

Collocated Collaboration Analytics: Principles and Dilemmas for Mining Multimodal Interaction Data

**Roberto Martinez-Maldonado^{1,2}, Judy Kay²,
Simon Buckingham Shum¹, Kalina Yacef²**

¹University of Technology Sydney, Australia

²The University of Sydney, Australia

ABSTRACT

Learning to collaborate effectively requires practice, awareness of group dynamics, reflection and, often, it benefits from coaching by an expert facilitator. However, in physical spaces it is not always easy to provide teams with evidence to support collaboration. Emerging technology provides a promising opportunity to make collocated collaboration visible by harnessing data about interactions and then mining and visualizing it. These *collocated collaboration analytics* can help researchers, designers, and users to understand the complexity of collaboration and to find ways they can support collaboration. This paper introduces and motivates a set of principles for mining collocated collaboration data, and draws attention to trade-offs that may need to be negotiated en route. We integrate Data Science principles and techniques with the advances in interactive surface devices and sensing technologies. We draw upon a seven year research program which has involved the analysis of six different group situations in collocated settings with more than 500 users and a variety of surface technologies, tasks, grouping structures and domains. The contribution of the paper includes the key insights and themes which we have identified and summarized in a set of principles and dilemmas that can inform design of future collocated collaboration analytics innovations.

CONTENTS

1. Introduction	3
2. Collocated Collaboration Analytics: Data Science meets Collocated Groupware	4
3. Related Work.....	6
3.1. Describing Collaboration Data: Visualization Approaches.....	6
3.2. Diagnosing, Predicting, and Prescribing for Collaboration Data	7
4. The Collaborative Situations: Case Studies	8
4.1. Case 1: A Multi-display Optimization Challenge	9
4.2. Case 2: A Pen-based Tabletop for 'Mystery' Solving.....	11
4.3. Case 3: A Multi-touch Tabletop for Concept Mapping.....	11
4.4. Case 4: A Multi-display, Multi-touch Classroom	12
4.5. Case 5: An Educational Design Studio.....	12
4.6. Case 6: A Health-care Simulation Classroom	13
5. Challenges with Collocated Interaction Data.....	13
6. Principles and Dilemmas.....	14
6.1. Collaborative activity	15
6.1.1. Principle 1: Understanding collaboration complexity	15
6.1.2. Principle 2: Transferability	16
6.1.3. Dilemma 1: Authenticity vs Experimentation	19
6.2. Multimodal Data.....	21
6.2.1. Dilemma 2: Multimodality and Cost	21
6.2.2. Principle 3: Temporality	23
6.2.3. Dilemma 3: Data Availability	24
6.3. Data-driven Insight	25
6.3.1. Principle 4: Post-hoc vs On-the-fly.....	25
6.3.2. Principle 5: Units of Analysis	26
6.3.3. Principle 6: Theory/hypothesis-driven Context Encoding.....	29
6.4. Social and Contextual Issues	30
6.4.1. Principle 7: Notice and Consent	30
7. CONCLUSION AND DISCUSSION.....	32
7.1. Limitations and Future Work	34

1. INTRODUCTION

The many benefits of collaboration on learning, knowledge sharing and in workplace settings, especially when it is mediated by technology, have been well documented in the Computer-Supported Collaborative Learning (CSCL)/Cooperative-Work (CSCW) literature (Gross, 2013; Schmidt & Bannon, 2013; Stahl, 2006). Collaboration can encourage people to externalize their personal perspectives in order to reach a shared understanding and generate creative solutions (Hilliges et al., 2007). At the same time, it has the potential to support people in developing critical thinking, reduce task workload, and provide a way to integrate individual with group knowledge (Stahl, 2006). In particular, face-to-face (f2f) collaboration provides benefits that are not easy to achieve in other forms of group work (Johnson, Sutton, & Poon, 2000; Olson, Teasley, Covi, & Olson, 2002). These include rich natural channels for continuous, multimodal communication; exchange of non-verbal cues; and potential for an increased productivity in completing tasks (Olson, Teasley, Covi, & Olson, 2002). Moreover, learning to collaborate effectively is a key 21st century skill for workforce and lifelong learning (Lee, Tsai, Chai, & Koh, 2014), and a competency often under-developed in graduates (Shum & Crick, 2016).

Collaboration is a complex mechanism; it commonly calls upon all group members to engage in a coordinated effort to create a joint solution to a problem (Dillenbourg, 1998). This frequently leads people to find that collaboration requires additional time and effort, compared with individual problem solving. This has been explained in terms of the cognitive and communication load added as the group members need to coordinate actions, address information conflicts and negotiate individual differences to achieve convergence (Kolfshoten & Brazier, 2013). Therefore, learning to collaborate effectively requires practice, awareness of group dynamics, reflection and, often, could be enhanced with coaching by an expert facilitator (e.g. a teacher, a tutor or a trainer). It also demands the development of social, coordination and communication skills (Mahyar & Tory, 2014), along with metacognitive and epistemic skills to gain understanding of the most effective group processes, strategies and best practices. To sum up, collaboration provides both potential benefits and challenges. Collaboration is complex and often not easy to understand or explain. Thus, it commonly requires people to learn how to collaborate with others by making the best of the social dynamics, the epistemic-related aspects (e.g. the approach to the task and knowledge structures) and the right tools that can support their tasks.

When people collaborate, some of their mental processes (which are otherwise hidden) are externalized to others in the form of dialogue, arguments, drawings, writing, models, etc. (Stahl, 2006). Knowledge and meaning are generated by each individual's participation in patterns of behavior and dialogue that are embedded in a socially-bound set of concerns (Winograd & Flores, 1985). Thus, the artifacts generated by the group and the dialogue allow group members to establish common ground, explain their personal views, reach consensus or produce new group knowledge (Scardamalia & Bereiter, 2006). As a result, collaboration provides particular opportunities for computer support when some of these traces of group interaction, behaviors and meaning-making processes can be digitally captured. These traces that collaborators leave behind while they interact face-to-face through different channels, can be analyzed and used to make collocated interactions *visible*, to find patterns of behavior, and to broaden the understanding of the collaborative activity in various contexts and domains. This could benefit researchers, developers and end-users (e.g. collaborators or facilitators). It could accelerate researchers' analysis cycles and developers' design iterations. Moreover, making group data available to

collaborators and facilitators in a timely manner may promote self-reflection and guide interventions in situ, rather than post hoc.

In this paper, we use the term *collaboration analytics* to refer to the particular techniques and approaches needed to automatically (or semi-automatically) capture, analyze, mine and distil data about collaborators' interactions. Thus, collaboration analytics innovations can be used to create mechanisms that help people (researchers, developers, collaborators and/or facilitators) to gain understanding of the complexity of collaboration, during or after the group activity, to support it more effectively. We are interested in those situations where people collaborate in *collocated* settings, using rich communication modes, such as speech, eye-contact, posture, all of which present particular challenges in terms of data capture and sense-making. The paper proposes a set of principles that can be useful and identifies dilemmas that can be encountered when mining such f2f multimodal interaction data. We integrate Data Science principles and techniques with the advances of interactive surface devices and sensing technologies. We build upon a 7-year research program which involved the analysis of six different small group work collaborative situations in collocated settings with 569 users and a variety of surface technologies, tasks, group formations and domains. Although our research has mostly been conducted in educational settings, the nature of the collaboration has much in common with far broader small-group collaboration. This means that the lessons learnt may be relevant to a broad range of training and workforce collocated contexts.

The rest of the paper is structured as follows. The next section discusses the term collaboration analytics for collocated environments in light of relevant research and development (R&D) on collaborative work. Section 3 presents related work addressing the analysis and visualization of group data. Section 4 describes the cases in our seven-year research program. Section 5 presents the challenges we have identified for deploying collaboration analytics. Section 6 presents principles and dilemmas identified from reflecting on these cases. Section 7 discusses the implications of collaboration analytics for design and development, and possible avenues for future work.

2. COLLOCATED COLLABORATION ANALYTICS: DATA SCIENCE MEETS COLLOCATED GROUPWARE

There has been substantial progress in the development of technologies that enable people to collaborate, mainly mediated via online systems. Large amounts of data can be captured as a result of the interaction of collaborators through these systems. While researchers have long engaged in manually constructing representations of human activity, computational techniques hold promise for the automated identification of significant patterns in the data, and close to real-time feedback. Just as many other fields of science that have taken a data-intensive turn, Human-Computer Interaction (HCI) and CSCW (R&D) has, for years, had structured "online" data to analyze digitally-mediated interaction (Dumais, Jeffries, Russell, Tang, & Teevan, 2014). However, it is only in recent years that the cost of sensors (e.g. microphone arrays, motion sensors) and surface devices (e.g. multi-touch displays, handheld devices) has fallen within an affordable range, and that the availability of analytics tools has broadened to a point that opens up a new range of applications to facilitate and enrich face-to-face activities in all sorts of contexts. Some foundational work has proposed guidelines (e.g. Kharrufa, Martinez-Maldonado, Kay, & Olivier, 2013; Scott, Grant, & Mandryk, 2003), recommendations (e.g. Mark, Carpenter, & Kobsa, 2003; Morris et al., 2006) and analysis frameworks (e.g. Jetter, Reiterer, & Geyer, 2014; Martinez-Maldonado, Goodyear, Kay, Thompson, & Carvalho, 2016) to enhance the design of

groupware systems in order to promote more effective collocated collaborative interaction. The Ubiquitous Computing field has also made a significant contribution by dramatically broadening previous notions of human-computer interaction, but it has had limited use of Data Science techniques (mostly on geo-localized personalization, - e.g. Chittaranjan, Blom, & Gatica-Perez, 2013; or users' profiling, - e.g. Eagle & Pentland, 2006), and certainly not focused on providing new forms of feedback to participants, or collaboration analytics tools for researchers.

There has been a growing interest in HCI (with key examples presented at CHI 2016) in mining interaction logs (e.g. clicks streams) to gain understanding of user's behaviors (e.g. Banovic, Buzali, Chevalier, Mankoff, & Dey, 2016; Fast, McGrath, Rajpurkar, & Bernstein, 2016; and Wang, Zhang, Tang, Zheng, & Zhao, 2016) and strategies (e.g. Fern, Komireddy, Grigoreanu, & Burnett, 2010). Likewise, in CSCW there has been interest in mining workflows and processes (e.g. van der Aalst, 2007) or social networks/meetings (e.g. Eagle & Pentland, 2006; and Kim, Chang, Holland, & Pentland, 2008) from interaction logs. The emerging boom in Data Science and the broader interest in data are producing a variety of tools and methods to analyze, visualize and support sense-making of these data. In short, data analysis tools such as descriptive modelling, visualization, predictive modelling (Banovic, Buzali, Chevalier, Mankoff, & Dey, 2016), text mining (Fast, McGrath, Rajpurkar, & Bernstein, 2016) and data mining (Fern, Komireddy, Grigoreanu, & Burnett, 2010), have been used in recent work in areas like HCI and CSCW to shed light on complex behavioral patterns which can be associated with higher order human activities (e.g. problem solving strategies as in Fern, Komireddy, Grigoreanu, & Burnett, 2010; or aspects of collaboration as in Martinez-Maldonado, Yacef, & Kay, 2013).

There has also been a growing interest in exploiting users' data for teaching and learning. Research fields such as Educational Data Mining (EDM) (Baker & Yacef, 2009), Artificial Intelligence in Education (AIED) (Roll & Wylie, 2016) and Learning Analytics (Siemens, 2013) focus on exploiting student data to enhance self-regulation (for students); to provide automated feedback, to improve coaching and authentic assessment (for students and teachers); or to support post-hoc analysis of collaboration, interventions, etc. (for researchers). However, most of the R&D in these fields has focused on networked collaborative learning environments, and it is only in recent years that educational Data Science has been brought to bear on collocated interaction with interactive surfaces, digital tangibles and sensors (Martinez-Maldonado et al., 2016).

In this paper, we introduce the term *collaboration analytics* to describe knowledge gained through the systematic study of the "data" emanating from collaboration networks with complex relationships between their entities (people, tasks, tools). The term has been used previously in a narrower sense, to refer to a specific technique to cluster behavioral data from users interacting with online collaboration tools (Wang, Jin, & Liu, 2010). However, our use includes an umbrella of techniques and approaches that can be used to facilitate the sense-making around data from collaborators' interactions. Particularly in collocated contexts, these collaboration data can be exceedingly heterogeneous and unstructured—text, images, video, application logs, location data — which creates specific challenges for data collection, processing, analysis and sense-making. Addressing these challenges requires a holistic and integrative approach: we need to incorporate Data Science principles and techniques with the advances in collaboration groupware and pervasive technologies, such as interactive surfaces and sensors. This is driven by the goal of bridging the gap between the information generated during the f2f activity and the digital support that can be offered to understand and improve collaboration. Data Science goes beyond the application of statistics or data mining techniques on data; it also emphasizes the display of

information for human consumption (Dhar, 2013), the sense-making process (Johnson & Myatt, 2014), people's confidence in the analysis (Dhar, 2013), the accountability of the algorithms that distil the data (Diakopoulos, 2015), making insights actionable (Dhar, 2013) and the ethical and practical issues that may arise when working with people's data (Richards & King, 2014; Slade & Prinsloo, 2013).

In summary, combining the R&D of these complementary areas offers a great opportunity for a new level of support for collocated collaboration. However, it is challenging to create innovative solutions that can exploit the collocated activity data to support understanding and improve f2f collaboration. At present, there is a significant gap in our knowledge of how to design such support tools: there are no clear guidelines or principles (either from Data Science or HCI perspectives) that can assist designers, researchers or developers in deploying collaboration analytics innovations in lab or real scenarios. This paper draws upon extensive experience in harnessing and mining collaboration data to make progress in tackling this gap in the literature.

3. RELATED WORK

This section presents the emerging research trends and more classical work at the intersection of CSCL/CSCW and the automatic analysis of group interaction data. The section describes two main types of studies. We first discuss work that has focused on *describing* and visualizing different aspects of collaboration, in both networked and f2f settings. Secondly, we discuss examples of studies that have exploited collaboration data using statistical, data mining or artificial intelligence (AI) tools for *diagnosing* (e.g. drill down dashboards, correlations, clustering), *predicting* (e.g. regression analysis, forecasting, multivariate statistics, pattern matching, predictive modelling, forecasting) or *prescribing* (e.g. graph analysis, simulation, complex event processing, neural networks, recommendation engines, machine learning). These approaches have been mostly conducted in non-collocated environments and they have also been tackled from different research perspectives such as HCI, the Learning Sciences, Data Science and CSCW.

