A Collaborative Environment for Engaging Novices in Scientific Inquiry

Consuelo Valdes, Michelle Ferreirae, Taili Feng, Heidi Wang, Kelsey Tempel, Sirui Liu, Orit Shaer
Wellesley College
Wellesley, Massachusetts, United States
cvaldes@wellesley.edu

ABSTRACT
We describe the design, implementation, and evaluation of GreenTouch, a collaborative environment that enables novice users to engage in authentic scientific inquiry. GreenTouch consists of a mobile user interface for capturing data in the field, a web application for data curation in the "cloud," and a tabletop interface for exploratory analysis of heterogeneous data. This paper contributes: 1) the design, implementation, and validation of a collaborative environment which allows novices to engage in scientific data capture, curation, and analysis; 2) empirical evidence for the feasibility and value of integrating interactive surfaces in college-level education based on an *in situ* study with 54 undergraduate students; and 3) insights collected through iterative design, providing concrete lessons and guidelines for designing multi-touch interfaces for collaborative inquiry of complex domains.

AUTHOR KEYWORDS
Tabletop; multi-touch interaction; collaborative learning.

ACM CLASSIFICATION KEYWORDS
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
The introduction and increasing availability of multi-touch, high-resolution displays in the form of handhelds, tabletops, and whiteboards open the opportunity to consider these technologies as a prominent alternative to current learning technologies. Several studies have examined the effects of interactive surface parameters on collaborative learning, investigating their benefits and deficits in the context of formal and informal learning; however, most of these studies have focused on children. To date, little research has been devoted to investigating the strengths and limitations of utilizing interactive surfaces in college-level learning. We are particularly interested in investigating the application of interactive surfaces for supporting collaborative inquiry of complex domains. Our main research question is: how can interactive surface technologies be used to help college students learn complex concepts through collaborative inquiry?

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ITS'12, November 11–14, 2012, Cambridge, Massachusetts, USA.
Copyright 2012 ACM 978-1-4503-1209-7/12/11...$15.00.

To address this question, we developed and evaluated GreenTouch: a pervasive environment that utilizes handheld and tabletop interfaces to mediate collaborative inquiry. GreenTouch enables undergraduate students to engage in authentic scientific inquiries in phenology – the study of how periodic plant and animal life cycle events are influenced by variations in climate.

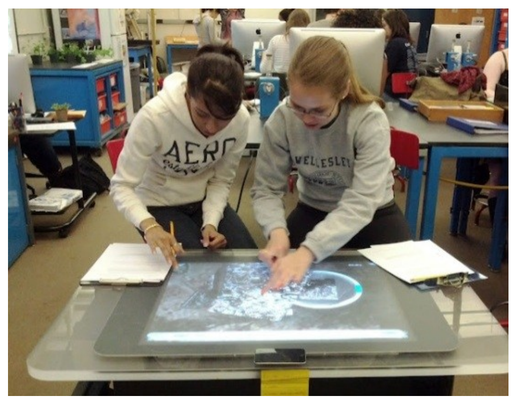

Figure 1: Collaborative exploration of heterogeneous data using GreenTouch.

Engaging undergraduate students in authentic scientific inquiries has many documented benefits including improved ability to apply scientific thinking and complex problem solving, preparedness for graduate study, and increased retention rates [12,13,28]. Furthermore, it provides students an opportunity to engage in knowledge building [25], producing ideas that are of value to the broader scientific community. However, in data-intense fields [6] such as genomics, environmental studies, and phenology, scientific investigations rely on large data sets that can be understood only through the use of sophisticated computational methods and interfaces. Such interfaces do not support important aspects of college-level science learning such as inquiry-based high-level reasoning, the development of process knowledge, and collaborative learning [15, 29, 30]. As a result, in data-intensive areas, undergraduate students rarely engage in stages of the scientific inquiry beyond data collection [23]. We propose that emerging multi-touch and tabletop interaction techniques present an opportunity for enhancing learning in fields that require understanding complex data sets, which are difficult for non-experts to comprehend.

In this paper, we reflect on the design, development, and evaluation of GreenTouch. We describe our 18-month long effort from design strategy to iterations of design, development, and evaluation. The paper presents three main contributions: 1) the design, implementation, and validation of a collaborative environment which lowers the threshold for scientific data capture, curation, and inquiry; 2) empirical evidence for the feasibility and value of integrating interactive surfaces in college-level education; and 3) insights collected through iterative design, providing concrete lessons and principles for designing multi-touch interfaces for inquiry of complex domains.

This paper is organized as follows: we first discuss related work, followed by a description of the GreenTouch system. We then outline our multi-tiered *in situ* evaluation with 132 students and present results from a recent study with 54 students, focusing on users' engagement with data and learning through collaborative inquiry. We conclude with design implications for utilizing interactive surface technologies for collaborative inquiry of complex domains.

RELATED WORK

Ubiquitous and Mobile Systems for Field Study

A number of mobile applications have been developed to enhance field study [1, 4, 22, 23, 32, 34]. For example, LilyPad [23] is a mobile application for field study that promotes learning through integrating task-based and inquiry processes *in situ*. However, we found that extreme field conditions, as well as the complexity of the data set, often drive the separation of data collection in the field from inquiry processes in the lab. Out There In Here (OTIH) [1] is a distributed information resource system for learning geology in higher education settings. Similarly to GreenTouch, it utilizes a variety of devices to provide situated access to information. However, it was designed to support remote collaboration, while our focus is on design for co-located collaboration. ButterflyNet [38] is a mobile capture and access system for field biologists that integrates paper notes with digital photographs. It shares our challenge of collecting, transforming, and organizing heterogeneous field data. However, ButterflyNet was created to enhance the work of *expert* users (e.g. field biologists) and facilitate the collection of rich and expressive field data, while our goal is to enable *novice* users to collect concise, accurate, and consistent data. Thus, while ButterflyNet utilizes paper-based input to allow users to generate free-form data such as descriptive text and sketches during an extended stay in the field, we seek to guide and constrain user input to a well-defined data collection protocol determined by the primary investigator (PI). Our goal was to minimize the time required to collect field data, while accommodating extreme weather conditions. Thus, we chose to trade the rich affordance of the paper notebook in favor of a small device that can be easily protected from water and allows for quick touch input for selecting options in a strict data collection protocol, in addition to note, video, and photo capture. ButterflyNet also provides a desktop browser for accessing field data, allowing users to collaborate *asynchronously* through shared notebooks. Since our focus is on promoting the participation of novice users in scientific inquiry, we consider support for *co-located* collaboration essential.

