, S. (2011). Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)* (pp. 3023–3032). New York: ACM.
- Holz, C., & Baudisch, P. (2013). Fiberio: a touchscreen that senses fingerprints. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '13)* (pp. 41–50). New York: ACM.
- Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research*, 81, 11–24.
- Järvelä, S., & Hadwin, A. F. (2013). New frontiers: regulating learning in CSCL. *Educational Psychologist*, 48(1), 25–39.
- Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., Järvenoja, H. (2014). Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools. *Educational Technology Research and Development*, 63(1), 125–142.

- Järvelä, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially shared regulation by using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction, 42*, 1–11.
- Järvenoja, H., & Järvelä, S. (2011). Socially constructed self-regulated learning and motivation regulation in collaborative learning groups. *Teachers College Record, 113*(2), 350–374.
- Järvenoja, H., Järvelä, S., & Malmberg, J. (2015). Understanding regulated learning in situative and contextual frameworks. *Educational Psychologist, 50*(3), 204–219.
- Jermann, P., Muhlenbrock, M., & Soller, A. (2002). Designing computational models of collaborative learning interaction. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community* (pp. 730–732). Boulder, Colorado: International Society of the Learning Sciences.
- Jermann, P., Zufferey, G., & Dillenbourg, P. (2008). Tinkering or sketching: apprentices' use of tangibles and drawings to solve design problems. In *Proceedings of the European conference on Technology Enhanced Learning (EC-TEL '08)* (pp. 167–178). Berlin.
- Kempler Rogat, T., Linnenbrink-Garcia, L., & DiDonato, N. (2013). Motivation in collaborative groups. In C. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, & A. O'Donnell (Eds.), *The International Handbook of Collaborative Learning* (pp. 250–267). Routledge.
- Kharrufa, A., Balaam, M., Heslop, P., Leat, D., Dolan, P., & Olivier, P. (2013). Tables in the wild: lessons learned from a large-scale multi-tabletop deployment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)* (pp. 1021–1030). New York: ACM.
- Khosa, D. K., & Volet, S. E. (2014). Productive group engagement in cognitive activity and metacognitive regulation during collaborative learning: can it explain differences in students' conceptual understanding? *Metacognition and Learning, 9*(3), 287–307.
- Kin, K., Agrawala, M., & DeRose, T. (2009). Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In *Proceedings of Graphics Interface (GI '09)* (pp. 119–124). Toronto: Canadian Information Processing Society.
- Klinkhammer, D., Nitsche, M., Specht, M., & Reiterer, H. (2011). Adaptive personal territories for co-located tabletop interaction in a museum setting. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '11)* (pp. 107–110). New York: ACM.

- Kouros, C., & Abrami, P. C. (2006). How do students really feel about working in small groups? The role of student attitudes and behaviours in cooperative classroom settings. In *Annual Meeting of the American Educational Research Association* (pp. 1–56).
- Kumar, R., & Rosé, C. P. (2011). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, 4(1), 21–34.
- Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Proceeding of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, 383–390.
- Lajoie, S. P., & Lu, J. (2012). Supporting collaboration with technology: Does shared cognition lead to co-regulation in medicine? *Metacognition and Learning*, 7(1), 45–62.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1).
- Lee, A., O'Donnell, A. M., & Rogat, T. K. (2015). Exploration of the cognitive regulatory sub-processes employed by groups characterized by socially shared and other-regulation in a CSCL context. *Computers in Human Behavior*, 52, 617–627.
- Lonchamp, J. (2010). Customizable computer-based interaction analysis for coaching and self-regulation in synchronous CSCL systems. *Journal of Educational Technology & Society*, 13(2), 193–205.
- Maekawa, T., Masuda, A., & Namioka, Y. (2016). Identification from ceiling: unconstrained person identification for tabletops using multiview learning. *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, 273–284.
- Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior*, 52, 562–572.
- Marcos-Garcia, J. A., Martinez-Mones, A., Dimitriadis, Y., & Anguita-Martínez, R. (2007). A role-based approach for the support of collaborative learning activities. *E-Service Journal*, 6(1).
- Marquardt, N., Kiemer, J., Ledo, D., Boring, S., & Greenberg, S. (2011). Designing user-, hand-, and handpart-aware tabletop interactions with the TouchID toolkit. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '11)* (Vol. 8, pp.