3.1. Describing Collaboration Data: Visualization Approaches

There has been considerable work exploring the value of group visualizations to externalize the activity of groups and, in some cases, to reveal relationships between observable patterns, the quality of the group work or trends in collaborative student's activity. Frequently, simple approaches can enable group members to find very valuable patterns or motivate subsequent data exploration with more complex AI techniques (Witten & Frank, 2005). For example, Erickson et al. (1999) introduced the concept of social translucence to support computer-mediated communication by showing simple quantitative aspects of group participation. Authors described three properties, namely *visibility*, *awareness* and *accountability*, which systems can build upon to support self-regulation within groups. This work showed that simple visualizations of quantitative group data (within a chat room) can give a sense of the level of interaction, the amount of conversational activity, as well as indicating whether people are gathering or dispersing. Subsequently, other researchers have explored similar approaches in collocated settings by showing individual levels of individual participation (e.g. the Reflect table by Bachour, Kaplan, & Dillenbourg, 2010; the Conversation Clock by Bergstrom & Karahalios, 2007; the Second Messenger by DiMicco, Hollenbach, Pandolfo, & Bender, 2007; and the Meeting Mediator by Kim, Chang, Holland, & Pentland, 2008). More recent work has explored the impact of showing the same participation information in different formats (e.g. as a

competitive/ collaborative visualization as in Tausch, Ta, & Hussmann, 2016) or using visual metaphors (Kay, Maisonneuve, Yacef, & Reimann, 2006b; Tausch, Hausen, Kosan, Raltchev, & Hussmann, 2014). Whilst such work has mostly focused on visualizing conversations within small groups, the same principles can apply to visualize interactions in larger cohorts (e.g. Olguín et al., 2009).

Similar visualization techniques have been used extensively in the CSCL field to represent learner interactions and uncover some trends that could potentially be visually identified. Examples of these are the *sociograms*, which are unidirectional network graphs that depict interactions among collaborators that can be applied for small groups (e.g. Jermann et al., 2009) or larger networks (e.g. Sundararajan, 2010). Other work has explored the effects of mirroring visual representations of activity to the group, finding this useful for encouraging coordination and regulation of team members (Janssen, Erkens, Kanselaar, & Jaspers, 2007), identifying qualitative aspects of user's *participation* (Donath, 2002), observing changes on behavior between meetings (Roman, Mastrogiacomio, Mlotkowski, Kaplan, & Dillenbourg, 2012), or supporting reflection in long term team work (Kay, Maisonneuve, Yacef, & Reimann, 2006a). This previous body of work suggests that showing collaboration data to group members may play a critical role in facilitating their collaboration as long as the data is not too overwhelming and does not conflict with their collaboration dynamics. Overall, unlike the work on visualizing collocated conversations reported above, most of the CSCL work has been conducted in computer-mediated settings for remote collaboration.

Some investigations have been conducted to automatically or semi-automatically exploit data obtained from f2f settings where people interact with large shared displays. Some examples have explored the visualization of touch activity on interactive surfaces to illustrate aspects of collaboration such as territoriality (e.g. VisTaco by Tang, Pahud, Carpendale, & Buxton, 2010) and the impact of the arrangement of collaborators around the device on performance and roles (e.g. Rick et al., 2009). Al-Qaraghuli et al. (2011) developed more complex visualizations that display the actions that students performed in a timeline. In short, there is evidence that well-designed, compact information provided by visualizations can make various aspects of group interaction available (e.g. performance, social interactions, use of the tools). As a result, visualizing group data has the potential to help collaborators reflect about their activity or researchers to gain insights about the collaborative activity. The cases that we present in this paper have gone beyond this previous work on visualization and will be presented fully in Section 4.

3.2. Diagnosing, Predicting, and Prescribing for Collaboration Data

The use of machine learning techniques in collaborative settings has proven successful in gaining insights into the interactions within groups that lead to high-quality results in terms of conflict resolution (Prata et al., 2009), teamwork (Perera, Kay, Koprinska, Yacef, & Zaiane, 2009) and quality of the group outcome (Talavera & Gaudioso, 2004). In early research that applied data mining techniques to the study of collaborative learning processes, Soller et al. (2002) used Hidden-Markov Models to identify episodes when students shared knowledge when collaborating over a networked system. Talavera & Gaudioso (2004) formally pioneered the intersection between CSCL and data mining to study student's collaborative interactions to find patterns of behavior. They applied a clustering technique on e-learning data to build student profiles based on a set of features associated with each learner's interactions with the system. Building on this work, Anaya & Boticario (2011) proposed supervised classification and unsupervised clustering

techniques for grouping students according to their level of collaboration, assigning a value to each student to facilitate comparison of student's behavior. Tackling a similar problem, Duque & Bravo (2007) presented a fuzzy model that classifies the different forms of collaboration that lead to high quality solutions. Prata et al. (2009) presented an automated detector of the nature of the utterances written at a math online system to detect conflict between peers.

Some researchers have tackled the analysis of collaboration using sequence analysis. Perera et al. (2009) explored the use of sequence pattern mining and clustering to find trends in interaction that were associated with effective group-work behaviors in the context of software development collaboration. This approach enables sequential mining algorithms to discover frequent patterns of interaction including information from multiple users over several media. The patterns found can then be associated with behaviors that may distinguish strong from weak groups. Other techniques have also been used to mine sequential patterns from collaborative data including Hidden Markov Models (Soller & Lesgold, 2007), Social Network Analysis (Casillas & Daradoumis, 2009) and Process Mining (Reimann, Frerejean, & Thompson, 2009). Some of these analytics techniques are now being scaled up to be applied to Massive Open Online Courses (MOOC's, e.g. Zheng, Rosson, Shih, & Carroll, 2015).

Most of the previous examples have studied remote collaboration supported by online systems such as chat rooms, forums, wikis, intelligent tutoring systems and, more recently, MOOC's. In these contexts, where most (and often all) of the communication is mediated by the system, system can automatically log users' actions. This is in stark contrast with the challenges of capturing collaboration data in collocated environments. There is no previous work that has analyzed the fine-grained interweaving of student's speech and physical activity when working at f2f shared devices using data mining techniques besides the work presented in the cases in the following section.

4. THE COLLABORATIVE SITUATIONS: CASE STUDIES

We derived a set of principles from the lessons we learnt over a seven-year research program at the overlap of the areas: CSCW/CSCL, HCI (surface computing) and Data Science (data mining, analytics and AI). This section describes this research program by providing details of six different case-scenarios in Figure 1. All were collocated situations mostly focused on developing people's collaboration, professional and/or problem solving skills and supported by the use of multiple (shared and/or personal) devices. In all cases, multiple channels of data were captured from the collaborative activity in the form of click streams, touch data, application logs, mobility information, etc. In this section, we briefly describe each of the cases in terms of: the task, the social aspects, the tools used, the different *data classes* involved, the *purpose* of applying analytics innovations and the *target users* of the collaboration analytics (e.g. researchers, developers, collaborators and/or facilitators).

Data Classes. The term data class can be used to describe a type of data source commonly used for data analytics where people interact with and through computers. We used the following data classes to help us categorize our case studies: *process* (usage behavior including the time, location and context in which an action occurs in the collaborative process), *check points* (logged access to relevant resources and tool functions or completion of tasks), *performance* (qualitative or quantitative measures of user performance in completing the task), *communication* (evidence of information shared between collaborators for example, voice, conversation patterns or text) and *self-reports* (measures of satisfaction, preferences or other information reported by users).

Analytics Purpose. The purpose of the analytics is critical for identifying which particular analysis tools and techniques can and should be used to generate some value for a given collaborative situation. Ultimately, the potential value of a data-driven approach should justify the added effort of capturing and analyzing collaborators' data. At the same time, the purpose of the analytics plays a major role in aligning the technical elements (e.g. data classes, tools and analytics techniques) and the human aspects that need to be supported (e.g. sense-making, generation of understanding or the group activity itself). Examples of analytics purposes include predicting the quality of collaboration, mining patterns from behavioral traces indicative of group strategies, exploring the feasibility of a data-driven approach for a specific scenario, etc.

Case & technology	Data	Cohort	N	# Groups	Analytics techniques	Devices	Purpose of the analytics	Target users
1: A Multi-display Optimization Challenge	Voice, clickstream, app logs	Adults	39	13 triads	Prediction & classification	Vertical displays & pc's	Predicting collaboration	Researchers
2: A Pen-based Tabletop for 'Mystery' Solving	Pens, app logs	Children	18	6 triads	Prediction, clustering & visualization	Tabletop & pens	Mining collaborative strategies	Researchers
3: A Multi-touch Tabletop for Concept Mapping	Touch, voice, app logs	Adults	60	20 triads	Sequence and process mining	Tabletop, kinect & mic array	Predicting collaboration and differentiating strategies	Researchers & developers
4: A multi-display, Multi-touch Classroom	Touch, classroom logs, mobility, dashboard logs	HE students	376	22 classrooms 4 groups each	Sequence mining, clustering & visualization	Tabletops, vertical displays, tablets, kinects & mic arrays	Enhancing awareness and feedback	Facilitators
5: An Educational Design Studio	Human observations, mobility, tools usage	Designers & teachers	20	4 dyads and 4 triads	Visualization	Tabletop, interactive whiteboard, dashboard, tablets, pc & physical objects	Providing instant feedback	Collaborators
6: A Health-care Simulation Classroom	Mobility, voice, manikin logs, simulation status, self-reports	Health students	56	11 groups of 5-10 medical trainees	Exploration	Medical manikins & sensors	Providing delayed feedback	Collaborators

Figure 1: Overview of the six collaboration cases that will be discussed in the paper.

4.1. Case 1: A Multi-display Optimization Challenge

Task: In this case, groups were asked to perform the Job Shop Scheduling (JSS) task (Wallace, Scott, Stutz, Enns, & Inkpen, 2009), a problem that consists of optimizing the scheduling of six jobs, each composed of six ordered operations using six resources that can only be in use by one operation at a time. **Tools and setting:** In addition to a large, shared display projected on a nearby wall, participants were provided with interconnected laptops through which they could perform individual actions on the shared interface (see Figure 2, Case 1). The laptop interface provided a personally tailored view of the workspace, where the resources that the user could interact with were presented as more salient than the others. The large, shared display provided an overview of the group's shared interface. **Social aspects:** A total of 39 participants were organized in 13 groups of 3 members each. Participants were mostly students enrolled in university Math, Science or Engineering courses and aged 18-27 years.

Data: Each group was required to develop solutions for the JSS task 2 or 3 times. Data from 29 trials were collected and coded. Groups spent 17 minutes per trial on average and executed between 100 and 600 clickstream actions per solution, for a total of 9,800 recorded mouse click operations. All the verbal utterances were transcribed from the video recordings and the quality of collaboration was assessed by multiple observers through video analysis. These transcripts included a total of 4,836 verbal participations. Thus, the *data classes* involved in this case are *communication*, *process* and *performance*. The *purpose of the analytics* was to help researchers explore whether the quantitative traces of peers' interaction in a multi-display environment could be used to automatically predict the quality of collaboration observed by a human.



Case 1: A multi-display optimisation challenge



Case 2: A pen-based tabletop for 'Mystery' problem solving



Case 3: A multi-touch tabletop for concept mapping



Case 4: A multi-display, multi-touch classroom



Case 5: An educational design studio



Case 6: A health-care simulation classroom

Figure 2: The collocated collaborative situations. Written permission has been granted for the use of the two top-most figures for Case 1 (Wallace, Scott, Stutz, Enns, & Inkpen, 2009) and Case 2 (Kharrufa, Leat, Dolan, & Olivier, 2010).

4.2. Case 2: A Pen-based Tabletop for 'Mystery' Solving

Task: Participants were given an open question associated with a subject (such as mathematics or history) and a number of slips which may hold direct clues for solving the mystery, background information, or even red-herrings. They are asked to read the slips, discuss them and formulate an answer to the 'mystery' (Kharrufa, Leat, Dolan, & Olivier, 2010). **Tools and setting:** The tool called Digital Mysteries supported the mystery solving process. It was implemented using a pen-based tabletop that made it possible to identify the author of each action (see Figure 2, Case 2). **Social aspects:** The study involved 18 elementary school students aged between 11 and 14 years. They formed 6 groups of 3 participants. Some groups solved more than one mystery, generating a total of 12 sessions.

Data: A total of 17,130 tabletop pen-based actions were automatically logged. Out of the 12 sessions, 5 were coded as low achieving groups of students, 5 as high achieving groups and 2 as average groups. The level of achievement was coded based on: the quality of the discussions, the degree of logical thinking and the soundness of the justification for the solution of the mystery. No information about the conversation activity was recorded. Thus, the only *data classes* involved are *process* and *performance*. The *purpose of the analytics* was to perform a post-hoc research analysis aimed at exploring whether the stream of very low level actions (the pen interactions only) can provide any information about higher level collaborators' strategies that differentiate groups according to their performance in terms of quality of collaboration assessed by a human (the teacher). Thus, the analytics purpose consisted of using sequence mining the strategies that differentiate high from low performing groups.

4.3. Case 3: A Multi-touch Tabletop for Concept Mapping

Task: Participants were asked to read a one-page article on dietary guidelines and draw a concept map individually at a personal computer. They were then asked to join a group and build a joint concept map at an interactive tabletop which gave participants access to the concepts, linking words and an image of their individual maps. Finally, each participant was asked to draw an individual map again. **Tools and setting:** The tabletop interface allowed students to add concepts from their lists, create new concepts, open their individual maps, create propositions and change the concept map layout. The tabletop was augmented to identify which user did each touch input and the speech of each person. This was done automatically and unobtrusively using an overhead sensor and a microphone array (see Figure 2, Case 3). **Social aspects:** Sixty adults were recruited to participate in the study, in groups of three. All participants had the same role, thus the learning situation was intended to be egalitarian and symmetric in terms of roles, goals, expertise and participation.

Data: The raw data for each group consists of two long sequences of time stamped actions, including: verbal utterances and touch actions by each user. Possible actions on the tabletop included: creating, deleting, editing or moving a concept or link; and open, move, enlarge or close individual maps. The original filtered sequence obtained for each group had 434 to 1467 touch actions and 83 to 627 utterances. Each session was rated by two observers according to a collaboration rubric (Meier, Spada, & Rummel, 2007). The *data classes* involved in this case are *communication*, *process* and *performance*. The *purpose of the analytics* was to design an automatic approach to distinguish, discover and distil patterns of multimodal interaction to associate them with higher level group's strategies based on three data mining techniques: a classification model; sequential pattern mining; and clustering (thus predicting collaboration and

differentiating group strategies). The outputs of the analytics were aimed at generating understanding about group dynamics (for research purposes) and also at identifying frequent patterns that could inform the re-design of the group interface (for development purposes).

4.4. Case 4: A Multi-display, Multi-touch Classroom

Task: This study was run in-the-wild. Two sets of tutorials were taught for two university level undergraduate management units. There were 14 and 8 classroom tutorial slots respectively for each unit of fifty minutes each. The task involved discussions of a case study and the graphical representation of a solution of that case in the form of a concept map (using the same system described in Case 3) to support the small group discussions. **Tools and setting:** The setting was a multi-tabletop classroom (see Figure 2, Case 4) designed to capture the interactions of students and provide the teacher with the infrastructure to control, monitor and assess group activities. The system allowed the capture of user-identified actions and the interconnection of the components (tabletops, a teacher's dashboard, vertical displays and a database). All 4-5 tabletops were enhanced with a system for recognizing which student is touching the interactive surface. The teacher had a tablet showing a dashboard with visualizations, notifications and buttons to control the tabletops by sending commands to lock them or display messages. **Social aspects:** The units had 236 and 140 students respectively. In each tutorial session, the 15-20 students attending were randomly assigned to small groups at each table (3-6 students in each). Students knew their peers well as they had met over several weeks before.