Figure 2: Capturing data in the field using GreenTouch.

Tabletop Interfaces for Scientists and Learning

Several systems illustrate the potential of supporting science education through multi-touch and tabletop interaction: Augmented Chemistry [2] and CheMo [35] are tangible user interfaces for chemistry education. Schkolne et al. [26] developed a tangible immersive interface for the design of DNA molecules. However, these systems focus on the representation and manipulation of objects that have an inherent physical structure. Involv [7] is a multi-touch tabletop interface for exploring the Encyclopedia of Life that shares our challenge of creating effective interaction techniques for large data spaces. However, it targets museum and informal learning settings, which are different from formal learning environments. Piper and Hollan [19] conducted a study with undergraduate students, comparing the affordance of tabletop displays and paper handouts for studying neuroanatomy. However, their study utilized a tabletop prototype with minimal functionality. We present the design and evaluation of a feature-rich interface that supports complex analytical tasks and utilizes various devices. Schneider et al. [27] developed Phylo-Genie, a tabletop user interface for collaborative learning of college-level phylogeny. However, to date, this system has not been evaluated in authentic classroom settings. G-nome Surfer [29, 30, 31] is a tabletop interface that supports collaborative exploration of college-level genomics. While G-nome Surfer focuses on the hypothesis forming stage, we present the design and implementation of a pervasive system that supports data collection, curation, and exploratory analysis. Finally, to date, a few systems have been developed to facilitate collaboration among scientists across large displays and tabletops (e.g. [24, 37]). However, these systems target experts rather than novices.

Applications for Citizen Scientists

Numerous mobile and web applications have been created to engage non-scientists in scientific data collection. Several of these programs have demonstrated the value of mobile devices for data collection (e.g. OilReporter[1], and iNaturalist[2]). Here, we only discuss those most relevant to our work. Project BudBurst[3] engages citizen scientists in phenology data collection. It provides online reference, allowing users to collect data in different locations using standard paper-based journals or mobile phones. Creek Watch [11] is a mobile application for data capture that was designed to ensure the usefulness of the data collected. While empowering citizen scientists to collect data, both BudBurst and Creek Watch maintain a traditional model where citizen scientists collect data for scientists to study. Pathfinder [14] is an online collaboration environment for citizen scientists which challenges this traditional model. While Pathfinder shares our goal of empowering non-scientists to contribute to authentic research, it targets a distributed user population studying transportation patterns.

GREENTOUCH: DESIGN AND IMPLEMENTATION

Design Strategy

Our goal was to develop an interactive system for engaging novice users in both data collection and exploratory analysis of phenology research. Due to the complexity of this domain, our design strategy combined user-centered and participatory design methods. We established design partnerships with a botany research group led by the director of botanical gardens in our institution. The group consisted of four faculty members and six undergraduate student researchers. To formulate requirements and design goals, we collected data through a series of recurring individual hour-long meetings with members of this group. A review of related literature and teaching materials provided us with additional insight. Our design partners were an integral part of our team. They participated in brainstorming sessions and provided concrete feedback on a series of prototypes. This participatory and iterative process was instrumental for increasing the fidelity and ecological validity of our system. Based on the data we collected, we identified a set of requirements for supporting novice users in scientific inquiry of phenology:

R1: Simplifying and accelerating the capture of heterogeneous, consistent, and accurate data

Phenology researchers collect data about various biological events in specific locations repetitively over a long period of time. Researchers capture heterogeneous data that include GPS data, quantitative data (e.g. height), weather conditions, digital photos, and notes. We identified three core challenges for engaging students in data collection. 1) *Quality of data*: capturing scientific data requires observation, technical skills, and judgment that are often developed through experience and guidance. Thus, data collected by students are often less accurate and prone to errors. The quality of the data is also affected by varying levels of commitment and effort. 2) *Time*: repetitive data collection in the field is time consuming and often takes portions of class time that could have been dedicated to analysis activities. 3) *Field conditions*: data collection is often conducted under harsh conditions such as heavy rain or the presence of insects. Often, students are required to carry materials and instruments to access a remote site.

Considering these challenges, we defined R1: simplifying and accelerating the capture of heterogeneous data while increasing accuracy and consistency. A technological solution must also be flexible enough to accommodate various data collection protocols and instruments.

R2: Transforming disparate data into a cohesive collection

The current practice of data collection includes the use of paper journals along various instruments. Thus, a core challenge for data curation is transforming data into an analyzable form. Most often, this means manually typing or copying information into a spreadsheet, a process that is time consuming and error-prone. We found that data collected by students is often not used, since it is not transformed into digital form. Furthermore, there is only limited computational support for coordinating, associating, and organizing the heterogeneous data collected in the field [7, 40]. Thus, we identify R2: Providing a solution for transforming disparate data into a *cohesive* collection of analyzable and related data sets rapidly and reliably.