- 21–30). New York: ACM.
- Martinez-maldonado, R., Kay, J., Yacef, K., Edbauer, M. T., & Dimitriadis, Y. (2013). MTClassroom and MTDashboard: supporting analysis of teacher attention in an orchestrated multi-tabletop classroom. In *Proceedings of the International Conference on Computer-Supported Collaborative Learning (CSCL '13)* (pp. 320–327). Madison, WI: ISLS.
- Martinez-Maldonado, R., Yacef, K., & Kay, J. (2013). Data mining in the classroom: discovering groups' strategies at a multi-tabletop environment. In *6th International Conference on Educational Data Mining*. Memphis.
- Martinez, M. (2001). Key design considerations for personalized learning on the Web. *Journal of Educational Technology and Society*, 4(1), 26–40.
- Martinez, R., Collins, A., Kay, J., & Yacef, K. (2011). Who did what? Who said that?: Collaid: an environment for capturing traces of collaborative learning at the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '11)* (pp. 172–181). New York: ACM.
- Martinez, R., Kay, J., & Yacef, K. (2011). Visualisations for longitudinal participation, contribution and progress of a collaborative task at the tabletop. In *Proceedings of the International Conference on Computer-Supported Collaborative Learning (CSCL '13)* (pp. 25–32). Hong Kong, China: ISLS.
- Martinez Maldonado, R., Kay, J., Yacef, K., & Schwendimann, B. (2012). An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In *Proceedings of the International Conference on Intelligent Tutoring Systems (ITS '12)* (pp. 482–492). Berlin, Heidelberg: Springer-Verlag.
- McMahan, E., Funk, R., & Day, S. (1998). *The Elements of Writing About Literature and Film*. Longman Publishing.
- Meyer, T., & Schmidt, D. (2010). IdWristbands: IR-based user identification on multi-touch surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)* (pp. 277–278). New York: ACM.
- Morris, M. R., Huang, A., Paepcke, A., & Winograd, T. (2006). Cooperative gestures: multi-user gestural interactions for co-located groupware. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '06)*. New York: ACM.

- Morris, M. R., Ryall, K., Shen, C., Forlines, C., & Vernier, F. (2004). Beyond “social protocols”: multi-user coordination policies for co-located groupware. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 1–4). New York: ACM.
- Muller-Tomfelde, C., & Fjeld, M. (2012). Tabletops: Interactive Horizontal Displays for Ubiquitous Computing. *Computer*, 45(2), 78–81.
- National Governors Association Center for Best Practices. (2010). *Common Core State Standards*. Washington, DC.
- National Research Council. (2015). *Guide to Implementing the Next Generation Science Standards*. Washington, D.C.
- Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods*. New York, NY: John Wiley & Sons.
- Panadero, E., & Järvelä, S. (2015). Socially shared regulation of learning: a review. *European Psychologist*, 20(3), 190–203.
- Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, 13(3), 423–451.
- Piper, A. M., O’Brien, E., Morris, M. R., & Winograd, T. (2006). SIDES: a cooperative tabletop computer game for social skills development. In *Conference on Computer Supported Cooperative Work (CSCW '06)* (pp. 1–10). New York: ACM.
- Pontual Falcão, T., & Price, S. (2010). Interfering and resolving: how tabletop interaction facilitates co-construction of argumentative knowledge. *International Journal of Computer-Supported Collaborative Learning*, 6(4), 539–559.
- Ramakers, R., Vanacken, D., Luyten, K., Coninx, K., & Schoning, J. (2012). Carpus: a non-intrusive user identification technique for interactive surfaces. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '12)* (pp. 35–44). New York: ACM.
- Rappolt-Schlichtmann, G., Daley, S. G., & Rose, L. T. (2012). *A research reader in universal design for learning*. Cambridge, MA: Harvard Education.
- Richter, S., Holz, C., & Baudisch, P. (2012). Bootstrapper: recognizing tabletop users by their shoes. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (pp. 1249–1252). New York: ACM.