Data: The data captured included: tabletop logs, snapshots of the evolution of each group's solution, and the actions performed by the teacher on the orchestration tool. The concept maps were automatically assessed by comparing them with a teacher 'ideal' map. Additionally, this work used participant's data, as well as observations, in order to capture the teacher's intentions, actions and reflections. No audio or video could be captured. The **data classes** in this case are *process*, *check points*, and *performance*. The **purpose of the analytics** was the provision of support for teachers (as facilitators), by enhancing their awareness of student's collaboration via a real-time dashboard, thus, enhancing teacher's awareness in the classroom on-the-fly. There was also an interface with other visualizations for after-class reflection.

4.5. Case 5: An Educational Design Studio

Task: Participants had to produce two or more candidate designs, satisficing competing design goals, of an authentic 13-week undergraduate course. This involved f2f meetings to resolve conflicting information, agree on the different design options, comply with institutional restrictions and create the learning designs using a design tool. **Tools and setting:** The setting was a multi-device Educational Design Studio equipped with a range of digital and physical tools and surfaces to support the collaborative design activity of small teams. These tools included a multi-touch tabletop connected to an interactive vertical display, tablet devices, a personal computer projected to a wall, a writeable white-wall, and various pen and paper materials. An analytics dashboard provided real-time quantitative indicators of the candidate designs. The design editor was displayed at both the tabletop and the vertical display. It provided digital elements to be manipulated by direct touch, allowing bi-manual input and fluid interaction with the visual representations of the design (see Figure 2, Case 5). **Social aspects:** Participants were mostly teachers and educational designers organized in four teams of three members each. Each team member was given one of three possible roles (Lecturer, Learning Designer and Quality Assurance Officer). According to their role, each user had specific information about the course

and goals. Some of the goals allocated to the participants complemented others' goals, and some were conflicting.

Data: The indicators displayed in the dashboard included i) a list of patterns added to each candidate design, ii) a pie chart that shows the total student time dedicated to each learning space, and iii) a histogram showing the student load for each week of the semester. Data about tools usage, attention and mobility was also captured mainly via video analysis. The **data classes** in this case are *process* and *check points*. The **purpose of the analytics** was to provide collaborators with metrics about the actual designs rather than learners' data (thus, enhancing awareness and real-time feedback).

4.6. Case 6: A Health-care Simulation Classroom

Task: This study was run in-the-wild. A set of 3-hour tutorial sessions was designed by the main teacher for the undergraduate unit: Integrated Nursing Practice. Students had to collectively put into practice their health care knowledge and skills to oversee a patient manikin in a hypothetical clinical scenario. **Tools and setting:** The simulation classrooms were equipped with 5-6 manikins placed on clinical beds which produce indicators of a patients' health, respond to actions and can be programmed to deteriorate over time (see Figure 2, Case 6). Each manikin provides a handheld device that students can use to report when they complete tasks. Beds were additionally equipped with a depth sensor and a microphone array to track participants around the bed and capture directional audio. **Social aspects:** Each student was asked to play one of five possible roles (e.g. the team leader, nurses, a family member, the doctor and an observer).

Data: Data was collected from 5 classroom sessions. Only the activity happening in two beds was recorded in each session to allow students to opt out from the recording. The data included: physical actions performed on the manikin (e.g. checking pulse, blood pressure), the handheld device logs, student's location, directional audio and video. The **data classes** involved were *mobility*, *process*, *check points* and *communication*. The **purpose of the analytics** was to explore whether it was feasible to obtain meaningful patterns about nursing students' collaboration in order to provide them with automated feedback.

5. CHALLENGES WITH COLLOCATED INTERACTION DATA

The variety of the cases presented above enabled us to build an empirically-based understanding of the complexity of collocated collaborative situations and the particular challenges of exploiting the rich multimodal interaction data that can be captured from such situations. As a starting point for crystalizing the lessons taken from this emerging body of work, we identified four different levels of **sense-making challenges**. These are critical in capturing the particular challenges of collocated collaboration analytics. We then used these to identify and organize the principles and dilemmas that will be presented in the next section. The challenges are defined as follows:

Challenge 1: **Complexity of the collaborative activity** – small-group collocated collaboration is inherently deeply complex. In addition to this, the nature of the particular collaborative task will further shape, restrict or focus what information *should* be valuable to capture, how to make sense of the collaboration data and the drivers for designing the interfaces for the analytics, all this taking account of the particular collaborative context. The foremost question to define and answer is: How can users (e.g. researchers or collaborators) make sense of

the data in ways that address the particular demands of the collaborative situation and the activity as a whole?

Challenge 2: **Multimodal data collection and aggregation in collocated settings** – capturing logs of the established conventional computer-mediated group-activity is far easier than is the case for collocated interactions. In the latter, communication has multiple channels, some of which are subtle and hard to capture (e.g. multi-device interaction, conversation, gestures, postures, gaze, etc). Furthermore, generating an exploitable dataset from such varied channels is still a technically very complex task. The questions here are primarily about the technical challenges that may limit what is actually feasible at present in relation to what is desirable to address: what data streams should be captured; how can these be aggregated; and what information can these data processing provide?

Challenge 3: **Data-driven insight** – this challenge is about how to bring the human into the analytics loop. It includes questions about the ways in which data is gathered, analyzed and presented: How might analytics augment the ability of all different stakeholders to reflect on their process? For instance, how might analytics provide insights for facilitators (e.g. teachers or trainers) on groups' processes, or on the efficacy of their teaching? In what ways can visualizations provide researchers with answers to their research questions?

Challenge 4: **Social and contextual issues** (such as ethical, legal, logistic, individual preferences) can impose practical limitations on what can and should be recorded from f2f scenarios or what can be analyzed by others or used by systems to perform automated actions. This ranges from the broadest social concerns, including those captured in legislation (such as management of personal data) to the particular demands of one collaborative context, where the individual preferences must be respected. What broad social and narrower contextual and personal concerns, can directly or indirectly affect the technical aims and sense-making processes?

Essentially, Challenge 1 is about the *desirable goals* for the collocated collaboration analytics. This should be the driver for aspirations of designers creating new ways to support collocated collaboration with analytics. Challenge 2 considers the *technical limitations and possibilities*, a fast moving area as new sensors, devices and infrastructures become increasingly available and where collocated collaboration analytics could be a driver for the design and building of new technical solutions. Where Challenge 2 deals with the input from multimodal data, Challenges 3 calls for designers to consider the particular **analytics tools** and **interfaces** that can serve the needs of the various stakeholders; this is a deeply human-computer interaction design challenge. Finally, Challenge 4 particularly relates to the critical concerns for moving beyond the laboratory to real-world deployments where this challenge must mediate all the others in terms of *people's concerns and preferences*, including individual preferences about the use of their personal data and the role of technology.

6. PRINCIPLES AND DILEMMAS

This section presents the principles and dilemmas derived from the above six case studies unfolded over the seven years of the project on collaboration analytics. Each informed the design of the research questions for the subsequent ones and each contributed to a growing understanding of the possibilities and challenges of harnessing emerging sensor technology and interfaces to create new forms of small-group collaboration analytics. The current section

represents both the insights over the program and based on reflection about the work as a whole. The principles and challenges we now present emerged through this two-part process: the growing understanding gained at each stage and the holistic reflection. The resulting principles and dilemmas are organized in four groups, each mapping to the challenges (presented above) under the following headings: 1) Collaborative activity; 2) Multimodal data; 3) Data-driven insight; and 4) Social and contextual issues. The following subsections introduce the principles and dilemmas that are indicative of the contributions of the work. We will provide the key underpinning background and the empirical foundations for them by illustrating situations where each emerged in the different cases.

6.1. Collaborative activity

6.1.1. Principle 1: Understanding collaboration complexity

Principle 1- Understanding the richness of collocated collaboration requires a **holistic approach** that takes account of four dimensions: physical, social, epistemic, and co-configuration (in the runtime enactment) of the design.

This principle is grounded on the view that group activity is not only shaped by the tools or technology design, but also by the loose relationship between users, tasks, technology available and the ways users adapt all these during the activity. In order to facilitate the identification of these aspects of group activity, we draw upon a theoretical perspective that established four dimensions of group activity: social, epistemic, set and the enactment of the design. In our case, we used the Activity-Centered Analysis and Design (ACAD) framework (Goodyear & Carvalho, 2014) as the core of our approach. ACAD can be used to scaffold the design processes that involve selection of specific methods for analyzing group experience as a whole (Martinez-Maldonado, Goodyear, Kay, Thompson, & Carvalho, 2016). The first three dimensions identified in ACAD are: (1) *the set (physical) component* – which includes the place in which participants’ activity unfolds, the physical and digital space and objects; the input devices, screens, software, digital documents, material tools, awareness tools currently available, and artifacts that are created in the collaborative processes; (2) *the social component* – which includes the variety of ways in which people might be grouped together (e.g. dyads, trios, groups); the scripted roles, divisions of labor, etc.; and (3) *the epistemic component* – which includes both implicit and explicit knowledge-oriented elements that shape the participants’ tasks and working methods. At runtime, when the groups actually do the collaborative activity using the available technology, the physical, social and epistemic elements are dynamically entangled as the group activity unfolds. But at design time, the physical, social and epistemic can be treated as discrete design components. What is designed in advance is then customized, selected from, added to, re-interpreted or otherwise modified by the people involved in the ensuing activity. We refer to this fourth component as (4) *co-configuration*. This dimension moderates the first three components which are the ones that designers can manipulate but then are appropriated, often in very different ways, by the actual users as they tackle problems collaboratively.

Having a holistic view of collaboration activity can sensitize the designer of the analytics innovation to look at the physical, the digital, the social and the epistemic dimensions of activity and to distinguish these (partially) designable elements from emergent activity. It also completes the loop from observation/disentangling to sense-making of the analytics outputs and (re)design. We selected the ACAD framework because it provides a holistic view of group activity, taking careful account of the tasks that the users are tackling (epistemic), the ways they divide up labor

(social), the ways they use the various tools, surfaces and materials in the space (physical) and how ultimately the design intentions play in runtime (co-configuration). The description of the case studies in Section 4 clearly identified these dimensions of collaborative activity and in the columns of Figure 1 referring to the groups' features, and the technology, data and devices. This allowed us to consider any analytics outcome according to the different *designable* dimensions of the situation (physical, epistemic and social) in light of their runtime enactment and co-configuration.

6.1.2. Principle 2: Transferability




In non-located situations it has been a common practice to try to use the same analysis techniques or models to inspect users' click streams across different domains (e.g. see John, Prevas, Salvucci, & Koedinger, 2004). This may speed up the analysis process or enable the generalization of results (Koedinger, Cunningham, Skogsholm, & Leber, 2008). This can also be a step towards building standard or reusable tools (Romero & Ventura, 2010). It would clearly be very helpful if we can model the level of collaboration of groups, for example, across the quite diverse social, epistemic, and physical situations presented in Section 4. Would it be possible to transfer models from one context to another, and if so, what are the implications of doing so? Our next principle points to three aspects designers should consider to create new systems based on this transfer in terms of data classes and sense-making.

Principle 2- Transferring lessons and data models across group contexts can be valid and valuable for rapid data exploration but requires a deep understanding of: data classes, potential for generalization and the particularities of the context.

We illustrate this principle by presenting a model transfer between Cases 1 and 3 (Martinez-Maldonado, Kay, Wallace, & Yacef, 2011). A rule-based model of collaboration was learnt from the dataset collected in the multi-display setting described in Case 1 to categorize behavior in the interactive tabletop setting described in Case 3. The first requirement, to match both collaborative situations, is that both datasets were very similar in terms of definition of data classes. In both these cases, the differentiated individual interactions with the system could happen in parallel (*process* data class) in the form of click streams (Case 1) or identified touch data (Case 3). Each verbal participation was either coded by a human (Case 1) or automatically captured with a microphone array (Case 3). Each included the time and duration of each verbal intervention by each user (*communication* data class). The quality of collaboration was manually assessed for both datasets (*performance* data class). Drawing on the considerable work on collaborative learning theories (Dillenbourg, 1998) and collaboration supported by desktop computers (Jeong & Hmelo-Silver, 2010), *symmetry* was identified as a valuable indicator of collaboration in cases where the task was intended to involve egalitarian collaboration. For this, the audio and application log lines of the dataset of Case 1 (the multi-display case) were grouped forming blocks of logged activity of different time-frame sizes (for example, 30 seconds or 1 minute of activity). Classification models were then applied as a supervised machine learning problem (Best-First tree, C4.5 decision tree, Bayes-Net and naïve Bayes algorithms) with the aim of classifying each block of interaction as collaborative, non-collaborative or partially collaborative based on human video annotations.

Feature	Metric	Example blocks of activity					
		C	C	SC	SC	NC	NC
1-Physical participation	Number of actions	7	12	15	15	10	15
2- Verbal participation	Number of utterances	28	9	4	5	0	4
3-Talking time	Seconds	19.4	17	7.5	8	0	5
4-Number of talking participants	0, 1, 2 or 3	3	3	2	3	0	1
5-Talking time dispersion	Gini coefficient (G)*	.51	.28	.75	.6	1	1
6-Verbal participation dispersion	Gini coefficient (G)*	.36	.14	.75	.61	1	1
7-Physical participation dispersion	Gini coefficient (G)*	.88	.58	.53	.4	.2	.8

Classification outcomes
(blocks of 30 seconds of activity classified as):

-  **Collaborative:** all participated or they were aware of their peers' actions
-  **Somewhat Collaborative:** 1 or 2 members were unaware or if the group failed to communicate but they still tried to collaborate at some level
-  **Non-Collaborative:** split the task, working separately, or one did all the work

*Gini coefficient:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2 n^2 \mu}$$

Figure 3: Six example blocks of activity in Case 1 classified by the C4.5 decision tree algorithm based on seven features of activity indicating amount and dispersion of physical (click streams) and verbal (talking) participation. Blocks were classified as collaborative, somewhat collaborative or non-collaborative.

Figure 3 provides examples of the resulting classifications for six blocks of 30 seconds of activity and the features that were used to generate the model in Case 1. A similar table for Case 3 would have the same rows and columns (but, of course, different numbers). For example, the first row indicates that the measure of physical participation corresponds to the number of clicks performed by each group member in Case 1, which in Case 3 would correspond to the number of touches on the tabletop. The same applies for Rows 2 and 3 in both cases. Based on the tree-based models obtained from the first dataset, we obtained a set of simpler rules as follows:

$$(low)P_{talk} + (high)G_{talk} + (high)P_{physical} + (low)G_{physical} \rightarrow \text{Non Collaborative situations} \quad (1)$$

$$(high, medium)P_{talk} + (high)G_{talk} + (high)P_{physical} \rightarrow \text{Partial collaboration} \quad (2)$$

$$(high)P_{talk} + (low)G_{talk} + (low)P_{physical} + (high, med)G_{physical} \rightarrow \text{Collaborative situations} \quad (3)$$

where P_{talk} corresponds to verbal participation, $P_{physical}$ to physical participation (touches, clicks), G_{talk} to the Gini coefficient as indicator of symmetry of talk and $G_{physical}$ as an indicator of symmetry of physical actions.