R3: Facilitating exploratory analysis of heterogeneous data by novices, and R4: Fostering collaborative learning, reflection and discussion.

To understand complex systems, biologists utilize various generic tools, including statistical and visualization software, image processing packages, and geographic information systems (GIS). While providing rich functionality, these tools pose high threshold and are often not suitable for novices. One of the instructors stated: “Our GIS system has the capability of multimedia, but it has a steep learning curve. It will take hours to get that far.” In addition, generic tools do not support important aspects of college-level science learning such as high-level reasoning and collaborative learning. Another instructor describes: “For teaching research, having data analysis be more participatory is really important. It’s something that is kind of a black box now for students.” We thereby define R3: Facilitate exploratory analysis of heterogeneous data by novice users, and R4: Foster collaborative learning, reflection, and discussion.

Design Goals

In addition to the requirements above, we defined the following learning goals for GreenTouch:

[1] http://oilreporter.org
[2] http://inaturalist.org
[3] http://neoninc.org/budburst

L1 Developing observation and data collection skills
L2 Interpreting field data through critical examination
L3 Forming hypotheses based on exploratory analysis

We then translated our requirements and learning goals into specific goals aimed at informing our design:

G1 Allowing novices to collect, curate, and explore scientific phenology data.
G2 Reducing the mental workload associated with accessing and manipulating large amounts of heterogeneous data.
G3 Providing fluid transition between data collection, data exploration, and hypothesis forming.
G4 Fostering learning through discussion and reflection.

Following, we describe the development of GreenTouch; focusing on the rationale we followed to meet these goals.

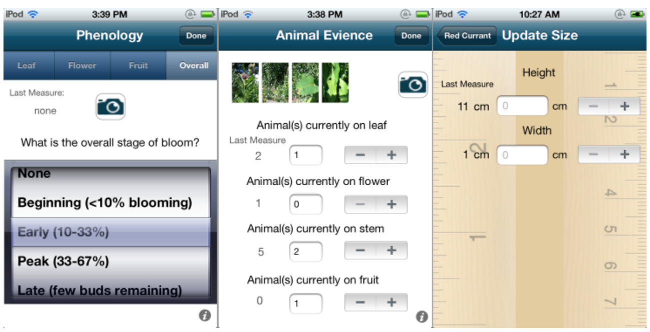

Figure 3: (a) A prompt to guide through categorical data collection. (b) A prompt for numerical data collection. (c) A ruler facilitates accurate measurements.

Design principles

In addition to our requirements and design goals, we propose the following generalizable design principles to foster collaborative inquiry of complex domains:

1) Reducing complexity - to lessen the mental workload associated with manipulating large amounts of information, we aimed to reduce the perceptive complexity of the interface while supporting rich inquiry cycles.

2) Supporting reflection - research indicates that reflection is essential for learning in the sciences [33]. Thus, we sought to support users in stepping back from “action mode” to achieve perspective, evaluate their experiences, and formally articulate insights.

3) Designing for large amounts of data - we aimed to create an environment that enables novice users to explore and make sense of large amounts of data. We applied various strategies including reducing clutter, highlighting connections, and supporting spatial manipulation.

4) Utilizing ecology of devices - to support fluid interaction across an inquiry cycle, we utilize ecology of devices, where each device used for a particular activity best supported by its affordance. For example, mobile devices facilitate portability and suit data capture in the field, while horizontal tabletops mediate collaborative discussion [4,6,10,24,35]. It is also important to consider metaphors for communicating relationships between different devices.

System Design

While there may be multiple ways to address each of these goals, our challenge was to find a way to address them all in a coherent design. Early in the design process, we realized that no single device is capable of meeting our design goals, so our system consists of three interconnected components: a mobile user interface for field data capture, a web application for automatic data curation, and a tabletop user interface for exploratory data analysis in the lab. Together, these components constitute a pervasive environment that enables novice science students to engage in scientific inquiry.

Mobile Interface for Heterogeneous Data Collection

The GreenTouch mobile interface augments traditional field data collection in several ways: 1) automatically tagging each entry with time, location, user names, and weather conditions (weather condition data is collected through Wi-Fi communication with a meteorological station on site); 2) presenting users with prompts that follow a particular field protocol while supporting the collection of heterogeneous data (e.g. measurements, photos, and notes); 3) applying mechanisms to ensure data quality.

To initiate interaction, users scan a QR code of a specific site or select a location from a list. Then, users are presented with a status update followed by a series of prompts that guide them through data collection. Figures 1 and 2 show data collection using GreenTouch. Upon completion, users submit their data via Wi-Fi.

Our design aims to reduce the threshold and provide guidance for capturing quality data (G1). To accelerate data collection under harsh weather conditions and to facilitate the collection of standardized data, we gave up the rich affordance of the traditional paper notebook in favor of a small computational device (i.e., iPod). Such devices can be easily protected from water and are capable of recording voice notes, capturing video, taking photos, and communicating through Wi-Fi. Through well-structured and standardized multiple-choice prompts, we reduced the need for text entry. However, users could choose to enter text in dedicated notes fields. We applied a multifaceted approach to ensure data quality, which considered completeness, consistency, and accuracy. To increase consistency and accuracy, we provided access to situated reference information, sample data, and an accurate ruler. We also checked for data completeness. To foster a sense of ownership, we associated a user name with each data entry (G4). Figure 3 shows screenshots of the mobile interface.