- Rick, J., Harris, A., Marshall, P., Fleck, R., Yuill, N., & Rogers, Y. (2009). Children designing together on a multi-touch tabletop: an analysis of spatial orientation and user interactions. *Proceedings of the International Conference on Interaction Design and Children (IDC '09)*. New York: ACM.
- Rogat, T. K., & Adams-Wiggins, K. R. (2014). Other-regulation in collaborative groups: implications for regulation quality. *Instructional Science*, 42(6), 879–904.
- Rogat, T. K., & Linnenbrink-Garcia, L. (2011). Socially shared regulation in collaborative groups: an analysis of the interplay between quality of social regulation and group processes. *Cognition and Instruction*, 29(4), 375–415.
- Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69–97). Heidelberg, Germany: Springer-Verlag.
- Rose, D. H., & Meyer, A. (2002). *Teaching every student in the Digital Age: universal design for learning*. Alexandria, Va.: Association for Supervision and Curriculum Development.
- Roth, V., Schmidt, P., & Benjamin, G. (2010). The IR ring: authenticating users' touches on a multi-touch display. In *Proceedings of the ACM symposium on User Interface Software and Technology (UIST '10)* (pp. 259–262). New York: ACM.
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, 3(3), 265–283.
- Sinha, S., Rogat, T. K., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning*, 10(3), 273–307.
- Smagorinsky, P. (2008). The method section as conceptual epicenter in constructing social science research reports. *Written Communication*, 25(3), 389–411.
- Stahl, G. (2006). A model of collaborative knowledge building. In *Group cognition: computer support for building collaborative knowledge*. Cambridge, Mass.: MIT Press.
- Tanase, C., Vatavu, R., Pentiu, Ș., & Graur, A. (2008). Detecting and tracking multiple users in the proximity of interactive tabletops. *Advances in Electrical and Computer Engineering*, 8(2), 50–53.
- The Partnership for 21st Century Skills. (2009). *P21 Framework Definitions*.
- Thomas, D. (2003). Fern Hill. In *The Poems of Dylan Thomas*. New York: New Directions

- Publishing Company.
- U.S. Department of Education. (2010). *Transforming American education: Learning powered by technology*. Washington, DC.
- Volet, S., Vauras, M., & Salonen, P. (2009). Self- and social regulation in learning contexts: An integrative perspective. *Educational Psychologist, 44*(4), 215–226.
- Walker, E., Rummel, N., & Koedinger, K. R. (2011). Adaptive support for CSCL: is it feedback relevance or increased student accountability that matters? In *Computer Supported Collaborative Learning* (Vol. 1, pp. 334–341).
- Wang, H. C., Rosé, C. P., & Chang, C. Y. (2011). Agent-based dynamic support for learning from collaborative brainstorming in scientific inquiry. *International Journal of Computer-Supported Collaborative Learning, 6*(3), 371–395.
- Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, 203–211*.
- Wobbrock, J. O. (2013). Improving pointing in graphical user interfaces for people with motor impairments through ability-based design. In G. Kouroupetroglou (Ed.), *Assistive Technologies and Computer Access for Motor Disabilities*. Hershey, PA: IGI Global.
- Wobbrock, J. O., Gajos, K. Z., Kane, S. K., & Vanderheiden, G. C. (2018). Ability-Based Design. *Communications of the ACM*, (June).
- Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S., & Froehlich, J. (2011). Ability-Based Design. *ACM TACCESS, 3*(3), 1–27.
- Wobbrock, J. O., Kane, S. K., Harada, S., Froehlich, J., & Gajos, K. Z. (2011). Ability-based design: concept, principles and examples. *ACM Transactions on Accessible Computing, 3*(3), 27–91.
- Xu, J., Du, J., & Fan, X. (2013). “finding our time”: Predicting students’ time management in online collaborative groupwork. *Computers and Education, 69*, 139–147.
- Zhang, H., Yang, X.-D., Ens, B., Liang, H.-N., Boulanger, P., & Irani, P. (2012). See me, see you: a lightweight method for discriminating user touches on tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)* (pp. 2327–2336). New York: ACM.
- Zhang, Z., Zhang, F., & Chen, H. (2014). Left and right hand distinction for multi-touch tabletop

interactions. In *Proceedings of the 19th international conference on Intelligent User Interfaces (IUI '14)* (pp. 47–56). New York: ACM.

