Assessing Collaborative Processes via Instrumented Working Spaces
Yoav Bergner and Anne-Laure Fayard
January 2021

Rapid Community Reports
Workshop Outcomes
Assessing Collaborative Processes via Instrumented Working Spaces

What can person- and room-mounted sensors reveal about collaborative processes in project-based learning environments?

Authors

Yoav Bergner, New York University, yoav.bergner@nyu.edu
Anne-Laure Fayard, New York University, alfayard@nyu.edu

Abstract

We report outcomes from the Working Group on Instrumented Learning Spaces (WGILS), a two-day convening of experts from the fields of learning sciences, human computer interaction, organizational studies, and learning analytics. The working group focused in particular on understanding and supporting collaborative project-based work in multifunction environments such as makerspaces. Considering the current state of consumer-level technology, we explored what person- and room-mounted sensors can reveal about collaborative processes in project-based learning spaces. Going beyond the theoretical, we implemented a fast-and-cheap data collection using ourselves as research subjects in a collaborative task. Viewing instrumented learning spaces are socio-techno-cognitive systems, this report is aimed at researchers who wish to advance the potential of such systems to illuminate complex social learning phenomena.

Keywords

Multimodal analytics, learning analytics, collaborative learning, project-based learning, assessment, evidence-centered design

Introduction

The Working Group on Instrumented Learning Spaces (WGILS) was a two-day convening of experts from the fields of learning sciences, human computer interaction, organizational studies, and learning analytics, among others, that took place at New York University’s Tandon Makerspace in February 2019 following several rounds of remote planning.

The purpose of this workshop was threefold:

1. to establish relevant types of claims, evidence, instrumentation, and tasks for understanding collaborative processes in instrumented spaces;
2. to identify key challenges, opportunities, and emergent issues with the use of sensing technologies;
3. to curate a multi-modal dataset that other researchers can draw on to further research on assessing collaboration in unstructured and creative learning environments.

Educators and employers know that interpersonal and collaborative skills are critical to success in the knowledge economy. Assessment of collaborative processes is notoriously hard to do without disrupting the natural flow of interactions. Emerging from research labs, however, a variety of multi-modal sensors (e.g., audio, video, spatial, biometric) have been used to capture collaborative interactions as they happen. These sensors have the potential to transform what kind of skills get measured. Some of these technologies are now affordable enough to bring out of the lab and into the learning environments themselves.

The purposes of this workshop was threefold:

difficult to scale up, and quantitative studies, which can be both disruptive—e.g., having people respond to questionnaires at regular intervals—and overly narrowly in their focus. Developments in technology and combined expertise from a range of socio-technical research fields continue to lower the barriers to studying such interactions in authentic learning and work environments.

We chose a makerspace as a model environment for our study. Makerspaces are sites of learning and work, often collaborative, involving a combination of design and prototyping activities. The open-ended, collaborative activities that take place in these venues pose unique challenges in comparison to more structured learning and performance environments. Tasks in makerspaces are often ill-defined, with emergent activities and evaluative criteria that are unclear or non-existing. That said, instrumenting spaces in service of understanding and supporting collaborative work goes beyond makerspaces, and our exploration was intended to inform those other environments as well.

Regarding collaboration and group dynamics for instrumented spaces, we considered in turn the facets of interest, the nature of informative evidence, and the context- and stakeholder-dependence of both of these.
Understanding collaborative processes in an instrumented learning space calls on convergent knowledge and methodology from learning sciences, psychology, computer science, and design fields. We invited researchers working in focus areas of collaborative learning, enactive learning spaces (project-based learning, makerspaces, FabLabs, etc.), human dynamics, human-computer interaction, organizational studies, and multi-modal learning analytics. Overall, seventeen scientists were involved in a face-to-face workshop.

Attendees participated in a live simulation of collaborative work in an instrumented space. Mode-switching—the researchers becoming research subjects—was a distinctive feature of our working group.