The next step was to evaluate these rules using the dataset of Case 3 (the tabletop case). The log lines were grouped in blocks of the same size and the same attributes were calculated as in the first dataset. Then, a clustering technique (k-means) was used to reveal if there was any relationship between the rules from the rule-based model and the extent of collaboration (coded from the video as a gold-standard by human experts) of the new groups. The blocks of activity were clustered into three groups and clusters and were inspected to determine the presence of collaborative, non-collaborative or partly collaborative blocks. Results showed that these very simple rules proved to have considerable power in identifying extremes cases of collaboration (Rules 1 and 3). That is, the non-collaborative blocks were characterized by low level of talk, asymmetry in the conversation and high levels of physical action. By contrast, collaborative

blocks were characterized by high levels of symmetric conversation and less physical interactions with the system. The rules did not prove powerful enough to identify those collaborative situations that were not at the extremes.

The rules obtained seem plausible based on the fact that collaboration calls for action to coordinate group progress, and that this context which was intended to be egalitarian, called for all group members to participate. This is aligned with the theoretical knowledge about collaboration which indicates that it is not easy to define when a group is collaborating, even if experts observe directly the activity of groups (Dillenbourg, 1998). This suggests that these rules may well be reusable, both because they were effective in both these cases and because it is possible to explain what they mean so that a person could decide how to make use of an interface showing this analytics information. For example, the same rules could be put into play in collaborative scenarios to trigger notifications if a threshold number of blocks of low level of collaboration is detected. Figure 4 depicts examples of the coding of some groups from Case 1. Rows with many blue blocks (C), some in orange (SC) and few in light yellow (NC) tend to correspond to collaborative sessions (e.g. rows A to C). By contrast, rows with few blue or orange blocks suggest these were non-collaborative groups (e.g. F and G). This shows that some quite simple group aspects, such as symmetry, can be automatically processed to give useful measures of collocated group collaboration that might be incorporated into interfaces for some stakeholders. For example, even the presentation in Figure 2 may be useful for a teacher who needs to decide where to put her attention in a class. An interface with a suitable presentation of this information seems promising as it may be able to enable the teacher to see which groups seem to be failing to collaborate.

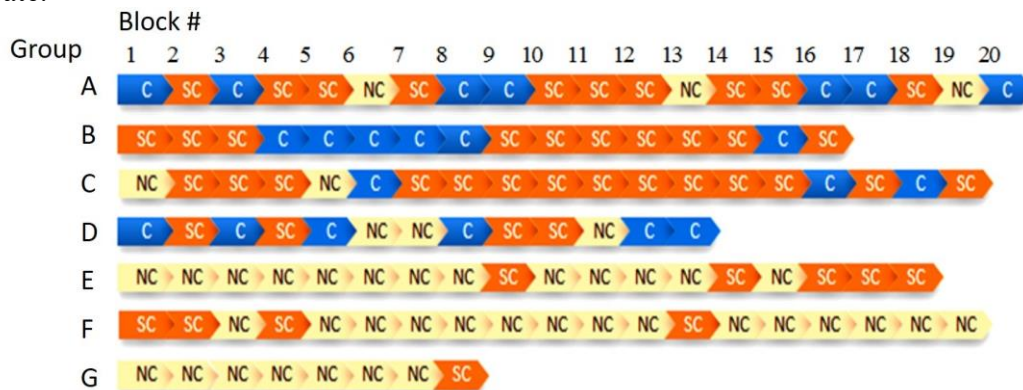


Figure 4: Analytics presentations of the coding of some analyzed sessions. Each 30 seconds of group work was tagged as “collaborative” (blue), somewhat collaborative (orange) or Non-collaborative (yellow). Diagram created using the process mining ProM (van Dongen, de Medeiros, Verbeek, Weijters, & van der Aalst, 2005) tool.

As specified in the first principle, collaboration, especially collocated collaboration, is far too complex to be determined by standard rules because it is strongly shaped by the nature of the task, the technology and the social rules. The models proved effective in helping identify extreme cases (e.g. clearly high or low collaboration). This means that there was high agreement with the classifications from the rules and those from the human expert in the gold-standard training corpus. However, these rules fall short in identifying groups whose behavior falls between the extremes of the spectrum of high or low collaboration (e.g. row D in Figure 4). This suggests we still need more work to determine whether it is feasible to create even more effective mining tools so that they can also deal with the intermediate cases, potentially by drawing on and integrating information about the domain and contextual knowledge into data mining algorithms.

6.1.3. Dilemma 1: Authenticity vs Experimentation

Another aspect of collaborative activity that can strongly shape the data capture process and the purpose of the analytics is whether the activity unfolds under controlled conditions (e.g. as an experiment, in the lab) or in-the-wild (e.g. in an authentic training environment, workplace or classroom). What limitations are imposed by the ‘authenticity’ of the scenario in terms of tasks, data capture, users’ goals, group dynamics, etc.? How can analytics approaches be transferred from lab experiments to in-the-wild settings? What is gained and what is lost in terms of data capture, analytics and sense-making? This section deals with the dilemma of performing collaboration analytics in the lab or under authentic conditions and the issues that may arise.

Dilemma 1: The dilemma between **authenticity** and controlled **experimentation** in data intensive collocated settings.

We illustrate this dilemma in terms of Cases 3 and 4. The tools that supported the collaborative activity in these two cases were the same: a multi-touch concept mapping application displayed on an interactive tabletop that could automatically identify the user touching the interface. The task design called for students to take egalitarian roles and share a common goal. The tasks were also similar, as both called upon users to build a concept map in response to a focus question. However, Case 3 was conducted under experimental conditions and Case 4 was conducted in an authentic classroom.

The first difference is in terms of data capture. In Case 3, users were recruited to participate in the activity, all groups were of the same size (20 groups of three members), they could be asked to respond to questionnaires, the activity was video recorded, directional audio was cleanly captured in a sound-proof room and the activity unfolded without interruptions for an hour or more, as needed. By contrast, when this activity was brought to the classroom (Case 4), the setting altered many of these aspects, so affecting the actual collaborative activity and the data collection. For example, audio could not be recorded because the system that worked in the lab could not operate in the very noisy conditions of regular classrooms. Nor was it acceptable to require each student to wear a microphone. There were other important changes: the teacher interrupted the activity, according to needs of each class as a whole; the group sizes varied according to the demands of classroom management, with students sometimes arriving late, joining groups already formed, etc. In short, this dilemma captures the conflicts a researcher faces in moving out of the lab. For example, authenticity may mean that one can no longer collect data that was available in the controlled lab setting. This, in turn impacts the kind of data analytics that can be applied.

In terms of analytics, the classroom teacher’s needs are critical design drivers. We illustrate the impact of this, as well as the limited data sources in the authentic classroom in Figure 5 which presents visualizations which, for simplicity, compare just two groups (although the actual interfaces in both cases had more groups). For Case 3, the two groups are called A and B and the interfaces for Case 4 are labelled for groups X1 and X2. The visualizations in Figure 5(1) are participation radars which show, at a glance, how much each group member participated. Although these visualizations show similar data, the visualizations used in the classroom (bottom left) do not show information about speech participation as in the top radars (where there are two triangles, blue for speech participation and red for touch data); and are simpler (e.g. guiding lines removed). This example shows how the context of use strongly affects the visual design.

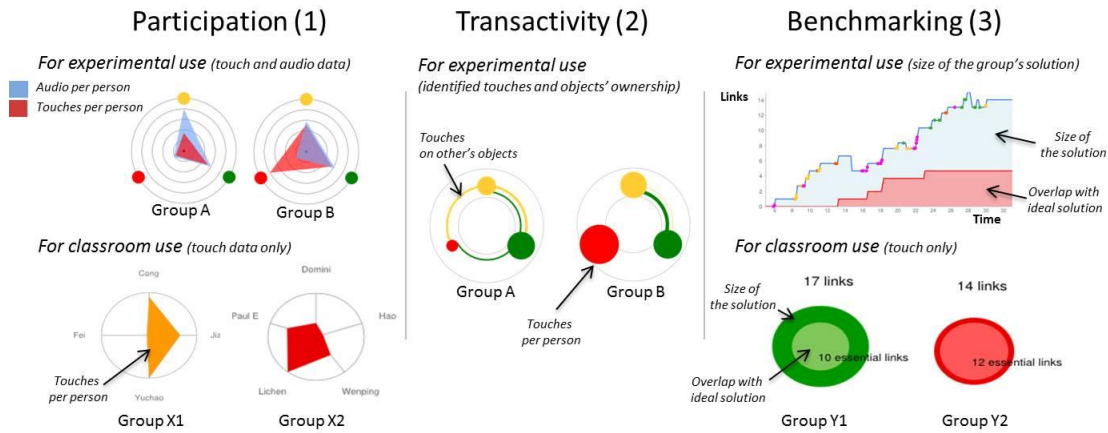


Figure 5: Left: radars of participation in the lab (top) and in the classroom (bottom). Middle is a lab only visualization comparing touch actions by each member of two groups. Right: Visualization showing the extent to which users interacted with others' objects in the lab (top) and a simpler visualization indicating the size and the distance of each map from the teacher's map for classroom use (bottom).

The visualizations at the center of Figure 5 show a different dilemma when comparing the two settings. The visualizations in Figure 5-2 depict the number of times that one user interacted with objects created by others on the tabletop. This measure is indicative of a collaborative process called “transactivity” (building on each other’s contribution), sometimes associated with highly collaborative groups. However, whilst visualizing transactivity proved useful for research purposes, for the study in the classroom, the teacher indicated that simpler information indicating the progress and some qualitative aspect of the students’ maps (such as benchmarking their maps with a teacher’s ideal solution) would be more useful in the hard constraint of a 1-hour scheduled class. As a result, the teacher dashboard had a simpler visualization showing the size of each group’s concept map and its distance from the teacher’s model solution, as in Figure 5(3) for groups Y1 and Y2 (bottom right) as opposed to what was offered to the teacher after a classroom session (e.g., Figure 5(3), top).

In terms of more sophisticated analytics, both cases also featured the analysis of patterns that could emerge from the long chain of actions recorded by the tabletop and/or the microphone. A sequential pattern mining technique was applied to both datasets to identify subsequences of actions that could be associated with higher level strategies. For example, some of patterns found are the following:

Sequential patterns from the lab dataset (Case 3): actions and speech

[Utterance]>[Utterance]>[MoveConcept-UtteranceByOther]>[AddLink-UtteranceByOther] (4)

[MoveConcept-NoUtterance]>[AddLink-NoUtterance]>[MoveConcept-NoUtterance] (5)

Sequential pattern from the classroom dataset (Case 4): actions and turn-taking

[MoveConcept-Other]>[AddLink-Same]>[MoveConcept-Same]> [AddLink-Parallel] (6)

[ShortInactivity]>[MoveConcept-Other]>[AddLink-Same] (7)

In Case 3, it was possible to combine the audio and touch data to mine patterns of speech and physical actions on the tabletop. For example, pattern (4) above illustrates a series of speech acts [Utterance] followed by two actions on the tabletop performed by one user [MoveConcept] while

a second user kept talking [UtteranceByOther]. Sub pattern (5) shows a series of three actions without any speech detected [NoUtterance]. Notably, these patterns were frequently found only in high and low collaboration groups respectively, thus, they were useful in differentiating cohorts. In the classroom, since the data were more limited (without audio) the sequences were simpler, focusing on who performed each action whether in parallel [Parallel] or taking turns [Same or Other – user performing the current action compared with the last action] (e.g. see pattern 6) or modelling periods without any touch activity detected as black boxes (see item [ShortInactivity] to represent this absence of touch data for a period of time in pattern 7).

Overall, moving from the lab into the *wild* not only has implications in terms of user interaction tailored to the classroom teacher’s needs but also in terms of data capture (e.g. what group data can be captured or how noisy these can be), data processing (e.g. what kind of analytics can be performed on-the-fly) and sense-making (e.g. what data representations can be really useful to support the group processes in situ).

6.2. Multimodal Data

6.2.1. Dilemma 2: Multimodality and Cost

As discussed briefly in the literature review, analytics initiatives that have looked at patterns from collaboration data have been mostly focused on only one dimension of the collaborative activity, namely the stream of actions of users to the system, conversation, or actions performed on artifacts. However, all these types of actions may occur in parallel. People communicate using voice but also use non-verbal cues (e.g. gaze, bodily gestures, postures, etc). Users can also interact with the system with shared devices (e.g. through multi-touch surfaces) or through their own devices (e.g. mobiles and tablets). These multiple modes of interaction with and through computers allow users to decide on different ways to approach the task, work in parallel, or divide the tasks according to their preferences and roles. This raises questions for analytics and underpins our following principle and two dilemmas in terms of: data multimodality, temporality and availability.

<p>Dilemma 2- The trade-off between richness of multimodal analytics and cost: the more collaboration dimensions are considered the better the story that can potentially be told, but at a cost.</p>
--

Ideally, there is a potential benefit from more data as this should give a better and more complete view of collaboration. It should enable the construction of better models and account for different dimensions of collaboration. For example, for some groups and for certain tasks, most of the collaboration occurs verbally (e.g. for tasks that involve discussion or argumentation). However, for other cases, actions may be more important as in tasks that require co-construction of artifacts, hands-on design or physical coordination. The main factors that we have identified limiting the multimodal data that can be captured in a collocated collaborative situation are the *cost* of data capture, the *feasibility* (limitations in current sensing technology and its deployment), the degree of *intrusiveness*, and the extent of *transferability* from the lab to the wild.

Cost. This refers to the cost of including the sensors in the setting and the algorithms that process raw data into meaningful information. For example, the easiest way to capture activity logs in the examples above was providing each user with a device (e.g. a personal computer or a

digital pen as in Figure 6, left). But shared devices that allow more flexible or natural interaction, such as tabletops, do not automatically log each user's data. To get the same kind of user-identified log data, additional methods are needed and these add to the cost of hardware and software system (e.g. adding a depth sensor to the tabletop like in Case 3 and 4, Figure 6, center).