APPENDIX A – EXAMPLE RELATIONSHIP DEFINITION XML

```

<?xml version="1.0"?>
<Elements>
  <Element id="Drag Drop"/>
  <Element id="Embed Video"/>
  <Element id="Mobile Site"/>
  <Element id="Image Slider"/>
  <Element id="Popup Calendar"/>
  <Element id="Background Image"/>
  <Element id="Game"/>
  <Element id="Glossary">
    <Related id="Search Query"/>
    <Related id="Keyboard"/>
  </Element>
  <Element id="Keyboard">
    <Related id="Search Query"/>
    <Related id="Glossary"/>
  </Element>
  <Element id="Search Query">
    <Related id="Keyboard"/>
    <Related id="Search Results"/>
    <Related id="Glossary"/>
  </Element>
  <Element id="Search Results">
    <Related id="Search Query"/>
    <Related id="Search Rating"/>
  </Element>
  <Element id="Search Rating">
    <Related id="Search Results"/>
  </Element>
  <Element id="other"/>
</Elements>

```

APPENDIX B – SURVEY OF STUDENT ATTITUDES TOWARD GROUP WORK

Group Work Survey⁹

Indicate whether or not you agree with the following statements. Your responses are confidential and will not impact your grade.

1. When I work in a group I do better quality work.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

2. When I work in a group I end up doing most of the work.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

3. When I work with other students I am able to work at my own pace.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

4. The work takes longer to complete when I work with other students.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

5. My group members do not respect my opinions.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

6. I enjoy the material more when I work with other students.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

7. My group members help explain things that I do not understand.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

8. When I work in a group I am able to share my ideas.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

9. My group members make me feel that I am not as smart as they are.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

⁹ Adapted from (Kouros & Abrami, 2006)

10. The material is easier to understand when I work with other students.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

11. My work is better organized when I am in a group.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

12. My group members like to help me learn the material.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

13. The workload is usually less when I work with other students.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

14. I feel I am part of what is going on in the group.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

15. One student usually makes the decisions in the group.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

16. I find it hard to express my thoughts when I work in a group.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

17. I try to make sure my group members learn the material.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

18. I learn to work with students who are different from me.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

19. My group members do not care about my feelings.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

20. I let the other students do most of the work.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

21. I feel working in groups is a waste of time.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

22. I have to work with students who are not as smart as I am.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

23. When I work in a group, there are opportunities to express your opinions.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

24. When I work with other students the work is divided equally.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

25. We cannot complete the assignment unless everyone contributes.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

26. I help my group members with what I am good at.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

27. My group members compete to see who does better work.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

28. The material is more interesting when I work with other students.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

29. When I work in a group my work habits improve.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

30. I like to help my group members learn the material.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

31. It is important to me that my group gets the work done on time.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

32. I learn more when I work with other students.

Strongly disagree *Disagree* *Neutral* *Agree* *Strongly agree*

33. It takes less time to complete the assignment when I work with others.

Strongly disagree Disagree Neutral Agree Strongly agree

34. I also learn when I teach the material to my group members.

Strongly disagree Disagree Neutral Agree Strongly agree

35. Everyone's ideas are needed if we are going to be successful.

Strongly disagree Disagree Neutral Agree Strongly agree

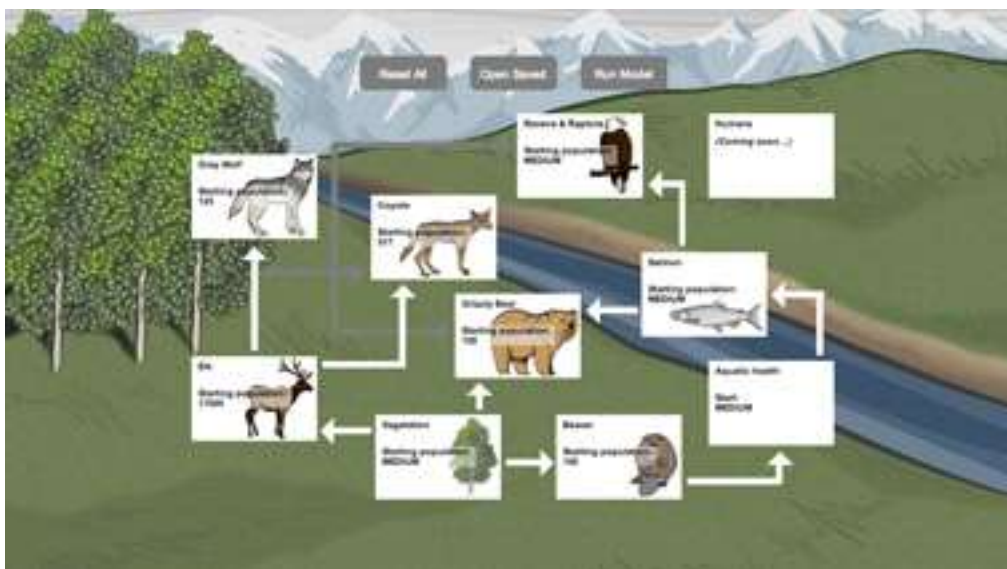
36. When I work with other students we spend too much time talking about other things.