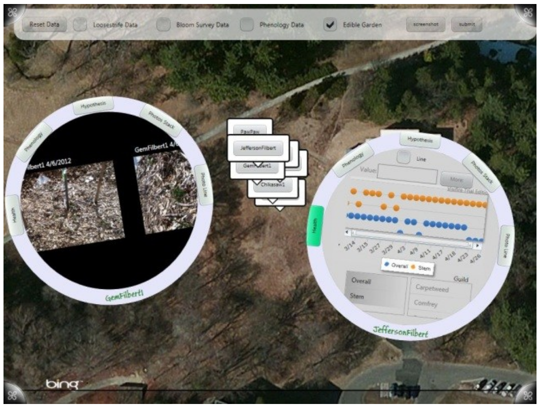

Figure 4: Map-based visualization with lenses for exploring heterogeneous data.

Web application for data curation

The web application makes data available in the "cloud" for both humans and machines. Data passes between mobile devices and the web application using a Wi-Fi network installed on campus (G3). The tabletop interface draws data from the web application and makes it available for users to manipulate. It then passes new information (e.g. users' annotations) back to the web application. Users can also access the data directly by downloading a spreadsheet. The application organizes the spreadsheet by data site. For each site, it curates quantitative and categorical data as well as links to relevant digital photos. The photos are stored on Flickr and can be accessed directly by users. Curating the data in the "cloud" lowers the threshold for maintaining, sharing, and accessing large amounts of data (G1).

Tabletop Interface for Collaborative Exploration

Based on existing research indicating that tabletop interfaces support collaboration through visibility of actions, egalitarian input and physical participation [5, 8, 20, 21, 30], facilitate active reading [16], and afford distributed cognition [18, 30], we decided to utilize tabletop interaction to facilitate collaborative data exploration (G4).

The GreenTouch tabletop interface supports the integration of heterogeneous data. To initiate interaction, users select a data set. The interface then displays a map-based visualization with highlighted sites. When a user touches a site, a semi-transparent circular "lens" appears around the site (see Figure 4). Each lens has several filters that can be applied to explore different facets of information. For example, Figure 4 shows two lenses: the first (right) displays an interactive time-series visualization, while the second lens (left) shows an interactive timeline of photos. Filters can be easily added or removed. The interactive time-series visualization in Figure 4 shows the overlay of two phenology data sets. The visualization also presents descriptive statistics for each time series that include mean, standard deviation, number of entries, and the minimum and maximum values. Users can pan left and right to explore large data sets and touch individual data points to display additional information (e.g. user name or weather).

The application provides the following filters: time series, interactive timeline, photo library, and hypotheses. Lenses can be moved and rotated upon the surface, as well as, easily removed. Users can annotate individual data points, particular time series, and different photos (see Figure 5). Users can also save photos or entire lenses for later reference by dragging the items into a "drawer" at the bottom of the surface. Users can "open the drawer" to revisit saved items at any point. Figure 5 shows a semi-open drawer. To combine data sets, users simply overlay two lenses, which are merged into a single lens.

Figure 5: A merged dataset and semi-open drawer with 3 annotated photos.

While the GreenTouch tabletop application eliminates some of the advanced functionality offered by existing statistics packages and GIS tools, the focus of our design was not on quantitative analysis but rather on enabling flexible but rich comparisons. Thus, we implemented: 1) seamless integration of quantitative and qualitative data, 2) fluid transition between data sets including merging data sets and side by side comparison, and 3) built-in visualizations that help users quickly see patterns.

Implementation

The GreenTouch mobile application is implemented using iOS and runs on iPod Touch devices. The QR code reading is implemented using the ZXing library. The application saves data locally until a Wi-Fi connection is available and then passes it to the web application. Photos are submitted to Flickr using the ObjectiveFlickr API. Photos are tagged by site, plant, and description. This allows users to view and share photos using any desktop computer. The tabletop interface is implemented using Microsoft Surface SDK 1.0 and updated to Surface 2.0 SDK for the SUR40. The web application is implemented in Python using Google AppEngine. The application receives data in the JSON format and upon request prints out a comma-delimited document (.csv) that can be opened by various tools. The application can also print to a webpage as well as use JSON format to send data to the mobile and tabletop interfaces.

EVALUATING GREENTOUCH

Our goal was to gain an understanding of the strengths and limitations of GreenTouch in supporting college-level learning through collaborative inquiry. We thereby utilized a variety of quantitative measures and qualitative indicators, that combined together, highlight multiple facets of the collaborative learning process.

Throughout the development process of GreenTouch, we applied a multi-tiered evaluation framework [31], which consisted of three *in situ* studies with a total of 132 participants. All studies were conducted by deploying the interface in *intro to botany* courses in our institution; participants were undergraduate students enrolled in the course (all female, age range 18-22). GreenTouch was used to replace a paper-based *data collection* activity included in the syllabus. Due to the challenges discussed in the requirement analysis (R2, R3) the original syllabus did not include an *exploratory analysis* activity. However, during the three semesters in which GreenTouch was studied, the instructors added an exploratory analysis activity facilitated by GreenTouch. All studies took place in our institution's botanic gardens. Partial Wi-Fi coverage was available.

Study	Metrics & indicators	Methods	Iteration
Spring '11 (n=43) Usability	• Accuracy • Completion rates • Errors • Satisfaction • Time on task	• Logging • Observations • Questionnaires • Debrief	Mobile: • Navigation • Situated reference • Comments • Summary Surface: • Cloud metaphor • Lenses
Fall '11 (n=35) Usefulness	• Accuracy • Number and quality of hypotheses • Participation • Satisfaction • Time on task	(all of the above) • Video analysis • Interviews	Mobile: • Ruler • Sample data • Data checks Surface: • Drawer • Visualization
Spring '12 (n=54) Impact	(all of the above) • Talk categories • User actions • Information artifacts	(all of the above) • Discourse analysis	Surface: • Advanced statistics tools

Table 1: Metrics, methods, and design iterations.