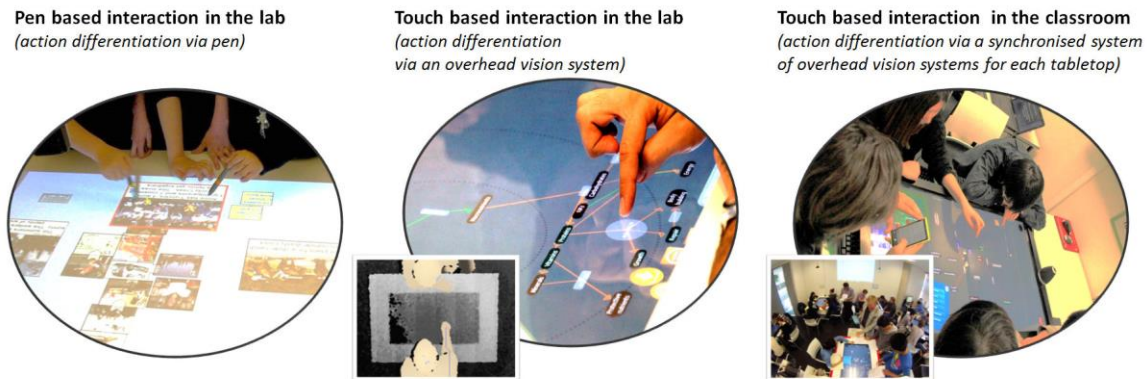


Figure 6: The trade-off of multimodal data capture and feasibility. Left: Pen-based interaction in Case 2. Centre: Using a Kinect (depth image shown over the left edge) to link each touch to a user for touch logging and where conversation patterns were recorded in Case 3. Right: Light-weight touch logging in the classroom in Case 4.

Feasibility. This refers to whether it is possible to actually collect data from a specific dimension of collaboration. For example, gaze data has proven to be valuable in providing information about where collaborators are looking and the degree of synchronicity as key indicators of collaboration (Schneider et al., 2016). However, the current state of technology makes this infeasible. For example, it may be possible if each user wore special glasses that track their gaze but this may be unacceptable. Similarly, other dimensions of collaboration data such as micro-mobility and group formations (Marquardt, Hinckley, & Greenberg, 2012), or using sensors to recognize body postures, facial expressions and other important non-verbal communication channels have been explored but only in very controlled experimental collaborative scenarios (Blikstein & Worsley, 2016).

Intrusiveness. This factor refers to the cases where the sensors can either affect the *behavior* of the group members, or *other epistemic or social aspects* for the sake of data capture. Examples of the former are the decision to use a microphone array instead of attaching lapel microphones to each user to record differentiated speech, or the use of a vision system to detect user's touch automatically as in Cases 3 and 4. An example of the latter includes modifications made to the task for the sake of data capture in Case 6, where the teacher asked students to self-report whenever they completed a sub-task (via a tablet checklist). This added an extra dimension of data capture besides the logs recorded automatically by the health manikins, the depth sensor and the microphone array.

Transferability. This refers to the approaches that can be deployed in experimental settings but cannot easily be transferred to realistic scenarios. A clear example is the transfer of both the data collection apparatus and analysis from Case 3 to 4. The touch data collection and analytics approaches could be transferred to the in-the-wild studies but the microphone arrays proved inadequate for this scenario due to the normal noise that is produced in the classroom (although it was attempted, see Figure 6, right).

6.2.2. Principle 3: Temporality

Principle 3- Time matters: consider the **temporality** and group **progression** as well as group outcomes and achievements.

Most automated analysis of collaboration has focused on either categorizing group members within groups (e.g. Anaya & Boticario, 2009; and Perera, Kay, Koprinska, Yacef, & Zaiane, 2009) or providing an overall view of group members' performance (e.g. Talavera & Gaudioso, 2004, and the visual analytics of level of collaboration developed in Cases 1 and 3). However, considering the temporality of actions is key to understand progression and get insights about the group processes. It can enable researchers to find patterns in low level actions where the order and/or timing of each interaction matter. For example, conversation patterns (e.g. to reach consensus or groundings in D tienne, 2006) followed by hands-on actions may be indicative of effective approaches to the task. In our cases we established that considering the temporality, or at least the order of events, can provide useful insights that can be seen at three different levels: sequences of *low level interaction events*, *episodes of group activity*, and *higher level collaboration sub-processes*. Next, we describe each of these levels with some examples.

Low level interaction events. This level refers to mining the low level atomic actions that users perform on the device(s) that can be associated with patterns of user activity at a higher level. It also includes simple counts of actions like participation in discussion boards and group meetings (e.g. Bachour, Kaplan, & Dillenbourg, 2008; and Bergstrom & Karahalios, 2007) to respond to simple questions about individual's contribution, symmetry of the group, interaction patterns and other quantitative aspects. The growing HCI interest in mining click streams to gain understanding of user's behaviors has been mostly done at this level (Banovic, Buzali, Chevalier, Mankoff, & Dey, 2016; Fast, McGrath, Rajpurkar, & Bernstein, 2016; Wang, Zhang, Tang, Zheng, & Zhao, 2016). In our cases, an example of this level of collaboration analytics was the application of sequence pattern mining followed by the clustering step to identify sub patterns of logged multi-touch data linked to higher order collaboration strategies (e.g. Case 2, see Section 6.1.3).

Fixed-length episodes of collaborative activity. This level refers to looking at fixed-length defined episodes of interaction of a sufficient length to identify key features from the data that can be indicative of some aspect of collaboration, for such an episode. This approach can be helpful when it is desirable to aggregate low level logged actions and it is hard to automatically identify variable-length collaboration sub-processes. An example of this type of collaboration analytics was used in Case 1 by generating models of activity for each block of detected activity in the multi-display environment using clickstream and conversation data (see Section 6.1.2). Different block sizes were explored (e.g. 30, 60 and 90 seconds). Figure 4 illustrated the output of the prediction model using the C4.5 tree algorithm applied to classify all the blocks of group activity for a group. Although this approach did not provide a high fidelity description of what occurred in each episode, it provided a quick and automated overview of the likely level of collaboration of a group based on quantitative data. This could be used by a facilitator, for example to investigate more deeply in manageable chunks of time what happened during the group work.

Collaboration sub-processes. This level refers to identifying the (variable length) sub-processes of collaboration. Unlike the episodes of group activity, here we refer to the medium level sub-processes or aggregated actions that can be associated with specific sub tasks in the

collaborative process. In open-ended collaborative activities, these sub-processes can commence and end at different time frames and also co-occur. Examples of collaborative sub-processes can include negotiating different perspective, establishing common ground, brainstorming ideas, or hands-on work. Figure 7 shows two examples of visualizations of these group sub-processes generated by aggregating the logs of group data from Case 3. These feature aggregated actions for the concept mapping task, with sub-tasks such as *creation and laying out of concepts* (yellow and orange), *linking concept and revising links* (light/dark blue), *consultation of external resources* (broken red lines) and *absence of touch activity* (blank spaces). The two types of visualizations helped to recognize a discontinuous group process followed by some groups. For example, Group A presented three sub-processes corresponding to sharing individual knowledge – broken red lines, in Figure 7, top right- followed or accompanied by other sub-processes. By contrast, Group B followed a fairly linear process with clear transitions from one sub-process to the next and where group members never shared their individual maps (see visualizations in the bottom of the figure).

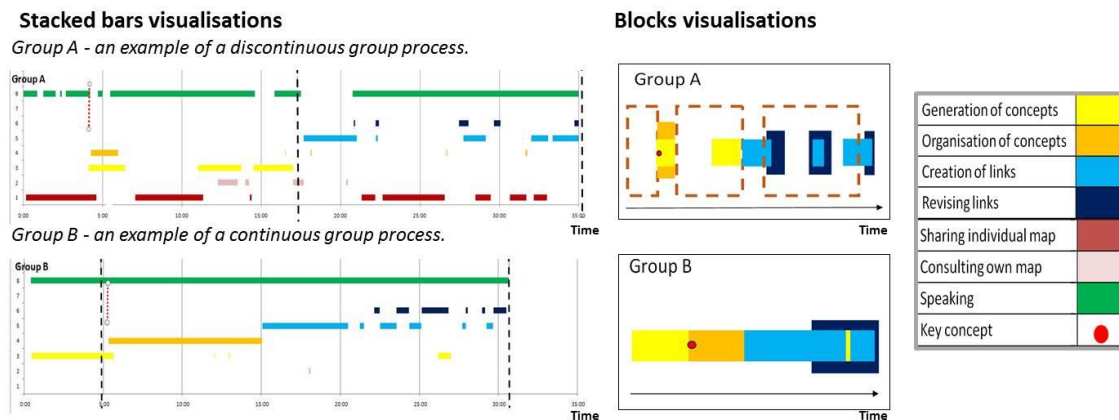


Figure 7: Visualization of group processes for two groups presenting A) a discontinuous and B) a continuous processes. Left: a stacked bar visualization. Right: a condensed block-based symbolic representation of the group process.

6.2.3. Dilemma 3: Data Availability

Dilemma 3- The dilemma of **data availability**: the trade-off between prioritizing high quality data capture and allowing user-preferred tools and practices.

Data availability is a crucial restriction for collaborative analytics. Data availability refers to the capability of a system of both capturing raw interaction logs or meaningful system actions, plus making these data accessible to third party entities for the purpose of data exploitation. However, to date, most software applications in general, and particularly educational technologies, are built without having a data collection feature in mind. This has a number of consequences. First, the data collected may not be rich and interpretable enough for useful analysis. Second, even if some data is collected, many systems (particularly off-the-shelf applications) are not open or do not provide an API to get access to relevant and fine-grained interaction data. To complicate the matter further, what if collaborators have their preferred “non-open” tools and have their own dynamics using these tools? How can the collaborative processes be supported based on data?

A potentially expensive solution is to create our own systems and convince collaborators to use them. For example, this was the scenario developed for Case 4: the multi-tabletop classroom. As we, developers and researchers, built all the software needed to run the learning activities

(including the tabletop software and backend) it was straightforward to exploit the data gathered in real time to create a database for generating visualizations and trigger notification on-the-fly. By contrast, although Case 6 presents a similar classroom scenario (with patient manikins rather than interactive tabletops), the kind of collaboration analytics that can be deployed are quite different. The manufacturer of manikins imposes limitations in terms of real time data access (data needs to be manually downloaded at the end of the session) making it impossible to offer real time feedback.

An alternative solution for the data availability problem that we have identified is to by-pass the system by adding sensing technology. In Case 6, for example, we used a depth sensor and a microphone to record additional information of user interaction besides what could be captured by the manikins. Another less ideal solution may be to adapt the practice or the task to force data availability, for example, in the same Case 6, the teacher requested students to self-report their key actions and created a role among the group members to make sure this task was completed.

It is currently too hard for people to harness their own data, particularly as data collection is rapidly becoming ubiquitous and pervasive (Abowd, 2012). Overall, there is a paradigm shift that is still needed for tool providers to overcome the inability to access to data collected by collaborative tools. The problem of data availability needs to be considered in the design phase in order to take the potential that data can offer to support collaboration to the next level. Moreover, this paradigm shift needs to be accompanied by the ethical and policy framing for data exploitation and management (Landau, 2015).

6.3. Data-driven Insight

A key part of the collaboration analytics process, in order to be effective, is to bring the human into the loop of the data/algorithm processing. Ideally, the results of the automated indicators or machine actions will be understood and inform decision making processes or prompt human actions or behaviors. There has been substantial research in the HCI and information visualization fields to bring final users in direct contact with data in terms of interaction design, perception, cognition, and decision-making (Holzinger, 2013; Ware, 2012). From a Data Science perspective, although there is a strong interest in using algorithms for finding previously unknown relationships in data, there is also the goal of helping human users in making sense of these relationships (Fayyad, Wierse, & Grinstein, 2002), storytelling with data (Kosara & Mackinlay, 2013), or provoking some actions based on data, thus, making data actionable. The next three principles define some of the actions that can be taken to enhance the opportunities for users to gain insights from data produced by diverse collaboration analytics algorithms.

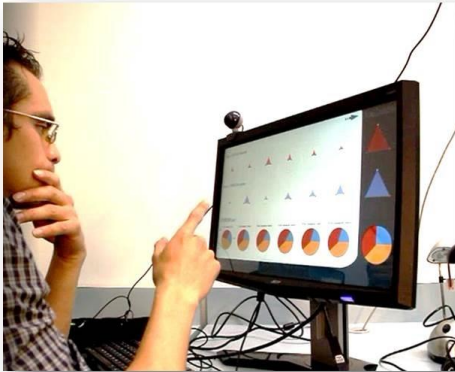
6.3.1. Principle 4: *Post-hoc vs On-the-fly*

Principle 4- The breadth and depth of the collaboration analytics provided to the user are strongly shaped by whether they support post-hoc or on-the-fly reflection.
--

The type, the levels of detail and even the ways the outputs of the collaboration analytics are visualized are shaped by the timing of usage of the user interface. For example, in an authentic classroom, a teacher has limited time and capacity for individual attention to make an intervention, and so needs timely information that can be almost instantly assimilated. Whilst more complex information will be interesting for research or to inform future versions of the task or learning materials, the types of interfaces with the data that can be produced need to be

designed to address practical demands in-the-wild. In this way, more detailed information can be provided for post-hoc analysis (e.g. see Figure 8, left, showing a dashboard to enable a teacher to scrutinize the interaction processes over time for one group of students in Case 3) whilst simpler and more succinct information was needed to support awareness on the fly (e.g. see Figure 8, right, showing the handheld teacher’s dashboard used in Case 4 where there were 5 groups in the class).

A dashboard for after class use
(post-hoc inspection of temporal Student’s data)



A dashboard for in-class use
(on-the-fly inspection of student’s performance data)



Figure 8: The different type of collaboration analytics provided for teachers to analyze post-hoc (left, as a controlled experimentation for Case 3) or in an authentic environment (right, in the classroom for Case 4).

6.3.2. Principle 5: Units of Analysis

Principle 5- Varying the unit of analysis (e.g. individuals, groups, cohorts, devices, artifacts/objects) widens the possible insights that can be gained from the collaboration analytics.

The analysis of interaction data in computer-mediated collaborative systems has typically focused on individual’s actions with the system and synchronous and/or asynchronous interaction with others (e.g. chat threads, messages sent, comments on other’s contributions). With these data, analytics researchers have mostly focused on individuals’ behaviors or actions performed on objects. From our cases, we identified six general units of analysis that can provide different insights in collocated collaborative settings. These units of analysis are: a) the individual; b) a role within a group; c) the group as a whole; d) cohorts; e) devices; f) artifacts/objects. Most of these are quite generic and can be applied to non-collocated settings as well. However, the fundamental point of this principle is on identifying the range of units of analysis that can be helpful for a designer to formulate collocated collaboration analytics innovations. Next, we illustrate how each of these units of analysis played a role in our cases to tackle different questions.

The individual: In collocated environments, differentiating individual’s actions may allow, for example, to: create visualizations showing each individual’s contribution or who interacts with whom; find patterns where authorship is considered (e.g. like the patterns shown in Section 6.1.3 for Cases 3 and 4) or classify students according to their individual behaviors that can be related to profiles within groups such as leaders, free-riders, doers, ‘speakers’, etc (e.g. these ‘profiles’ were identified in Case 3). However, shared devices commonly do support

authentication (log in), user-identification and the ability to track who performed each action (with some exceptions such as the Diamond Touch by Dietz & Leigh, 2001; or some experimental personal device handshake solutions such as the ones developed by Ackad, Clayphan, Martinez-Maldonado, & Kay, 2012; and Schmidt, Chehimi, Rukzio, & Gellersen, 2010; which may not be practical for authentic use). In our cases, differentiated individual's streams of actions were recorded in different ways such as users interacting through personal devices; using personal pointers; or using a vision based tracking sensors.