Strongly disagree Disagree Neutral Agree Strongly agree

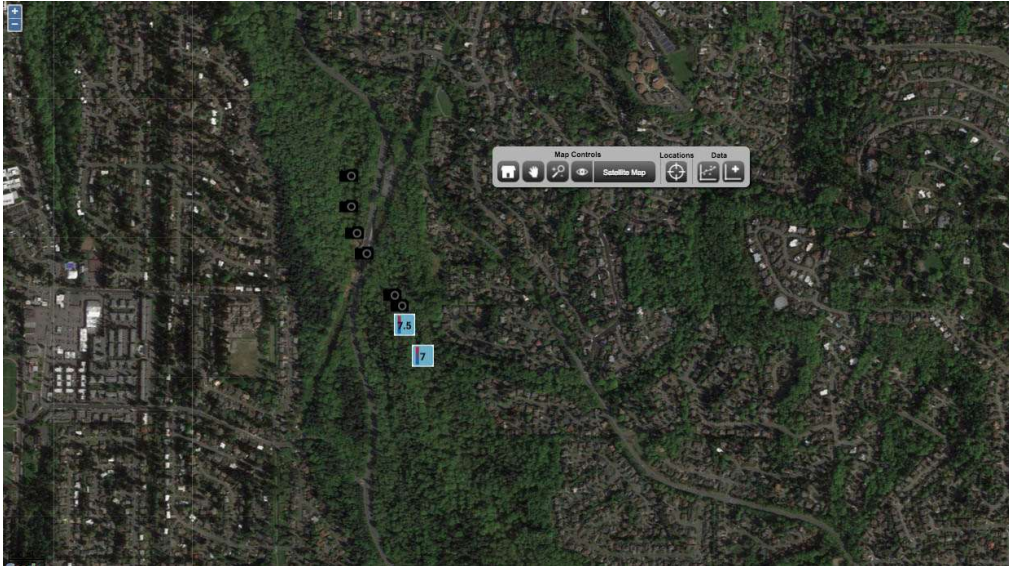
APPENDIX C – SCREENSHOTS OF LEARNING APPLICATIONS



The poetry analysis application used in the lab study described in Section 4.1. This application was built using the Microsoft PixelSense SDK.



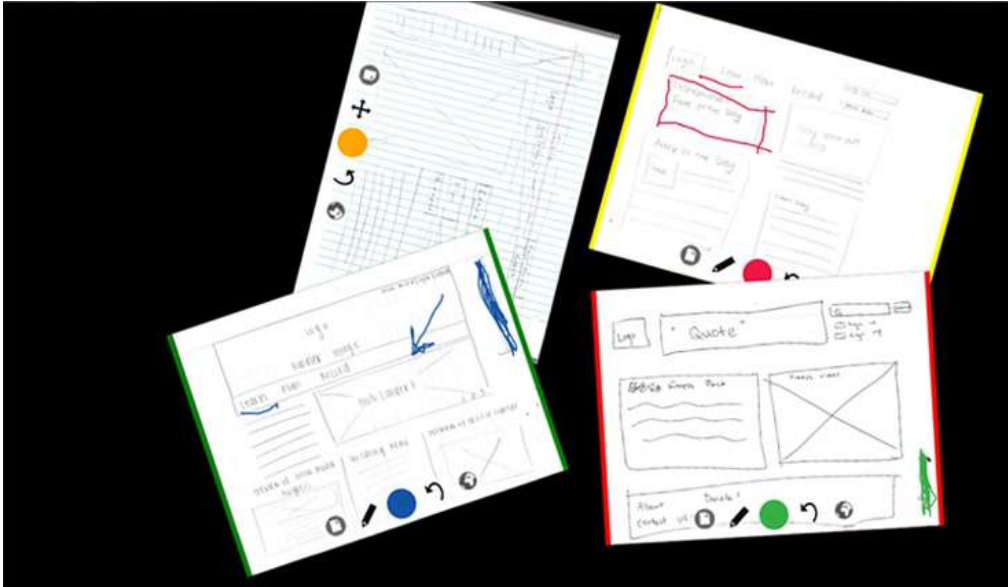
An ecosystem modeling application created for an after-school science program for middle school students but not deployed. This application was created with HTML5.