The first study applied a micro-perspective - focusing on the *usability* and functionality of concrete interaction techniques; the second study applied a macro-perspective - studying the *usefulness* of the system in the context of a full-scale task; the third study applied holistic perspective – examining the *impact* of the system on the collaborative-inquiry process. Table 1 describes the metrics, collection and analysis methods used in each study, and the major features addressed in following design iterations. In this paper we report findings from the third, most recent, study.

User study: Spring '12

We deployed GreenTouch in an intro to botany course for a period of 4 weeks in which students collected spring phenology data. In each of 4 weekly lab sections (held on different days and lasting 3 hours each), students collected data in the field. Overall, 54 students (25 pairs and 4 singletons) participated in the data collection phase, collecting data on 5 species. In week 4, following data collection, each lab section was divided into smaller groups which used the tabletop interface to explore the entire data set. Participants were asked to form two hypotheses for future research based on exploratory data analysis. Overall 49 students, 24 groups participated in the data analysis activity (21 pairs, 2 triples, 1 single). Two Microsoft Surface 1.0 and one SUR40 devices were set up for this activity in the botany lab. While three groups used the tabletop interface, others completed an unrelated activity. Upon task completion, the groups switched activities.

Mobile

Satisfaction

We used post-task questionnaires and interviews to evaluate usefulness and satisfaction. Table 2 shows the questionnaire results. Overall, users found the interface useful and enjoyed using it. The results also show moderate confidence levels and an indication that the application helped users to reflect on the purpose of the study. Through post-task interviews and an investigation of the data collected we found evidence that users improved their observation and data collection skills (L1). For example, one user described: "The app made my observations more systematic and organized." Users also commented on the utility of the mobile device: "Much easier to collect data and it saved paper!" In face-to-face interviews, conducted in weeks 2 and 3, the instructors confirmed that the data collected is useful and that the curation method is effective for both teaching and research.

Question	Mean (SD)
Using the app made my observations better.	4.19 (0.96)
Using the app made me consider the relationship between climate and phenology.	3.66 (0.82)
Using the app let me collect data faster than without the app.	4.52 (0.95)
Using the app made it easier to collect heterogeneous data.	4.37 (1.06)
I enjoyed using this app to collect data.	3.49 (1.05)
I was confident in the accuracy of the data I gathered.	3.39 (1.03)

Table 2: Results of post-task questionnaire for mobile interface, scale of 1 (strongly disagree) to 5 (strongly agree).

Collaboration

To get insight into the collaborative process during data collection, we analyzed observations collected by investigators positioned in each data site and transcripts of user dialogs recorded for each pair in each session (total 79 legible recordings). We coded for verbal participation, data entry events, and physical interaction with plants of study.

We found that in the first session most pairs collaborated through *turn taking,* sharing control of the device while demonstrating similar levels of involvement. Decisions were made through critical discussion. In subsequent sessions however, in about *half* of the pairs the partnership became less equitable. One participant would take control of the device early on and "drive" for the session. The driver would think aloud while the navigator was participating verbally, interrupting the driver to express different ideas. In most *driver-navigator* pairs, the same participant assumed the role of the driver in all sessions. Overall, we found that most pairs (21/25) exhibited effective collaboration (turn-takers 13/25 or driver-navigator 8/25) that was motivated by the *opportunities for discussion* provided by the mobile interface, while in a small number of teams (4/25 driver-passenger) only the driver was involved in the task. The data collection prompts and the situated reference data encouraged students to reach decisions through discussion, which, in turn increased their confidence in the data collected. As one student described: "Today, my partner was absent. I find it easier to work with someone and get a second opinion. It would have made me more confident about the results."

Tabletop

Performance

The 24 groups spent on average 27:37 minutes using the system (SD 7:40 minutes). The relatively large standard deviation can be attributed to natural variation between participants and groups. At the end of each session, each group was required to formulate two hypotheses about the data for future research, while providing supportive evidence. The instructors rated 22 of the hypotheses on a scale of 1 to 5, where one was assigned to a shallow and simple observation, and five represented an exceptionally insightful hypothesis. The instructor ratings were distributed as follows: one (0%), two (27%), three (27%), four (36%), and five (5%). Overall 68% of the hypotheses formed by students were valuable and based on evidence. This indicates that users were able to utilize the interface to make sense of the heterogeneous data sets and derive useful insights (L3) within a short time frame.

Satisfaction

To gage user satisfaction, we utilized a post-task questionnaire, which asked users to rate enjoyment, confidence in their findings, difficulty of the task, and frustration using the interface. We used a Likert scale of 1 (low) to 7 (high). Users were also asked to comment on their experience. We found that users enjoyed (4.38 mean, 1.85 std) using the interface, had a moderately high confidence in their hypotheses (4.28 mean, 1.35 std) and low frustration (2.93 mean, 2.06 std). One participant shared: "I liked manipulating the data and exploring the ways in which it was possible to visually compare it. It made the analysis much less time consuming and probably more accurate." Users rated the difficulty of the task as moderate (3.43 mean, 1.56 std). Among SUR40 users, we found lower enjoyment (3.58, 2.27 std) and increased frustration (3.17, 2.44 std) caused by over sensitivity, which registered hovers as accidental touches.

Collaboration

To gain insight into *how* users collaborate and solve problems using GreenTouch, we conducted a manual video analysis of 24 recordings. Our coding scheme included verbal and physical participation (e.g. touch events, and "offline" gestures such as pointing, reaching, blocking, note taking). We calculated levels of physical participation per participant by summing the number of touch events and offline gestures. We calculated verbal participation levels per participant based on session transcripts. We also coded the interaction with information artifacts, recording their source, type, manipulation and configuration.