Before the face-to-face workshop, we organized remotely into four thematic subgroups dedicated to (a) explication of constructs and theory, to focus inferences, (b) structure and content of data to support inference, (c) task design, to elicit meaningful behaviors, and (d) instrumentation itself, to collect the data. From the outset, we set out to narrow the landscape of “things one could observe happening in learning spaces.” We imagined that the purpose of instrumenting learning spaces was ultimately to be able to understand and affect behaviors, attitudes, and other learning and performance outcomes for the benefit of individuals and groups. Working within the conceptual assessment framework of evidence-centered design (Mislevy, Steinberg, & Almond, 2003) as a structural outline, we set out first to answer the question, “evidence about what?” What claims would we want to be able to make about work, especially group work, in project-oriented learning spaces? These claims would in turn help to guide what kind of evidence would be necessary to warrant such claims and what kinds of tasks would be suitable to elicit such evidence.

Recognizing also that the experiment we would carry out as a working group in person was not quite an authentic experiment—that we the researchers were not authentic subjects—we distinguished between more and less artificial inferences. For example, inferences about a professor learning how to use a makerspace tool would be inherently less interesting than a real student doing the same. But inferences about a group of professors designing an object together would still involve judgments relevant to
collaborative processes in a student population. For example, questions like these would still apply: How were decisions made? How was activity distributed among participants? Did individuals assume particular roles in the collaboration?

Time constraints of the in-person data collection necessarily affected potential inferences. Given just a few hours of activity, it would be hard to justify strong claims about learning gains or about the quality of prototypes. The logistical and practical constraints helped us turn away from outcomes, i.e., evaluation of products, and to focus rather on collaborative processes. The group drew from several existing frameworks for collaborative problem solving (e.g., Meier, Spada & Rummel, 2007; Hesse et al., 2015; Graesser et al., 2018) and narrowed the focus to four key aspects: negotiation, coordination, role-taking, turn-taking. To that end, we also designed for some between-group differences, described below.

In February 2019, seventeen researchers from around the country convened in person at the NYU Tandon Makerspace in Brooklyn, NY. In contrast to the designed and fixed setups of many research labs, we set out to build a kind of pop-up or temporary installation. We set up the instrumentation in a real learning space in one day. Selection of instruments were guided by specific purpose, which had to be narrowed to a reasonable scope for a short working group, but also by cost targets for replicability in real-world classrooms (in dollar terms, thousands but not tens of thousands). We thus imagined a range of technologies: from audio and video—augmented by algorithms for automated processing—to wearable position and proximity detectors.

The gathering consisted of four phases: (1) orientation, (2) instrumented prototyping activity, (3) data assembly, and (4) postmortem. After a brief welcome, participants were individually equipped with voice recorders and position trackers, assigned to their groups (four groups of four) and presented with their challenge for the day: to design and prototype a cup and saucer using the materials of their choice.

To introduce variation in the collaboration constructs of interest, we used a two-by-two design that systematically varied degree of prescriptive structure and role assignment of each group: one group was provided with a detailed activity script that prescribed their workflow and role cards for each member (e.g. questioner, time keeper), one group received only role cards, one group was assigned only to the activity script, and the fourth group was not given any kind of assignment.

Once the groups were assigned and all of the sensors were on, we more or less forgot about them and became experimental subjects for a day. Participants had approximately six hours to complete the challenge, including a break for lunch. During the prototyping, groups chose to experiment with different machines provided in the space.
such as 3D printers, CNC milling machines, and laser cutters. One group took a radically different approach and made their prototype entirely out of food!

At the end of the day groups presented their final product and evaluated one another using the following criteria: functionality, creativity, aesthetics, affordability, and environmental friendliness. (These criteria were made known ahead of time.) The next day, we gathered to reflect and debrief, collect notes, and inspect the data to see which instruments worked as planned. We also collectively constructed a multi-modal data archive, to be released for public use.

Figure 1: Group work during the prototyping activity in the makerspace. Participants are wearing vests equipped with voice recorders and position-tracking transmitters. A table-mounted directional microphone array is obscured by the craft materials. A Kinect sensor by the window is also tracking body positions. Photo credit: Ofer Chen
Key issues

As we reflected, four different aspects of research on instrumented learning spaces emerged: technological challenges, assessment limitations, learning design tensions, and the difficulty of going from sensing to meaningful feedback.