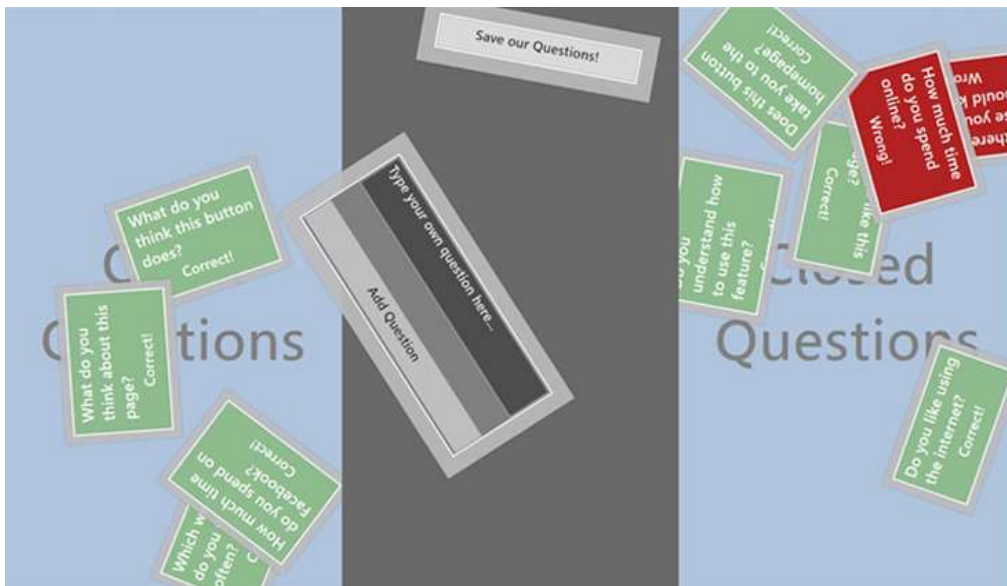
The mapping application used to collect touch data as described in Section 5.1.1. This application was built in HTML5.



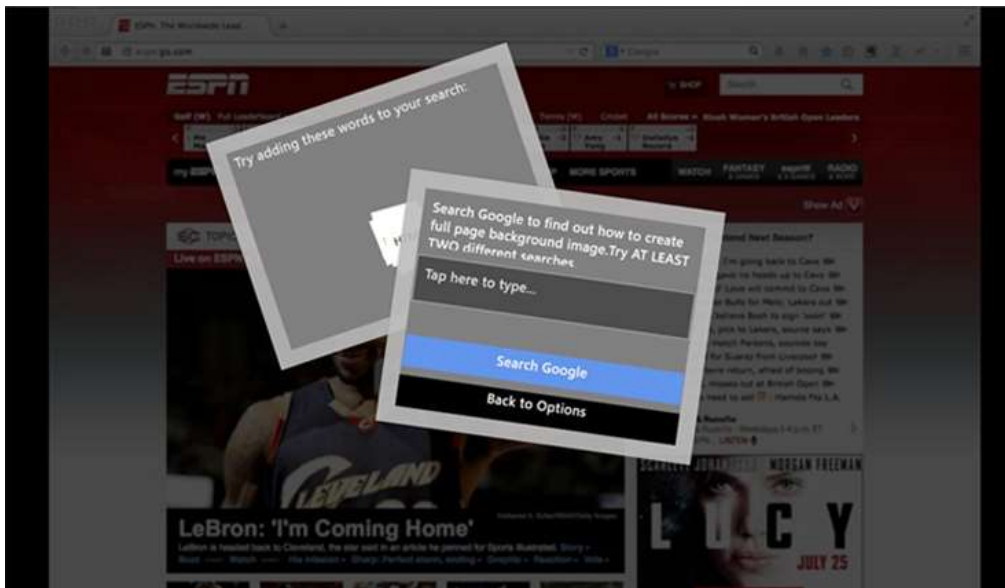
A climate change modeling application created for an after-school science program for middle school students but not deployed. This application was created with HTML5.