Objects and Artifacts: An alternative way to analyze group interactions is to focus on the actions performed by individuals involving the objects or artifacts displayed on the interactive interface. For example, in Case 2, we performed sequential mining to find patterns of actions performed on specific digital objects located on the interactive tabletop in order to identify group strategies such as whether one or more users interacted with the same object consecutively, continuously or for sporadic periods of time. Alternatively, in Case 4 we provided visual analytics about the group artifacts (measures of size and quality of concept maps). This approach works mostly for artifacts that remain active during the whole collaboration activity.

Roles: Similarly to identifying individual patterns, the unit of analysis could be focused on identifying differences among established roles enacted by different users. For example, in Case 5 analysis of mobility was performed to identify how individuals used the collaborative space according to their role, producing colored mobility maps like the ones in Figure 9 (left). These data facilitate insights on the impact the epistemic and social scaffolding had on the collaborative activity. In this example, each user had different goals and tasks, according to their roles. The lecturer and designer were the more active participants in terms of hands-on activity and the participants playing the quality assurance officer role remained mostly as observers and engaged with the analytics dashboard more often as it supported the main goals for this role.

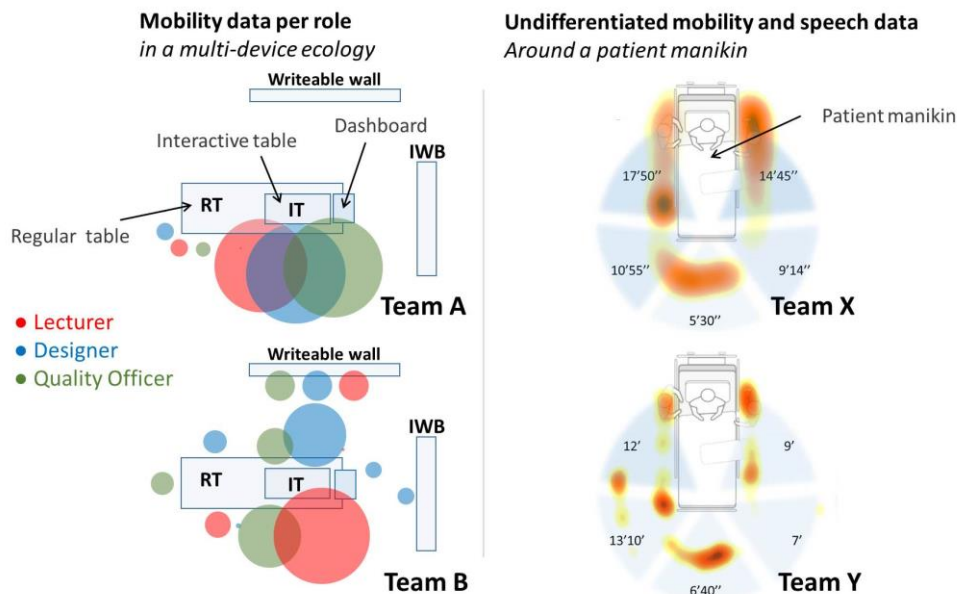


Figure 9: Varying the units of analysis. For example, mobility data, focusing on roles as in Case 5 (left) or on small groups as a whole without differentiating individuals as in Case 6 (right).

The group as a whole: In practical terms, the information generated by a group of individuals as a whole (e.g. without user-differentiation of actions or contributions) can provide key insights about the overall collaboration (e.g. to determine if a group is collaborating as in Cases 2 and 3) or to generate visualizations of group progress that, in our classroom studies in Case 4, were more effective in providing insights to the teacher in the classroom than information about individuals within a group. Additionally, undifferentiated individual traces of activity can illustrate group processes, also overcoming the limitations of the sensors in obtaining differentiated logs. For example, Figure 9 (right) shows the mobility data of two groups of students working around a patient manikin in Case 6. In this case, the data does not differentiate students but just identifies when someone was in a certain position at certain time. The top-right heatmap shows one group (X) that spent more time near the patient than the other (Y, bottom-right) which may be meaningful for a teacher or the students themselves, especially in conjunction with data about problems in their activity, pointing to adjustments required in the activity for one of these groups.

Cohorts: Both the individuals and the groups can further be grouped in terms of performance measures, preferences or other features. With this information, they can be grouped into higher level clusters to perform cohort analysis. In such a case, the unit of analysis still is either the individual (e.g. high or low achieving or active students) or the group (e.g. groups that presented low or high levels of achievement or collaboration) but the analysis is performed to identify differences among groups. An example of this type of collaboration analytics was performed in Cases 1, 3 and 4 where we applied sequence mining to identify patterns of behavior associated with high or low group achievers according to their performance (Case 1) or quality of collaboration (Cases 3 and 4).

Devices: In multi-device collocated settings the different devices utilized can capture different traces of information. In our multi-device setting presented in Case 5, the focus of attention can be on how collaborators interact with the ecology of devices (e.g. tablet devices, vertical displays, the interactive tabletop or personal computers) for different tasks. For example, Figure 10 shows aggregated tool usage and individual's visual attention on the different tools and displays within the Design Studio of Case 5. This illustrates how four groups using the ecology of devices, accomplished the same task, in quite different extents under similar social and epistemic situations.

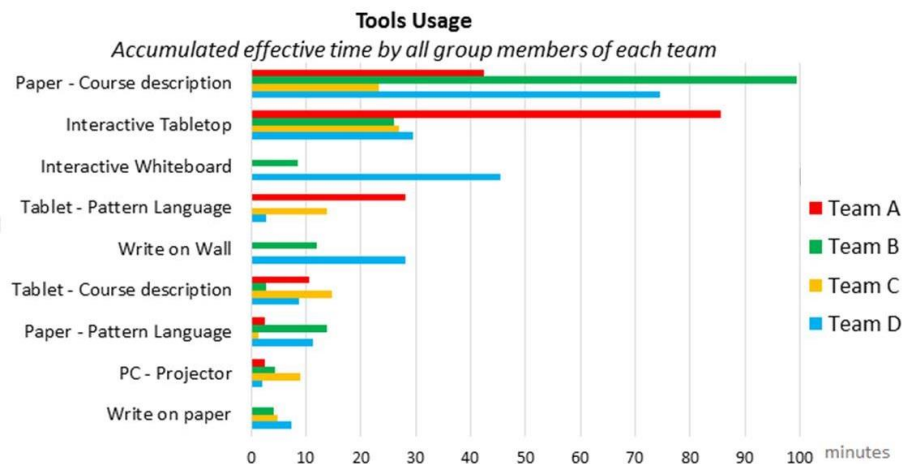


Figure 10: Tools usage and attention dedicated by teams to the different tools and displays within the Design Studio of Case 5.

6.3.3. Principle 6: Theory/hypothesis-driven Context Encoding

Principle 6- Using alphabets for **encoding contextual information**, driven by theory or hypotheses, can facilitate making sequential results meaningful.

Multimodal collaboration data is particularly complex. This complexity comes from the need to combine data of multiple users, devices and/or artifacts, each generating different classes of data. For example, Figure 11 illustrates the logs generated in Case 3. This shows three users interacting with the tabletop, discussing, making changes to a digital artifact, and an algorithm actively assessing it to detect misconceptions. The key question here is how to pre-process these data in ways that can be understandable by current algorithms and, at the same time, in ways that can help to provide meaningful insights.

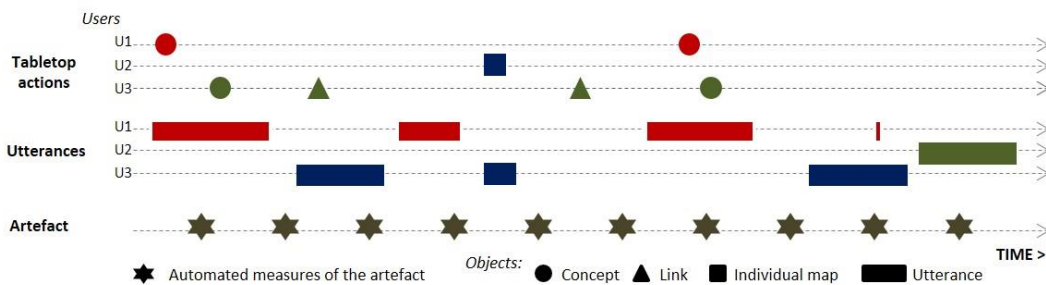


Figure 11: Example of dimensions of collocated interaction for the concept mapping interactive tabletop case (Case 3) by modality (e.g. touch, verbal communication, changes in the group artifact) and differentiated individual's action in time.

One of the data mining techniques used to investigate the impact of the order of student's actions is sequence pattern mining. The raw data for each group initially consists of two long sequences of actions: evidence of verbal speech by each learner and identified touch actions. These were transformed into a list of time-stamped actions performed on specific objects and verbal utterances. Although the raw format of these may appear simple, the level of complexity of these data is actually high when contextual information is considered. For example, each student action on the multi-touch device can be performed while she/he or other students are speaking, similarly, the same sequence of actions can be performed by different students in multiple combinations. The challenge was how to pre-process the data to make it ready for a sequence pattern mining algorithm and include contextual information for each action to facilitate the human consumption of the algorithm outputs. A possible solution is to code each student's action with all the possible contextual information (e.g. if the action was performed on his/her own object of one of others', if there is talking happening in parallel, who performed the previous action, etc.). However, this action would be very high dimensionality, making it unlikely to find meaningful patterns (e.g. there is more chance to find more slightly different patterns that may not provide any insight into the collaboration activity).

Thus, we identified hypotheses, some inspired by theoretical knowledge of collaborative learning (e.g. Stahl, 2006), to guide the pre-processing of the data. The first hypothesis referred to the possibility of finding sequential patterns that distinguish more collaborative from less collaborative groups by the interwoven stream of students' verbal and physical participation. The link between this hypothesis and the data is made through an *encoding alphabet*. An alphabet can be used to encode student actions into sequential items that capture contextual information. For

example, Figure 12 (left) shows the alphabet that was used to encode the dataset to investigate the hypothesis. A short example of how three actions would be encoded with this alphabet is shown in Figure 12 (right). The sequence starts with an *add concept* action performed by the first user (U1) accompanied by an utterance by the same person which was encoded as: *Add-Concept-SAAuthor*. Then, U3 modifies the text of a link while the first user is still talking: *Edit-Link-SOther-OtherAuthor*. U3 then opens her individual concept map while speaking: *Open-Map-SAAuthor-SameAuthor*. Afterwards, U1 verbally responds with a short utterance: *Speech-Short*.

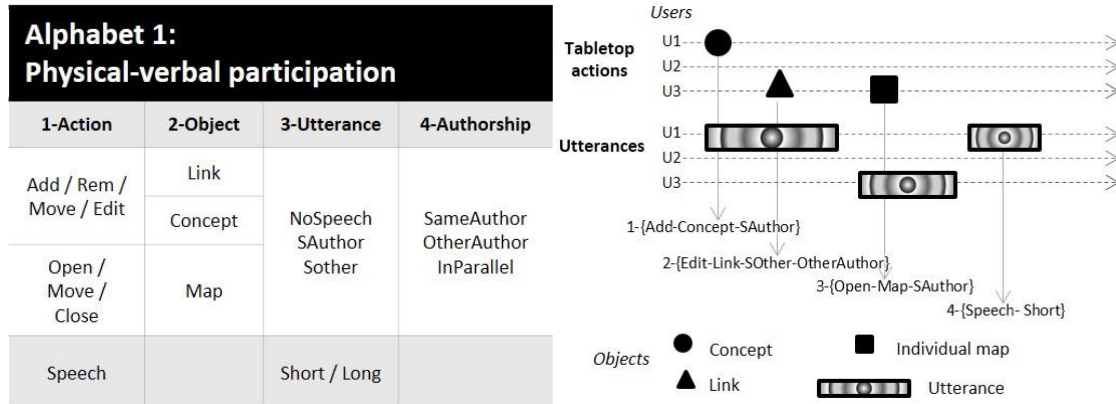


Figure 12: Left: Example alphabet used to encode group members' types of actions (1) performed on objects (2); with contextual information, linked to (3) speech - if action occurred while no-one was talking –NoSpeech, the same person was talking –SAuthor, or a another was talking -SOther; and (4) authorship - if the same or another person performed the previous action – SameAuthor, OtherAuthor – or two actions in parallel -InParallel. Right: Encoding 4 group's events using alphabet 1.

By applying the sequence pattern mining algorithm on this processed dataset, we obtained valuable insights about group collaboration at the tabletop. For example, we discovered that the less collaborative groups had a predomination of patterns with physical interactions, high levels of physical concurrency and greater parallelism than the more collaborative groups (with patterns such as *Move-Concept-NoSpeech-SameAuthor*>*Add-Concept-NoSpeech-InParallel*). By contrast, the more collaborative groups had more verbal discussions in conjunction with physical actions, especially in the initial phases (with patterns such as *Speech-Short*>*Move-Concept-SOther-SameAuthor*>*Move-Concept-SOther-SameAuthor*). Thus, drawing on theory or hypotheses (as in this example) for pre-processing the collaboration data can increase the possibilities of obtaining meaningful patterns from the multi-dimensional dataset.

6.4. Social and Contextual Issues

6.4.1. Principle 7: Notice and Consent

Although there is an increasing interest in data driven approaches to understand human behavior, there are already some cases, particularly in training and learning contexts, that have been halted by reason of concerns raised by stakeholders, governmental entities or civil rights groups about privacy and ethics in regards of personal data handling (Drachsler & Greller, 2016). These concerns apply to large datasets collected through online systems but privacy, consent and data management issues also need to be taken into consideration for deploying analytics in authentic collocated settings.

Principle 7- Particular **notice** and **consent** strategies must be operationalized to address organizational, ethical and/or data privacy issues emerging in authentic deployments of small-group collaboration analytics.

Developing risk management strategies is a common practice to address any possible ethical issues in general contexts (Richards & King, 2014; Slade & Prinsloo, 2013). In collocated settings, it is crucial to address ethics and privacy concerns particularly because personal data may not be easy to be de-identified (for example, if the data capture involves video or similar vision systems recording or conversation audio capture) and multimodal data capture where the user may not always be aware about what the system is or is not recording (e.g. when capturing gestures, mobility, postures, gaze, etc.). Even further, people may not realize how they will appear in the analytics; for example, a person who has had low levels of contribution may not be aware of that until they see the analytics interface. Another key challenge, beyond research studies, is that a whole group's data may only be used if every member agrees to this; should a person decide to withdraw from a study for any reason, this means a loss of that whole group's data. These concerns apply in laboratory studies but are even more challenging in authentic deployments. For example, in a classroom, students may be concerned about possible effects on their grades. Ways to address these issues commonly involve a comprehensive consideration of strategies to *inform* or give *notice* users about all the aspects of the data management in order to allow them to give explicit *consent* without affecting their regular activities (Landau, 2015). However, for authentic deployments, where data innovations are intended to become part of the day-to-day activities, different ethical and organizational strategies must be used, bringing different opportunities or limitations for the collocated data analytics innovations.