Based on this analysis, we found that during the session, each of the groups applied one of two distinct *problem solving strategies*: 1) Vertical (9/24) – where investigation focused on one heterogeneous data set (i.e., lens). The data was carefully examined using the various filters to explore multiple data facets of an individual site. Once several observations were made, another dataset was introduced and merged into a new data set. Conclusions were drawn through side-by-side comparison of the original data sets and careful analysis of the merged data set. The original data sets were stored on the surface or in the drawer. If needed other datasets were introduced one at a time for detailed examination. Data sets that were determined not interesting were discarded from the table. 2) Horizontal (15/24) - users brought up as many as four data sets at a time, and compared them to one another by aligning them next to each other. Users explored multiple facets of the data with the goal of identifying general trends from as much data as possible. Inter-coder reliability based on a random sample of 20% of the data was measured using Cohen's Kappa as 1.0 (95% CI: [0.12, 1.8]), indicating strong agreement. In both strategies we observed that participants moved information artifacts on the tabletop frequently and used the horizontal surface for spatially arranging the lens in meaningful ways, for example utilizing proximity to reflect potential connections. Also, side-by-side comparison was a key feature in both strategies. Some groups used the "drawer" as their workspace to arrange and compare lenses, while others used it for saving and revisiting information artifacts. Most teams didn't use the hypotheses filter for recording their observations through note taking until the end of the session when they were asked to record their hypotheses.

Studying physical and verbal participation data, we found that in the beginning of the session, often one user was unsure about trying out the new technology and thus waited until the partner took physical control of the system. When the timid user became comfortable with the technology, or was pushed into action due to their partner's inability to complete a task, they began interacting with the system physically. This would often result in turn-taking collaboration for the rest of the session. In other cases, users switched roles between driver (who participates both physically and verbally) and navigator (who guides the driver through verbal cues and "offline" gestures). We identified that in general, pair collaboration patterns could be described using 3 profiles (we excluded the singleton user): turn-takers (9 of 23), driver-navigator (10 of 23), and driver-passenger (2 of 23). The 2 groups of triples could be described as driver-navigator-passenger, and turn-takers and passenger. Inter-coder reliability based on a random sample of 20% of the data was measured using Cohen's Kappa as 0.62 (95% CI: [-0.19, 1.42]), indicating substantial agreement. These profiles draw on those presented by Shaer et al. [31] but were adjusted to describe interaction in triples. Interestingly, while most pairs collaborated effectively, in the triples one user was less engaged (i.e. passenger). This may result from our design, which uses a single orientation and thus affords side-by-side interaction leading to the exclusion of a third user.

To better understand the *nature of the dialogue* around the tabletop, we conducted a discourse analysis using transcripts of 23 groups (we excluded the singleton user). We iteratively developed a coding scheme that classifies talk into seven general categories. Table 3 defines each talk category and provides an example from the transcript for each. It also shows the proportional distribution of talk categories. Inter-coder reliability based on a random sample of 20% of the data was good with 81% agreement. These results suggest that users were mostly engaged in effective discussion, which focused on the task, as evident by the proportion of talk dedicated to problem solving and reflection (together 48%) and the low percentage of disengagement utterances (3%). We also found within the problem solving and reflection utterances evidence that users interpreted the data through critical examination (L2) (see examples in Table 3). Our discourse analysis further indicates that users established task division through continuous discussion as evident by the proportion of coordination talk (25%) and found the interface easy to use as evident by the low proportion of syntax talk (8%).

DISCUSSION

We applied holistic approach in examining the impact of GreenTouch on the collaborative-inquiry process, focusing on performance, satisfaction, and collaboration.

Our findings provide *empirical evidence for the feasibility and value* of utilizing ecology of devices for helping college students learn *complex* concepts through collaborative inquiry. In particular, we showed that GreenTouch: 1) allows novices to collect consistent and useful heterogeneous data that is curated automatically (G1); 2) enables novice students to participate in exploratory analysis that produces valuable insights through effective collaboration (G2, G4); and 3) facilitates nearly seamless transitions between stages across the workflow (pending connectivity) with no manual intervention (G3). While these results are specific, taken together they make up a compelling case study of using interactive surfaces for collaborative learning of complex concepts.

Category	Definition	Examples from transcripts
Problem Solving 43%	Planning, relevant questions or pointing out facts.	1: It didn't change degree. 3: Oh, okay. And then if we look at this, then we have quite a fluctuation, but both going towards a more positive trend upwards, so maybe over a longer period of time we can see an average upward
Coordination 25%	Turn taking, navigating, verbal shadowing, or task division.	2: Let's just focus on health because that is like growing 1: Wait, we should choose another plant. 2: OK. Let's compare it with 1: Which one?
Reflection 5%	Referring to the purpose of the exercise, the data collection, analysis, or hypothesis forming stages.	1: Okay, so it looks like someone didn't know what they were doing when they were doing the data. It's quite possible. Everybody was doing it for the first time and it's hard to know whether it is a leaf bud, a flower bud…
Peer Teaching 1%	One participant teaches the other how to analyze or interpret data.	1: Yeah, but she<the instructor> was saying that you already know the actual x-values…so you don't look at it as a line but rather as individual points so that gives you…there… <how to read a graph>
Syntax 8%	Referring to how to use certain features.	1: Oh. Wait. How'd you say you could combine them?
Brief Response 15%	Short responses to suggestions	1: Yeah 2: Yeah 3: mumbles
Disengagement 3%	Non-task related	2: And our report is due the 23rd. 3: No it said the 30th

Table 3: Dialogue coding scheme.