Multimodal sensing of collaboration is technically challenging

Our “on-the-cheap” approach was only partially successful. For example, while the Pozyx (pozyx.io) position tracking sensors (cost $150/person) provided a relatively accurate map of each participant’s location throughout the exercise, many of the personal voice recorders ($18/person) either failed to record entirely or turned off at some point during the activity. One of the table-mounted audio arrays ($100) also failed. Reconstruction of who said what to whom and where, an essential task of multi-modal analysis, thus became much harder. Although state-of-the-art technologies for voice and position tracking are becoming available, they are still beyond reach for most classrooms with the budget in our target. $150 personal voice recorders would have surely been more robust than our $18 models, but too expensive in a classroom of 20-30 students. In addition to the financial limitations, significant technical skill is required to parse the synchronized streams of voice, position, and gesture data that we collected using the OpenSSI framework (hcm-lab.de/projects/ssi/). This hurdle is likely too high for a classroom or makerspace teacher, and there are currently no easy tools or interfaces to help teachers and students make meaning from these data. Thinking about a user-centered solution, including front end design, for instrumented learning spaces is an important direction for future work.

Participating in a simulation revealed assessment limitations

Being “part of the experiment” made the signal underdetermination problem salient. Sometimes a shrug means “I don’t know” and sometimes it means “I don’t care.” Even philosophers of science could not escape the conclusion that evidence always “under-determines” a scientific theory; that there can always be competing theories that adequately explain a set of observations. As both researchers and research subjects, participants were able to switch modes between knowing their own intentions and feelings and wondering whether the instrumentation signals would be able to uniquely identify those experiences to another observer. In the workshop debrief, participants wondered, for example, would the data differentiate between a participant going over to look at a machine and the participant actually using the machine? Similarly, did the data make it possible to reconstruct the frustration teams encountered in the design process and while using the equipment? Could the data capture
other kinds of emotions, such as surprise or enjoyment, engagement and disengagement? During a timed prototyping task, emotional experiences—even among our pseudo-research subjects—ran the gamut. Some rejected the assigned roles, others questioned the task value (much like real students would), while others wrestled with balancing metacognition with a need to “get things done.” Such reflections shed light on broader questions regarding what kinds of information the instruments in their existing form can capture and what is most useful for students and teachers to know. Combining a variety of multi-modal sources, such as video and position tracking with audio diaries, can help ensure that the data collected can provide answers to the right questions. An instrumented space holds promise for revealing more of what is hard to measure, but we are still struggling to know what students are thinking.

Instrumentation design vs. learning design

The technical challenges of using multi-modal sensing technologies can easily overwhelm the design efforts of instrumented learning spaces, pushing aside important questions of “purposeful evidence.” The constraints of what we can measure end up taking precedence over what we want to measure in order to draw desired inferences about the learning experience. For example, our working group elected to focus on collaborative processes rather than individual learning trajectories due in part to limitations on the spatial resolution of technology to capture the nuances of individual movements. Modeling individual hand positions is just harder than extracting the relative positions and rotation (turn and leaning angles) of members in a group.

The instrumentation was also in competition with the learning design in a more practical way: the instruments themselves proved to be somewhat burdensome. Each participant was asked to wear a number to facilitate the computer vision analysis of the videos, a voice recorder, and a position tracker. The position tracker had to be oriented in a specific direction in order to sync with the anchors placed around the room. Participants were given safety-vests with multiple pockets to facilitate the process of attaching the number and carrying the different sensors; however, these proved bulky and required frequent attention, such as making sure the recorder was on or that the sensor was positioned correctly. The trade-off between accuracy and disruptiveness in instrumented learning spaces is an important consideration for future research.

Meaningful feedback is its own challenge

Real-time feedback on collaboration in an instrumented learning space is its own computer-human interaction design problem. We have alluded to some of the challenges with interpretation of multimodal data. With time, we believe, both observational and experimental studies of instrumented learning spaces are likely to yield real insights. Ideally these will even go beyond correlational claims (e.g., students/groups who do X are also
more likely to do Y) into causal claims (e.g., certain types of scaffolds lead to more harmonious and productive collaboration). Even so, converting such insights into real-time automated feedback will remain a design challenge. What kind of automated feedback is both enjoyable and effective?

Suppose that evidence from a group’s behavior suggests that its members are struggling to negotiate roles or spinning wheels about a particular technique. It might be a good time, for example, to suggest a temporary role assignment, or to point out that expertise is available outside of the group. But a system that is capable of making intermittent suggestions can also be annoying and intrusive. As one workshop participant recalled, referring to a widely reviled automated assistant in Microsoft Word from 1997-2007, “no one wants ‘clippy’ for collaboration.” Nevertheless, providing assistance with knowledge diffusion for prototyping techniques was seen as critical and worth the design effort to do it well. How and when to intervene in a collaborative task is a research question worth exploring.