Our cases provided at least three different approaches that can be followed to implement strategies to address organizational and ethical issues for the data collection (e.g. privacy, ethics, consent) and for the data management (e.g. data ownership, provenance, storage). For example, the students recruited in Case 3 (an experimental study) experienced a procedure similar to that of any ethically-aware experimental study follows (with the participants receiving an information sheet with all the information about how their data will be captured, stored and managed, signing a consent form and being advised of their right to withdraw at any time).

Case 4 was a real classroom where all students are to be given the same opportunities of learning (thus, the situation is not experimental but just part of their day-to-day learning activities). In line with normal practice, the use of their data to drive the teacher's dashboard did not require student approval. However, for the research, students were asked to give consent for data analysis only by researchers to mine students' interaction with the tabletops. To facilitate this process, no video data was captured and a simple consent form was sent to the students; this was important as it avoided eating into precious classroom time just for signing paper-based forms.

Similarly, for Case 6 (also a classroom/training setting) blanket authorization was granted by the ethics committee of the institution to pervasively collect de-identified data (e.g. this includes mobility data captured by a depth sensor or system's logs). Then, during the study, only two out of the 5 clinical beds were recorded (there were always unmonitored beds in each class). This allowed students to voluntarily participate, or not. Importantly, students in this faculty were used to the occasional recording session being organized as it was a common practice in this learning space. At present, this is more of an exception than the norm in other classrooms. So, work is needed to establish acceptable and manageable ways to request consent and clarify data

management issues without affecting already limited classroom time. In the end, this principle will differ over time and context, according to accepted practice in monitoring group activity.

7. CONCLUSION AND DISCUSSION

An intriguing question is how well current and emerging sensor technologies capture data needed to track collocated human interaction at the speed and resolution required to test theories (a researcher perspective), improve practice (coach perspective) and provide better feedback (employee/student perspective). Researchers have studied such processes with painstaking care using manual techniques, but technical developments mean that soon, the everyday infrastructure will be capable of generating such data. The key challenge for HCI researchers and designers is to demonstrate and validate examples of how this can work, and distil design principles that can guide the design of such data-driven approaches and systems. This paper proposes a step forward towards shedding initial light towards addressing this question.

This paper has proposed candidate principles, and moreover, documented some of the dilemmas that arise. Figure 13 presents an overview of the principles and dilemmas discussed earlier in the paper, and how the case studies served to illustrate each. The cases represent a rich and diverse range of collaboration analytics, crossing various goals for the analytics, different classes of multimodal data and analytics methods as well as settings. In some cases, we demonstrated that we were able to harness the richness of the multimodal data to make patterns of the interwoven verbal and physical interaction events visible (e.g. by applying sequential pattern mining in Case 3). By contrast, we also showed that even quite simple data can actually be very useful in contexts such as in a classroom (e.g. Case 4). One of the key differences between experimental, lab settings and more authentic situations, such as a classroom, is in the type of analytics interfaces that are acceptable and can be helpful for the different purposes of the analytics. For example, a lab setting may allow researchers to gain a deeper understanding of group processes (e.g. by classifying episodes of collaboration as in Case 1, or generating models of group processes such as in Case 3) or an educator to explore the data more deeply as they reflect on each group. But in the in-the-wild classroom studies, teachers wanted a dashboard for quick decision-making (e.g. Case 4) or for users to get immediate feedback (e.g. Case 6).

In terms of the principles that we proposed, embracing the complexity of collaboration is a first key step to generate understanding about collaboration in specific scenarios (Principle 1). This can allow identification of social, epistemic and toolset characteristics that may help in explaining why some groups work more effectively than others, or why some strategies are more effective in particular situations. This holistic view of collaboration also can help to generalize (or not) lessons learnt and the use of analytics tools across different scenarios (Principle 2). In more technical terms, the data that can be automatically captured from group collaboration can be viewed from multiple perspectives, particularly, if multiple modes of interaction are tracked (e.g. application logs, changes on the group artefacts, communication, etc). In all cases, considering the temporality of actions and group progression is as (or more) important as the final group achievements (Principle 3). No matter how ‘small’ the data captured from groups is, the breadth and depth of the collaboration analytics can be strongly shaped by the focus of attention on the analytics (e.g. individuals, groups, cohorts, devices, and artifacts/objects) and whether the support is being provided on-the-fly or for post-hoc reflection (Principles 4 and 5). Rich contextual information is critical for facilitating sense-making of analytical results, thus sensing and encoding techniques that include contextual information should be considered for inclusion into the analytics loop. Finally, the potential for inappropriate surveillance is clear, as is the risk

that the algorithms analyzing the data are opaque to relevant stakeholders and may be subject to bias (e.g. there are emerging forums dedicated to issues associated with fairness, accountability, and transparency in machine learning, such as FATML - fatml.org).

Challenge	Type	Principles and dilemmas	Illustrative cases	Theme
Collaborative Activity	Principle 1	Understanding the richness of collocated collaboration requires a holistic approach that takes account of four dimensions: physical, social, epistemic, and co-configuration of the design.	All	Understanding collaboration complexity
	Principle 2	Transferring lessons and data models across group contexts can be valid and valuable for rapid data exploration but requires a deep understanding of: data classes, potential for generalization and the particularities of the context.	1 & 3	Transferability
	Dilemma 1	The dilemma between authenticity and controlled experimentation in data intensive collocated settings.	3 & 4	Authenticity
Multimodal Data	Dilemma 2	The trade-off between richness of multimodal analytics and cost : the more collaboration dimensions are considered the better the story that can be told, but at a cost.	2, 3, 4 & 6	Multimodality
	Principle 3	Time matters: consider the temporality and group progression as well as group outcomes and achievements.	1, 2 & 3	Temporality
	Dilemma 3	The dilemma of data availability : the trade-off between prioritizing high quality data capture and allowing user-preferred tools and practices.	4 & 6	Data availability
Data-Driven Insight	Principle 4	The breadth and depth of the collaboration analytics provided to the user are strongly shaped by whether they support on-the-fly or post-hoc reflection	3 & 4	Post-hoc versus on-the-fly
	Principle 5	Varying the unit of analysis (e.g. individuals, groups, cohorts, devices, and artifacts/objects) widens the possible insights that can be gained from the collaboration analytics.	1, 2, 3, 4 & 5	User identification & units of analysis
	Principle 6	Using alphabets for encoding contextual information , driven by theory or hypotheses, can facilitate making sequential results meaningful.	2, 3 & 4	(Theory driven) context encoding
Social and Contextual Issues	Principle 7	Particular notice and consent strategies must be operationalized to address organizational, ethical and/or data privacy issues emerging in authentic deployments of small-group collaboration analytics.	3, 4 & 6	Notice and consent

Figure 13: Overview of the principles and dilemmas discussed in the paper.

In terms of the dilemmas that we identified, there was one critical challenge that we found in several research projects that utilized analytics tools and techniques to understand collaboration; this challenge is that insights obtained under controlled experimentation conditions can differ in important ways from what can be observed or even implemented in authentic scenarios (Dilemma 1). This is also linked with the cost of capturing multiple dimensions of collaboration (Dilemma 2). For example, some data that can be relatively easily tracked in a lab experiment (such as eye gaze, clean audio, and physiological indicators) can be much more challenging to capture in-the-wild with current technology. If such tracking is possible at all, it may incur a higher cost or introduce significant levels noise in the data. Ultimately, not all the collaborative tools and sensors make the data available (Dilemma 3). Some systems log user interactions but do not make these data available. If they do, they may impose certain barriers for data availability such as requiring complicated protocols to get the data or not providing physical or programmatic

ways to get the data. Most systems simply do not keep interaction logs at all, and many that do fail to have logs that have enough of the right information to be mined effectively.

7.1. Limitations and Future Work

Although the cases present diverse collocated collaboration contexts, the aim was not to be exhaustive in describing of all the possible collaboration analytics approaches that can be applied. For example, other collocated analytics, such as joint gaze (Schneider, et al., 2016) and posture analyses (Blikstein & Worsley, 2016), have also been explored nonetheless only under very controlled experimental conditions. In particular, we have not touched on the sort of analyses of the content of speech utterances (which is strong in non-collocated collaborative settings, e.g. Rosé et al., 2008) that can provide a quite additional dimension to the analysis of collaboration. This also brings additional concerns in tracking spoken language (Klasnja, Consolvo, Choudhury, Beckwith, & Hightower, 2009). This would also facilitate accounting for more complex analyses of group interactions that nowadays are commonly accomplished through qualitative analysis, for example, eliciting implicit group sub-processes from dialogues or identifying subtle within-group dependencies. Additionally, much more work needs to be done to explore the potential of negatively influencing group behavior by making traces of activity more visible as has been suggested, for example, by Wilson et al. (2006) and Bachour et al. (2008).

Finally, we can note that the cost of some of the technologies used in the studies showcased in the paper is dropping all the time, which will democratize access to what used to be esoteric research instrumentation. As noted in different parts of the paper, R&D based on data-driven approaches to understand or support collocated collaboration is still in its infancy if compared with analytics and data mining solutions in computer-mediated collaboration scenarios. We hope that the principles and dilemmas discussed in this paper will provide guidance to, and influence the methodologies followed by, designers, developers and researchers creating innovative systems that are intended to discover important trends or patterns from the digital traces of interaction; predict outcomes such as levels of collaboration; or simply present collaboration data in ways that can enable the users to digest the digital traces and make sense of key aspects of collaboration.

REFERENCES

- Abowd, G. D. (2012). What next, ubicomp?: celebrating an intellectual disappearing act. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. New York: ACM.
- Ackad, C., Clayphan, A., Martinez-Maldonado, R., & Kay, J. (2012). Seamless and continuous user identification for interactive tabletops using personal device handshaking and body tracking. *Proceedings of the Work-In-Progress CHI 2012 Conference on Human Factors in Computer Systems*. New York: ACM.
- Al-Qaraghuli, A., Zaman, H., Olivier, P., Kharrufa, A., & Ahmad, A. (2011). Analysing Tabletop Based Computer Supported Collaborative Learning Data through Visualization *Proceedings of the IVIC 2011 Conference on Visual Informatics* Berlin: Springer.
- Anaya, A., & Boticario, J. (2009). Clustering learners according to their collaboration. *Proceedings of the CSCWD 2009 Conference on Computer Supported Cooperative Work in Design*. IEEE.
- Anaya, A., & Boticario, J. (2011). Application of machine learning techniques to analyse student interactions and improve the collaboration process. *Expert Systems with Applications*, 38(2), 1171-1181.
- Bachour, K., Kaplan, F., & Dillenbourg, P. (2008). Reflect : An Interactive Table for Regulating Face-to-Face Collaborative Learning. *Proceedings of the EC-TEL 2008 European Conference on Technology Enhanced Learning*. Berlin, Heidelberg: Springer.

- Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Transactions on Learning Technologies*, 3(3), 203-213.
- Baker, R. S., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1(1), 3-17.
- Banovic, N., Buzali, T., Chevalier, F., Mankoff, J., & Dey, A. K. (2016). Modeling and Understanding Human Routine Behavior. *Proceedings of the CHI 2016 Conference on Human Factors in Computer Systems*. New York: ACM.
- Bergstrom, T., & Karahalios, K. (2007). Conversation Clock: Visualizing audio patterns in co-located groups. *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*. IEEE Computer Society.
- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220-238.
- Casillas, L., & Daradoumis, T. (2009). Knowledge extraction and representation of collaborative activity through ontology based and Social Network Analysis technologies. *BI and Data Mining*, 4(2), 141-158.
- Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, 17(3), 433-450.
- Détienne, F. (2006). Collaborative design: Managing task interdependencies and multiple perspectives. *Interacting with Computers*, 18(1), 1-20.
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64-73.
- Diakopoulos, N. (2015). Algorithmic Accountability. *Digital Journalism*, 3(3), 398-415.
- Dietz, P., & Leigh, D. (2001). DiamondTouch: a multi-user touch technology. *Proceedings of the UIST 2001 Symposium on User Interface Software and Technology* New York: ACM.
- Dillenbourg, P. (1998). What do you mean by 'collaborative learning'? In P. Dillenbourg (Ed.), *Collaborative Learning: Cognitive and Computational Approaches. Advances in Learning and Instruction Series.*, (pp. 1-19). Oxford: Elsevier Science.
- DiMicco, J. M., Hollenbach, K. J., Pandolfo, A., & Bender, W. (2007). The impact of increased awareness while face-to-face. *Human-Computer Interaction*, 22(1), 47-96.
- Donath, J. (2002). A semantic approach to visualizing online conversations. *Communications of the ACM*, 45(4), 45-49.
- Drachler, H., & Greller, W. (2016). Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. *Proceedings of the LAK 2016 International Conference on Learning Analytics & Knowledge*. New York: ACM.
- Dumais, S., Jeffries, R., Russell, D. M., Tang, D., & Teevan, J. (2014). Understanding User Behavior Through Log Data and Analysis. In S. J. Olson & A. W. Kellogg (Eds.), *Ways of Knowing in HCI*, (pp. 349-372). New York: Springer.
- Duque, R., & Bravo, C. (2007). A Method to Classify Collaboration in CSCW Systems. In B. Beliczynski, A. Zielinski, M. Iwanowski & B. Ribeiro (Eds.), *Adaptive and Natural Computing Algorithms*, (pp. 649-656). Berlin, Heidelberg: Springer.
- Eagle, N., & Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4), 255-268.
- Erickson, T., Smith, D., Kellogg, W., Laff, M., Richards, J., & Bradner, E. (1999). Socially translucent systems: social proxies, persistent conversation, and the design of "babble". *Proceedings of the CHI 1999 Conference on Human Factors in Computer Systems*. New York: ACM.
- Fast, E., McGrath, W., Rajpurkar, P., & Bernstein, M. S. (2016). Augur: Mining Human Behaviors from Fiction to Power Interactive Systems. *Proceedings of the CHI 2016 Conference on Human Factors in Computer Systems*. New York: ACM.
- Fayyad, U. M., Wierse, A., & Grinstein, G. G. (2002). *Information visualization in data mining and knowledge discovery*. San Diego, CA: Morgan Kaufmann.
- Fern, X., Komireddy, C., Grigoreanu, V., & Burnett, M. (2010). Mining problem-solving strategies from HCI data. *ACM Transactions on Computer-Human Interaction*, 17(1), 1-22.
- Goodyear, P., & Carvalho, L. (2014). Framing the analysis of learning network architectures. In L. Carvalho & P. Goodyear (Eds.), *The architecture of productive learning networks*, (pp. 48-70). New York, NY: Routledge.
- Gross, T. (2013). Supporting Effortless Coordination: 25 Years of Awareness Research. *Computer Supported Cooperative Work*, 22(4-6), 425-474.