We also highlight *how users collaborate and solve problems* using handheld and tabletop interfaces in the context of college level inquiry-based learning. We showed that *mobile interfaces* could motivate effective co-located collaboration by providing *opportunities for discussion.* We also showed that mobile interfaces can guide users in complex inquiry through *situated reference*. We further demonstrated that the large size and shared workspace of a tabletop display provide benefits for collaborative problem solving. For example, side-by-side comparison and spatial interaction played an essential role in the problem-solving process mediated by GreenTouch: users accumulated, arranged, and moved information artifacts on the tabletop frequently. Also, we found that most pairs collaborated effectively through turn taking and role switching (driver-navigator). Such collaboration styles were mediated by the large horizontal multi-touch display and were established through continuous coordination (see Table 3). To date, few projects have explored the use of interactive surfaces with college-level learners (rather than children); thus, these insights shed light on how *older* users collaborate and solve problems using interactive surfaces. These findings are important for driving the discussion on educational benefits of interactive surfaces, for college-level learning, forward.

Finally, our study has several limitations that point towards future work: we neither provided participants with the opportunity to engage in iterative analysis nor measured learning outcomes in a way that enables to assess individual learning and participants' ability to apply their learning.

Designing for collaborative inquiry of complex concepts

Four generalizable guidelines shaped our design:

First, *reducing complexity*: we applied design strategies for eliminating complexity including traditional methods such as, reducing functionality and hiding complexity [10], as well as using reality-based metaphors [9] that draw upon interaction with the non-digital world, such as: lenses, filters, and drawers. Our evaluation shows that the use of lenses for displaying multi-faceted data combined with spatial interaction facilitated three aspects that were important for collaborative problem solving: incremental addition of complexity, seamless integration of quantitative and qualitative data, and fluid transition between data sets.

Second, *providing space for reflection*: to support reflection we provided physical space for the conceptual construction of ideas. On the tabletop interface, we implemented a "drawer," which allows users to deposit and revisit information. We found that while some users used the "drawer" for storing and revisiting data, others used it as their workspace, preferring to separate data retrieval from analysis. We also provided a space for formally articulating observations (i.e. hypotheses lens) through note taking. This space was not used often since users did not like typing using the multi touch keyboard. We suggest exploring alternative modalities for recording observations such as voice and pen input. On the mobile interface, we allowed users to reflect on their last data entry by presenting, upon request, a summary of the data collected. While some users used this feature to reflect on the data, most users were focused on collecting data. We plan to explore alternative designs to encourage users to step back from action mode.

Third, *designing for large amounts of data*: to mitigate the workload associated with manipulating large amounts of data, we designed for reducing clutter, highlighting the connections among information artifacts, and providing space for spatial manipulation. Our findings indicate that this approach allowed for both vertical and horizontal problem-solving strategies. We also considered the size of user interface elements, optimizing target size for touch while seeking to effectively utilize limited screen real estate. This required us to make tradeoffs such as overlaying information, and limiting the resizing of some tabletop information artifacts (e.g. lenses).

Finally, *utilizing ecology of devices*: our design facilitates nearly seamless transition across stages of the inquiry cycle by utilizing ecology of devices consisting of a mobile interface for capturing data, a web application for data curation in the "cloud," and a tabletop interface for group collaborative analysis. By making information available in the cloud, we also allow for new devices to be integrated into the GreenTouch device ecology. We experimented with various metaphors for communicating the relationship between the devices. For example, our early prototype (evaluated in the first study) was designed around the use of a "container" metaphor. When placing a mobile device on the tabletop, information appeared on the tabletop surface. However, users were confused by this metaphor and were not sure when it was safe to remove the device from the Surface. As a result, mobile devices were left on the Surface, covering important screen real estate. The current prototype uses the "site" metaphor, all data collected in a particular site is stored in the "cloud" and can be accessed by tapping the site on the map-based visualization. Users had no problems understanding the site and cloud metaphors and were readily able to find and retrieve data.

While we applied these guidelines to a particular domain they are generalizable and can be used to inform the design of other interfaces for college-level collaborative inquiry.

CONCLUSION

The increasing availability of commercial tabletop and handheld systems coupled with falling hardware prices remove major barriers for using interactive surfaces in educational settings. Within this context, this research makes three contributions that highlight possibilities for engaging novice users in activities that were previously available only for experts: 1) the design, implementation, and validation of a collaborative environment which allows novices to engage in scientific data capture, curation, and analysis; 2) empirical evidence for the feasibility and value of integrating interactive surfaces in college-level education; and 3) insights collected through iterative design, providing guidelines for designing multi-touch interfaces for collaborative inquiry in complex domains.

We intend to further evaluate GreenTouch in a longitudinal study by deploying it in a student research program.

ACKNOWLEDGMENTS

We thank our collaborators Kristina Jones, Marcy Thomas, Janet McDonough, and Alden Griffith. We also thank the Wellesley Botanic Gardens and all our study participants. This work is partially funded by HHMI through the Wellesley College Science Center.

REFERENCES

1. Adams, A., Coughlan, T., Lea, J., Rogers, Y., Davies, S., and Collins, T. (2011). Designing Interconnected Distributed Resources for Collaborative Inquiry Based Science Education. In Proc. of Digital Libraries.
2. Fjeld, M., Fredriksson, J., Ejdestig, M., Duca, F., Bötschi, K., Voegtli, B., and Juchli, P. (2007). Tangible User Interface for

Chemistry Education: Comparative Evaluation and Re-Design. In Proc. of CHI.