Recommendations for future work

The working group uncovered opportunities and challenges associated with the use of affordable sensing technologies to assess collaboration and collaborative learning processes. As discussed in the insights, although our ability to set up a temporary simulation space “in the wild” in less than a day and with a small budget is promising, the challenges associated with the malfunctioning of some equipment, as well as the expertise required to extract and make sense of the different data streams, offer opportunities for researchers to further explore and refine how these technologies can be made more accessible to teachers and students.

In addition, the group’s reflections on the uses and limitations of the various sources of data collected raise important questions that researchers can explore both using our curated dataset and by collecting more data in future experiments. Key reflections to be explored further include:

1. **Extent of Capture:** How much of what happened throughout the day was captured? How much/what needs to be captured? Can the data capture a group’s “pivotal moments”? How might the data tell us the stories of different users— their level of expertise, how they engaged and collaborated, what they learned, and how they might have evolved throughout the process?

2. **Learning Goals:** How might the data look different if the goal had been different (e.g., to learn something, to collaborate meaningfully)? How might
what could be considered “off-task” behavior illustrate relationships being built or development of “soft-skills”?

“Key questions include the extent of capture and the match between learning goals and measurements.”

**Recommendations**

We propose three short-term recommendations (the next two years) to the field:

S1: With respect to the overall challenge of measuring collaborative group work, instrumentation research should help the field move from single instruments to well-chosen combinations. This shift will keep the focus on the underlying constructs and away from the if-you-have-a-hammer fallacy. Design guidance is needed to understand the tradeoffs between accuracy, affordability, and intrusiveness for different educational use-cases.

S2: The field should bring together scholars in learning sciences, organizational behavior, and learning analytics to help unpack the interaction between contextual factors, instrumentation technologies, and inferences about learning.

S3: Researchers should pursue scholarship that identifies the effect of context on the measurable properties of collaboration. Outcome measures and inferences about collaboration constructs are deeply entangled with the learning environment, the learners themselves, the task design, and the technologies used for observation.

Our additional medium-term recommendations (3-5 years) are:

M1: The field should develop teacher-facing interfaces to help educators turn multi-modal data into useful feedback in real-time.

M2: Researchers should explore richer socio-techno-cognitive models to better understand how technology, environment, and cognition interact in learning spaces.

Our longer term recommendations (5+ years) are:

L1: What we measure affects what we do in an instrumented learning space. Researchers and educators should make sure they are setting up to measure what should count and harnessing new technologies to assess hard-to-measure constructs.

L2: The field should move towards understanding instrumented learning spaces as cyber-physical “expert” systems with the potential to augment the formative and summative assessment capabilities of human teachers and facilitators.

For further information about this workshop and our discussion, please read the full report and slides.
References


Resources

Edutopia: Assessing Learning in Maker Education
https://www.edutopia.org/article/assessing-learning-maker-education

EdSurge: How Should We Measure the Impact of Makerspaces?

LearnAlITech: Introduction to Multimodal Learning Analytics
https://learnaitech.com/introduction-to-multimodal-learning-analytics/
Acknowledgements

The authors would like to acknowledge Victoria Bill and the NYU Tandon Makerspace staff for their flexibility and technical support and Patti Schank, Judi Fusco, and Digital Promise for financial and organizational support.

Rapid Community Reports are supported by NSF grant 2021159. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix

Workshop Participants:

Win Burleson (University of Arizona), Gustavo Almeida (University of Arizona), Xavier Ochoa (NYU Steinhardt), Kayla DesPortes (NYU Steinhardt), Mike Tissenbaum (UIC), Veronica Newhardt (UC Irvine), Lin Lin (University of North Texas), Matthew Berland (UW Madison), Marcelo Worsley (Northwestern University), Tom Moher (UIC), Bertrand Schneider (Harvard University), Yoon-Jeon Kim (MIT), Caitlin Martin (Digital Youth Network), Peter Wardrip (UW Madison), Ingrid Erickson (Syracuse University), Dani Herro (Clemson University), Paulo Blikstein (Teacher's College, Columbia University), Anne-Laure Fayard (NYU Tandon), Yoav Bergner (NYU Steinhardt).