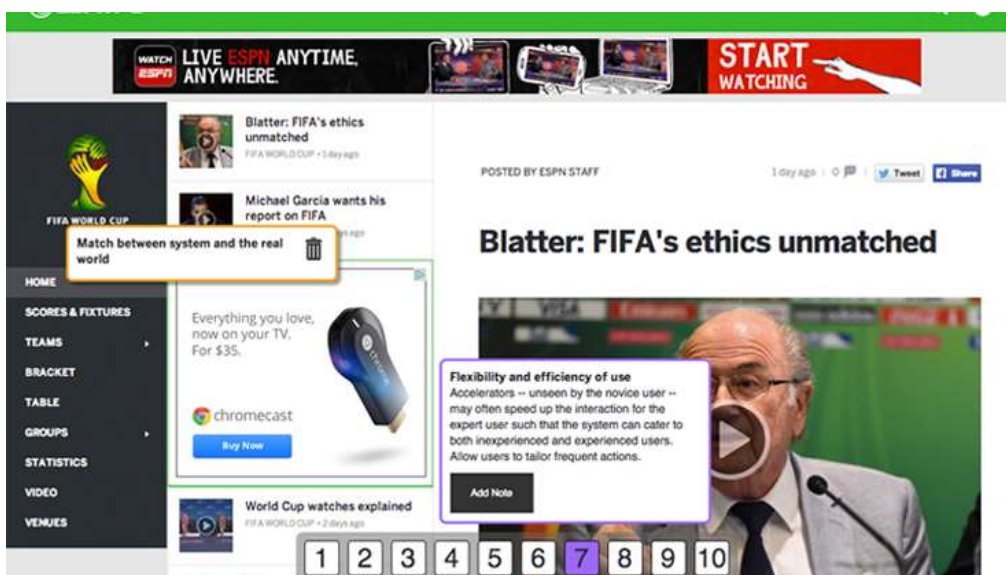
This application was built using the Microsoft PixelSense SDK and supported group design critiques in the user-centered design course described in Section 4.2.



This application was built using the Microsoft PixelSense SDK and helped students to prepare for usability tests in the user-centered design course described in Section 4.2.



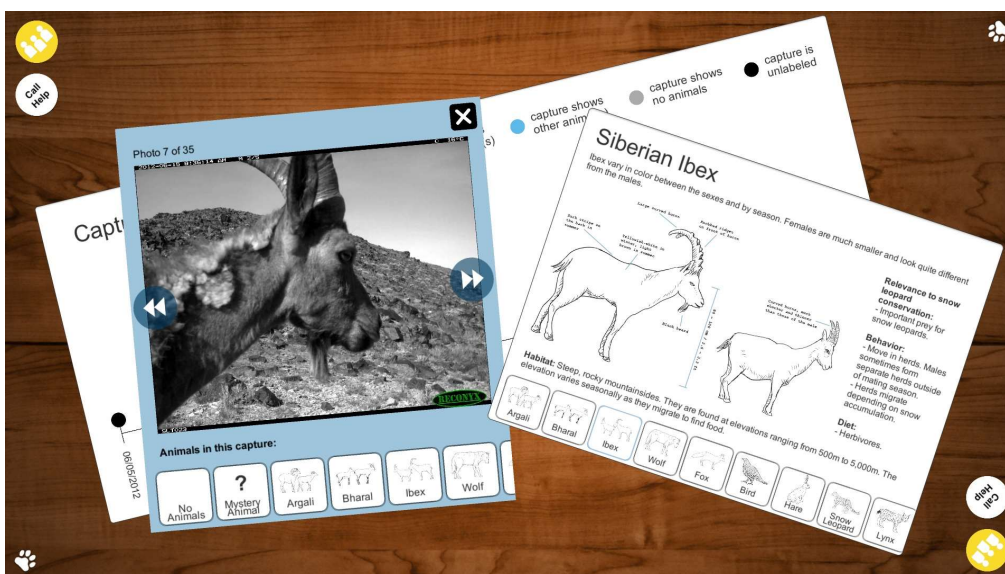
An application to help students improve their online search skills. This application was created with the Microsoft PixelSense SDK and was used in the user-centered design course described in Section 4.2.