3. Fleck, R., Rogers, Y., Yuill, N., Marshall, P., Carr, A., Rick, J., and Bonnett, V. (2009). Actions Speak Loudly with Words: Unpacking Collaboration Around the Table. In Proc. of ITS.
4. Grant, W.C. (1993). Wireless Coyote: a Computer-Supported Field Trip. Commun. ACM 36, 5 (pp.57-59).
5. Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., and Rogers, Y. (2009). Around the Table: Are multiple-touch surfaces better than single-touch for children's collaborative interactions? In Proc. of CSCL.
6. Hey, T., Tansley, S., Tolle, K. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, Redmond.
7. Horn, M.S., Tobiasz, M., and Shen, C. (2009). Visualizing Biodiversity with Voronoi Treemaps. In Proc. Sixth International Symposium on Voronoi Diagrams in Science and Engineering.
8. Hornecker, E., Marshall, P., Dalton, N.S., and Rogers, Y. (2008). Collaboration and Interference: Awareness with Mice or Touch Input. Proc. of CSCW.
9. Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., and Zigelbaum, J. (2008). Reality-based Interaction: a Framework for Post-WIMP Interfaces. In Proc. of CHI.
10. Janlert, L.E. and Stolterman, E. (2008). Complex interaction. ACM Trans. Comput.-Hum. Interact. 17, 2.
11. Kim, S., Robson, C., Zimmerman, T., Pierce, J., and Haber, E.B. (2011). Creek Watch: Pairing usefulness and Usability for Successful Citizen Science. In Proc. of CHI.
12. Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. CBE Life Science Edu. 6, (297--306).
13. Lopatto D. et al. (2008). Undergraduate research. Genomics Education Partnership. Science. 322:684-5.
14. Luther, K., Counts, S., Stecher, K.B., Hoff, A., and Johns, P. (2009). Pathfinder: an Online Collaboration Environment for Citizen Scientists. In Pro. of CHI.
15. Mirel, B. (2009). Supporting cognition in systems biology analysis: findings on users’ processes and design implications. J Biomed Discov Collab. 27(2).
16. Morris, M.R., Brush, A.J.B., and Meyers, B. (2007). Reading revisited: evaluating the usability of digital display Surfaces for Active Reading Tasks. Proc. of IEEE Tabletop (pp.79-86).
17. Okada, T., and Simon, H.A. (1997). Collaborative discovery in a scientific domain. Cognitive Science (pp.109-146).
18. Patten, J., and Ishii, H. (2000). A comparison of spatial organization strategies in graphical and tangible user interfaces, Proc. of Designing Augmented Reality Environments.
19. Piper, A.M., and Hollan, J. (2009). Tabletop displays for small group study: affordances of paper and digital materials. In Proc. of CHI.
20. Rick, J., Rogers, Y., Haig, C., and Yuill, N. (2009). Learning by doing with shareable interfaces. Children, Youth and Environments, (pp.321–342).
21. Rick, J., Marshall, P., and Yuill, N. (2011). Beyond one-size-fits-all: How interactive tabletops support collaborative learning. In Proceedings of IDC.
22. Rogers, Y., Price, S., Randell, C., Fraser, D.S., Weal, M., and Fitzpatrick, G. (2005). Ubi-learning integrates indoor and outdoor experiences. Commun (pp.55-59).
23. Rogers, Y., Connelly, K., Hazlewood, W., and Tedesco, L. (2010). Enhancing learning: a study of how mobile devices can facilitate sensemaking. Personal Ubiquitous Computing (pp.111-124).
24. Ryall, K., Forlines, C., Shen, C., Morris, M.R., and Everitt, K. (2006). Experiences with and Observations of Direct-Touch Tabletops. In Proc. IEEE Tabletop.
25. Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences (pp. 97-118). New York: Cambridge University Press.
26. Schkolne, S., Ishii, H., and Schroder, P. (2004). Immersive design of DNA molecules with a tangible interface, Visualization, IEEE.
27. Schneider, B., Strait, M. Muller, L., Elfenbein, S., Shaer, O., Shen, C. (2012). Phylo-Genie: Engaging Students in Collaborative ‘Tree-Thinking’ through Tabletop Techniques. In Proc. of CHI.
28. Seymour, E., et al. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. Science Education, 88(4).
29. Shaer, O., Kol, G., Strait, M., Fan, C., Grevet, C., and Elfenbein, S. (2010). G-nome surfer: a tabletop interface for collaborative exploration of genomic data. In Proc. of CHI.
30. Shaer, O., Strait, M., Valdes, C., Lintz, M., Feng, T., and Wang, H. (2011). Enhancing Genomic Learning through Tabletop Interaction. In Proc. of CHI.
31. Shaer, O., Strait, M., Valdes, C., Wang, H., Feng, T., Lintz, M., Ferreirae, M., Grote, C. Tempel, K., and Liu, S. (2012) "The Design, Development, and Deployment of a Tabletop Interface for Collaborative Exploration of Genomic Data", Int J of Human-Computer Studies.
32. Sharples, M., Corlett, D., and Westmancott, O. (2002). The Design and Implementation of a Mobile Learning Resource. Personal Ubiquitous Computing (220-234).
33. Singer, S. R., Hilton M. L., and Schweingruber, H. A. (2005). America's lab report: investigations in high School science, National Research Council.
34. Soloway, E. (2001). Log on education: Handheld devices are ready-at-hand. Commun (pp.15-20).
35. Song K, Kim G, Han I, Lee J, Park J, and Ha S. (2011). CheMO: mixed object instruments and interactions for tangible chemistry experiments. In Proc. of CHI.
36. Tanner, K., Chatman, L. S., and Allen, D. (2003). Cooperative learning in the science classroom - beyond students working in groups. Cell Biol. Educ.
37. Wigdor, D., Jiang, H., Forlines, C., Borkin, M., Shen, C. (2009). The WeSpace: The Design, Development and Deployment of a Walk-Up and Share Multi-Surface Visual Collaboration System. In Proc. of CHI.
38. Yeh, R., Liao, C., Klemmer, S., Guimbretiere, F., Lee, B., Kakaradov, B., Stamberger, J. and Paepcke, A., (2006). ButterflyNet: a mobile capture and access system for field biology research. In Proc. of CHI.