- Hilliges, O., Terrenghi, L., Boring, S., Kim, D., Richter, H., & Butz, A. (2007). Designing for collaborative creative problem solving. *Proceedings of the 2007 Conference on Creativity and Cognition*. New York: ACM.
- Holzinger, A. (2013). Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What Is the Benefit of Bringing Those Two Fields to Work Together? *Proceedings of the 2013 Conference on Availability, Reliability, and Security in Information Systems*. Berlin, Heidelberg: Springer
- Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education*, 49(4), 1037-1065.
- Jeong, H., & Hmelo-Silver, C. E. (2010). An Overview of CSCL Methodologies. *Proceedings of the ICLS 2010 9th International Conference of the Learning Sciences ISLS*.
- Jermann, P., Zufferey, G., Schneider, B., Lucci, A., Lepine, S., & Dillenbourg, P. (2009). Physical space and division of labor around a tabletop tangible simulation. *Proceedings of the CSCL 2009 International Conference on Computer Supported Collaborative Learning*. ISLS.
- Jetter, H.-C., Reiterer, H., & Geyer, F. (2014). Blended Interaction: understanding natural human-computer interaction in post-WIMP interactive spaces. *Personal and Ubiquitous Computing*, 18(5), 1139-1158.
- John, B. E., Prevas, K., Salvucci, D. D., & Koedinger, K. (2004). Predictive human performance modeling made easy. *Proceedings of the CHI 2004 Conference on Human Factors in Computer Systems*. New York: ACM.
- Johnson, D. M., Sutton, P., & Poon, J. (2000). Face-to-Face vs. CMC: Student communication in a technologically rich learning environment. *Proceedings of the ASCILITE 2000 17th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education*. ASCILITE.
- Johnson, W. P., & Myatt, G. J. (2014). *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*. Hoboken, NJ: Wiley.
- Kay, J., Maisonneuve, N., Yacef, K., & Reimann, P. (2006a). The Big Five and Visualisations of a Team Work Activity. *Proceedings of the ITS 2006 International Conference on Intelligent Tutoring Systems*. Berlin: Springer.
- Kay, J., Maisonneuve, N., Yacef, K., & Reimann, P. (2006b). *Wattle tree: what'll it tell us?* , (Technical Report). Sydney: University of Sydney.
- Kharrufa, A., Leat, D., Dolan, P., & Olivier, P. (2010). Digital mysteries: designing for learning at the tabletop. *Proceedings of the ITS 2010 International Conference on Interactive Tabletops and Surfaces*. New York: ACM.
- Kharrufa, A., Martinez-Maldonado, R., Kay, J., & Olivier, P. (2013). Extending tabletop application design to the classroom. *Proceedings of the ITS 2013 International Conference on Interactive Tabletops and Surfaces*. New York: ACM.
- Kim, T., Chang, A., Holland, L., & Pentland, A. S. (2008). Meeting mediator: enhancing group collaboration using sociometric feedback. *Proceedings of the CSCW 2008 International Conference on Computer Supported Cooperative Work*. New York: ACM.
- Klasnja, P., Consolvo, S., Choudhury, T., Beckwith, R., & Hightower, J. (2009). Exploring privacy concerns about personal sensing. *Proceedings of the IEEE PerCom 2009 International Conference on Pervasive Computing*. Berlin: Springer.
- Koedinger, K., Cunningham, K., Skogsholm, A., & Leber, B. (2008). An open repository and analysis tools for fine-grained, longitudinal learner data. *Proceedings of the EDM 2008 International Conference on Educational Data Mining*. IEDMS.
- Kolfschoten, G. L., & Brazier, F. M. (2013). Cognitive load in collaboration: convergence. *Group Decision and Negotiation*, 22(5), 975-996.
- Kosara, R., & Mackinlay, J. (2013). Storytelling: The next step for visualization. *Computer*, (5), 44-50.
- Landau, S. (2015). Control use of data to protect privacy. *Science*, 347(6221), 504-506.
- Lee, K., Tsai, P. S., Chai, C. S., & Koh, J. H. L. (2014). Students' perceptions of self-directed learning and collaborative learning with and without technology. *Journal of Computer Assisted Learning*, 30(5), 425-437.
- Mahyar, N., & Tory, M. (2014). Supporting communication and coordination in collaborative sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1633-1642.
- Mark, G., Carpenter, K., & Kobsa, A. (2003). A model of synchronous collaborative information visualization. *Proceedings of the IV 2003 International Conference on Information Visualization*. Los Alamitos, CA: IEEE.

- Marquardt, N., Hinckley, K., & Greenberg, S. (2012). Cross-device interaction via micro-mobility and formations. *Proceedings of the UIST 2012 Symposium on User Interface Software and Technology*. New York: ACM.
- Martinez-Maldonado, R., Goodyear, P., Kay, J., Thompson, K., & Carvalho, L. (2016). An Actionable Approach to Understand Group Experience in Complex, Multi-surface Spaces. *Proceedings of the CHI 2016 Conference on Human Factors in Computing Systems*. New York: ACM.
- Martinez-Maldonado, R., Kay, J., Wallace, J., & Yacef, K. (2011). Modelling symmetry of activity as an indicator of collocated group collaboration. *Proceedings of the UMAP 2011 International Conference on User Modeling, Adaptation and Personalization*. Berlin: Springer.
- Martinez-Maldonado, R., Schneider, B., Charleer, S., Shum, S. B., Klerkx, J., & Duval, E. (2016). Interactive surfaces and learning analytics: data, orchestration aspects, pedagogical uses and challenges. *Proceedings of the LAK 2016 International Conference on Learning Analytics & Knowledge*. New York: ACM.
- Martinez-Maldonado, R., Yacef, K., & Kay, J. (2013). Data Mining in the Classroom: Discovering Groups' Strategies at a Multi-tabletop Environment. *Proceedings of the EDM 2013 International Conference on Educational Data Mining*. IEDMS.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 63-86.
- Morris, M. R., Cassanego, A., Paepcke, A., Winograd, T., Piper, A. M., & Huang, A. (2006). Mediating Group Dynamics through Tabletop Interface Design. *IEEE Computer Graphics and Applications*, 26(5), 65-73.
- Olguín, D. O., Waber, B. N., Kim, T., Mohan, A., Ara, K., & Pentland, A. (2009). Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics*, 39(1), 43-55.
- Olson, J. S., Teasley, S., Covi, L., & Olson, G. (2002). The (currently) unique advantages of collocated work. In P. J. Hinds & S. Kiesler (Eds.), *Distributed work: New research on working across distance using technology* (pp. 113-136). Cambridge, MA: MIT Press.
- Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. (2009). Clustering and Sequential Pattern Mining of Online Collaborative Learning Data. *IEEE Transactions on Knowledge and Data Engineering* 21(6), 759-772.
- Prata, D., Baker, R., Costa, E., Rosé, C., Cui, Y., & de Carvalho, A. (2009). Detecting and Understanding the Impact of Cognitive and Interpersonal Conflict in Computer Supported Collaborative Learning Environments. *Proceedings of the EDM 2009 International Conference on Educational Data Mining*. IEDMS.
- Reimann, P., Frerejean, J., & Thompson, K. (2009). Using process mining to identify models of group decision making in chat data. *Proceedings of the CSCW 2009 International Conference on Computer Supported Collaborative Learning*. ISLS.
- Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest Law Reviews*, 49, 393-432.
- 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 IDC 2009 International Conference on Interaction Design and Children*. New York: ACM.
- Roll, I., & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, 26(2), 582-599.
- Roman, F., Mastrogiacomo, S., Mlotkowski, D., Kaplan, F., & Dillenbourg, P. (2012). Can a table regulate participation in top level managers' meetings? *Proceedings of the GROUP 2012 International Conference on Supporting Group Work*. New York: ACM.
- Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics*, 40(6), 601-618.
- Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 237-271.
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. S. (Ed.) (Ed.), *The Cambridge Handbook of the Learning Sciences*, (pp. 97-115). New York: Cambridge University Press.
- Schmidt, D., Chehimi, F., Rukzio, E., & Gellersen, H. (2010). PhoneTouch: a technique for direct phone interaction on surfaces. *Proceedings of the UIST 2010 Symposium on User Interface, Software and Technology*. New York: ACM.

- Schmidt, K., & Bannon, L. (2013). Constructing CSCW: The First Quarter Century. *Computer Supported Cooperative Work*, 22(4-6), 345-372.
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Detecting Collaborative Dynamics Using Mobile Eye-Trackers. *Proceedings of the ICLS 2016 International Conference of the Learning Sciences*. ISLS.
- Scott, S. D., Grant, K. D., & Mandryk, R. L. (2003). System guidelines for co-located, collaborative work on a tabletop display. *Proceedings of the ECCSCW 2003 European Conference on Computer Supported Cooperative Work*. Norwell, MA, USA: Kluwer Academic Publishers.
- Shum, S. B., & Crick, R. D. (2016). Learning Analytics for 21st Century Competencies. *Journal of Learning Analytics*, 3(2), 6-21.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400.
- Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, 57(10), 1510-1529.
- Soller, A., & Lesgold, A. (2007). Modeling the process of collaborative learning. In U. Hoppe, H. Ogata & A. Soller (Eds.), *The Role of Technology in CSCL*, (pp. 63-86). Berlin: Springer.
- Soller, A., Wiebe, J., & Lesgold, A. (2002). A machine learning approach to assessing knowledge sharing during collaborative learning activities. *Proceedings of the CSCL 2012 International Conference on Computer Support for Collaborative Learning*. ISLS.
- Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge*. Cambridge, USA: MIT Press
- Sundararajan, B. (2010). Emergence of the Most Knowledgeable Other (MKO): Social Network Analysis of Chat and Bulletin Board Conversations in a CSCL System. *Electronic Journal of e-Learning*, 8(2), 191-208.
- Talavera, L., & Gaudioso, E. (2004). Mining Student Data to Characterize Similar Behavior Groups in Unstructured Collaboration Spaces. *Proceedings of the ECAI 2004 Workshop on Artificial Intelligence in CSCL - European Conference on Artificial Intelligence*
- Tang, A., Pahud, M., Carpendale, S., & Buxton, B. (2010). VisTACO: visualizing tabletop collaboration. *Proceedings of the ITS 2010 International Conference on Interactive Tabletops and Surfaces*. New York: ACM.
- Tausch, S., Hausen, D., Kosan, I., Raltchev, A., & Hussmann, H. (2014). Groupgarden: supporting brainstorming through a metaphorical group mirror on table or wall. *Proceedings of the NordCHI 2014 Nordic Conference on Human-Computer Interaction*. New York: ACM.
- Tausch, S., Ta, S., & Hussmann, H. (2016). A Comparison of Cooperative and Competitive Visualizations for Co-located Collaboration. *Proceedings of the CHI 2016 Conference on Human Factors in Computing Systems*. New York: ACM.
- van der Aalst, W. M. (2007). Exploring the CSCW spectrum using process mining. *Advanced Engineering Informatics*, 21(2), 191-199.
- van Dongen, B. F., de Medeiros, A. K. A., Verbeek, H. M. W., Weijters, A. J. M. M., & van der Aalst, W. M. P. (2005). The ProM Framework: A New Era in Process Mining Tool Support. *Proceedings of the ICATPN 2005 International Conference on Application and Theory of Petri Nets*. Berlin: Springer.
- Wallace, J., Scott, S., Stutz, T., Enns, T., & Inkpen, K. (2009). Investigating teamwork and taskwork in single- and multi-display groupware systems. *Personal and Ubiquitous Computing*, 13(8), 569-581.
- Wang, G., Zhang, X., Tang, S., Zheng, H., & Zhao, B. Y. (2016). Unsupervised Clickstream Clustering For User Behavior Analysis. *Proceedings of the CHI 2016 Conference on Human Factors in Computing Systems*. New York: ACM.
- Wang, Q., Jin, H., & Liu, Y. (2010). Collaboration analytics: mining work patterns from collaboration activities. *Proceedings of the CIKM 2010 International Conference on Information and Knowledge Management*. New York: ACM.
- Ware, C. (2012). *Information Visualization: Perception for Design*. Waltham, MA, USA: Elsevier.
- Wilson, S., Galliers, J., & Fone, J. (2006). Not all sharing is equal: the impact of a large display on small group collaborative work. *Proceedings of the CSCW 2006 International Conference on Computer Supported Cooperative Work*. New York: ACM.
- Winograd, T., & Flores, F. (Eds.). (1985). *Understanding computers and cognition*. Norwood, NJ, USA: Ablex Publishing Corp.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Burlington, MA, USA: Morgan Kaufmann Pub.

Zheng, S., Rosson, M. B., Shih, P. C., & Carroll, J. M. (2015). Designing MOOCs as interactive places for collaborative learning. *Proceedings of the L@S 2015 Conference on Learning@ Scale*. New York: ACM.

Authors' Mini-bios:

Roberto Martinez-Maldonado (Roberto.Martinez-Maldonado@uts.edu.au, <http://roberto.martinezmaldonado.net/>) is a Data Science and HCI Researcher with an interest in collaborative interfaces, machine learning and analytics systems; he is a Research Fellow in the Connected Intelligence Centre of the University of Technology Sydney.

Judy Kay (Judy.Kay@sydney.edu.au, <http://sydney.edu.au/engineering/it/~judy/>) is a HCI researcher with a particular interest in life-long, life-wide learning and novel interfaces; she is a Professor and the Director of the Human Adapted Interaction Group of The University of Sydney.

Simon Buckingham Shum (Simon.BuckinghamShum@uts.edu.au, <http://simon.buckinghamshum.net/>) is a Psychology, Ergonomics and HCI researcher with an interest in learning analytics, collective intelligence and sensemaking; he is a Professor and the Director of the Connected Intelligence Centre of the University of Technology Sydney.

Kalina Yacef (Kalina.Yacef@sydney.edu.au, <http://sydney.edu.au/engineering/people/kalina.yacef.php>) is a Data Science researcher with a particular interest in Artificial Intelligence in Education, Educational Data Mining and HCI; she is an Associate Professor at the Human Adapted Interaction Group of The University of Sydney.

Background. This article draws upon a seven year research program which has involved the analysis of six quite different group situations (cases) in collocated settings with more than 500 users and a variety of surface technologies, tasks, grouping structures and domains. Two of these cases were part of the Ph.D. thesis of the first author. The rest of the cases were conducted in 4 different institutions in collaboration with other authors. The article contribution includes the key insights and themes which have been identified and summarized in the set of principles and dilemmas that can inform design of future collocated collaboration analytics innovations based on this programme.

Acknowledgments. The authors would like to thank all of the participants in the studies presented in this article including: volunteers, students, teachers and facilitators, for the time and help given throughout. We would also thank collaborators that were involved in the research program, particularly to Dr. Ahmed Kharrufa and Dr. James Wallace for allowing very productive collaborations. We would like to also thank Dr. Maresa Edbauer, Prof. Yannis Dimitriadis, Prof. Peter Goodyear and Dr. Tamara Power for their valuable input and contributions to the program.

Funding. This research program has been partly funded by: Smart Services CRC, CONACYT (Consejo Nacional de Ciencia y Tecnologia), Fundación Pablo, The University of Sydney (International Office) and the Australian Research Council (Grant FL100100203). The studies were conducted under the following Human research Ethics protocols: 2012/2794 (The University of Sydney) and ETH16-0582 (University of Technology Sydney)