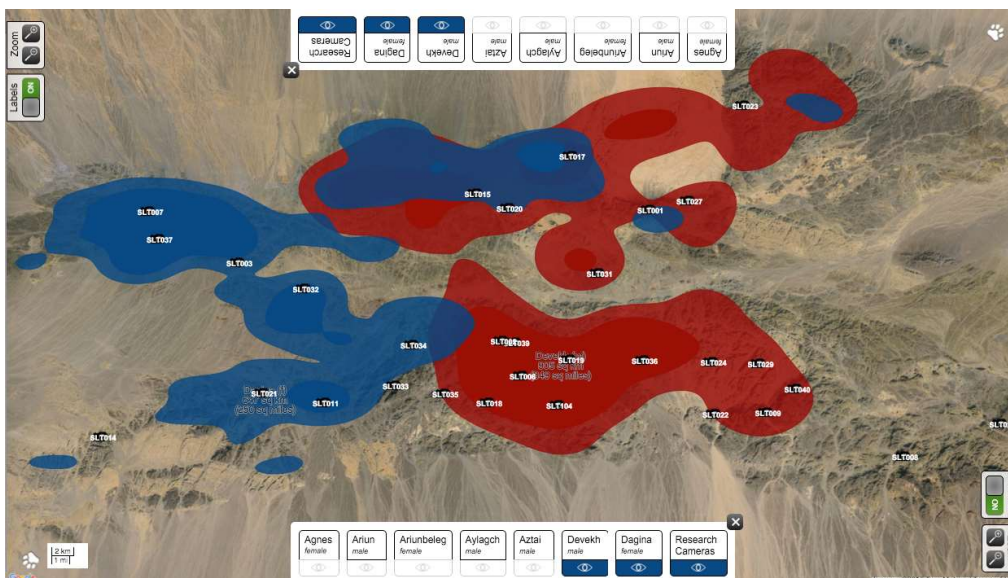
This application enabled students to annotate real-world examples of Nielsen's usability heuristics. The application was built in HTML5 and was used in the user-centered design course described in Section 4.2.



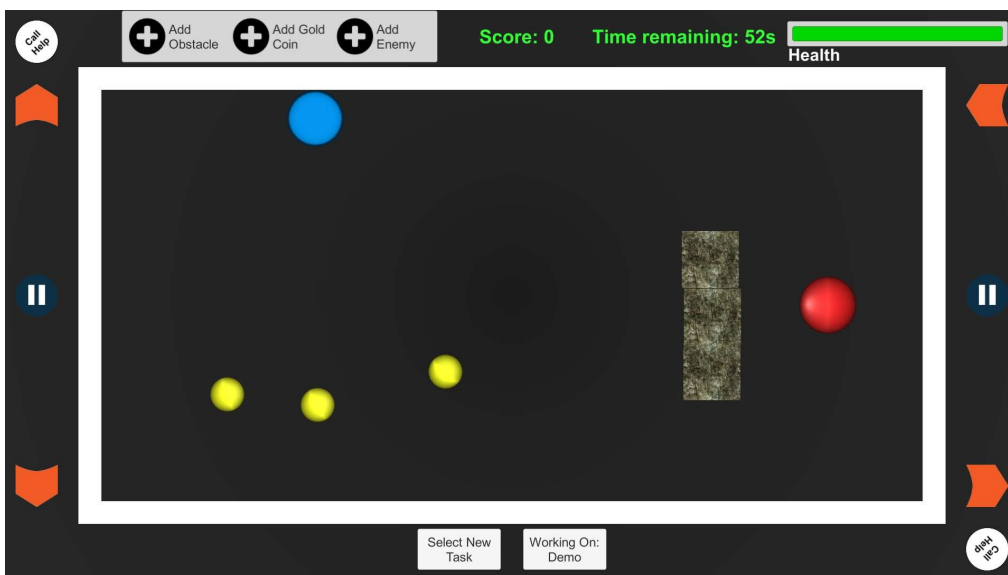
The *Snow Leopards 101* application used in the videogame development course described in Chapter 8. This application was built in Unity 3D.



The *Help a Scientist* application used in the videogame development course described in Chapter 8. This application was originally built in HTML5 then recreated in Unity 3D (pictured).



Identify Snow Leopards was built as a follow-up to *Help a Scientist*. *Identify Snow Leopards* was not used in the videogame development course because it did not align with the course learning goals. The application was originally built in HTML5 (pictured) then recreated in Unity 3D.



The **Game Challenge** application used in the videogame development course described in Chapter 8. This application was built in Unity 3D.

VITA

Abigail Evans was born in Norwich, England in 1981. She received a Bachelor of Arts in English from Bowdoin College in 2003 and a Master of Education from Harvard Graduate School of Education in 2009.