Working Group on Instrumented Learning Spaces

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

Introduction

This paper reports outcomes from the Working Group in Instrumented Learning Space (WGILS), 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 remotely-coordinated planning. The goal of the workshop was to explore the opportunities and boundaries of person- and room-mounted sensors for studying collaborative processes in project-based learning environments. The original motivating scenarios involved educational makerspaces, though the insights gleaned can be applied to a variety of unstructured, collaborative settings.

The report begins with a description of the working group motivation, design, and planning process. After describing the two-day live simulation, we synthesize key insights and reflections from the experience, noting the continued challenge of balancing fidelity of sensing capture with affordability and accessibility of the technology and risks of competition between instrumentation and the learning design. We conclude with immediate, near-term, and longer-term recommendations for the field, highlighting opportunities for more interdisciplinary collaboration and supporting the further development of instrumentation that enhances the learning design and student experience, and advocating for keen attention to what we are choosing to measure and why, remembering the adage that what we measure affects what we do.

Some of the key takeaways are as follows:

- Multimodal sensing of design activities is still technologically very challenging. *Thinking about a user-centered solution, including front end design for real classroom teachers in instrumented learning spaces, is an important direction for future work.*
- Automated analysis of collaborative processes is only part of the solution; converting inferences into feedback brings its own set of challenges. *How and when to intervene in a collaborative task is a research question worth exploring.*
- Being "part of the experiment" made the signal underdetermination problem salient for the working group members. *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 can be in competition with learning design. *The trade-off between accuracy and disruptiveness in instrumented learning spaces is an important consideration for future research.*
- Instrumented learning spaces are socio-techno-cognitive systems. *Communication and collaboration opportunities between organizationists, educationists (of both formal and informal contexts), and technologists can pave the way for context-rich models of learning.*
- Instrumented learning spaces are cyber-physical systems with the potential to change what gets counted in educational spaces. *The groundwork ought to be explored for expert systems used by teachers of the future to connect multi-modal data signals as evidence for constructs of interest.*

Motivation or Purpose

The purpose of this workshop was threefold:

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

A key objective of the NSF Cyberlearning 2017 gathering was to encourage the formation of teams to tackle future challenges and continue the conversation about innovative genres for learning at the human-technology frontier. In this spirit, we proposed to assemble a Working Group on Instrumented Learning Spaces (WGILS) using multi-modal data streams. We had in mind some specific use cases, namely project-based learning environments, such as makerspaces, where learners work, often collaboratively, through a combination of design and prototyping activities. These environments, and in particular the open-ended, collaborative activities that take place in them, pose unique challenges in comparison to more structured learning and performance environments. Tasks are often ill-defined, activities emergent, and criteria to evaluate the work unclear or non-existing.

Educators and employers know that interpersonal and collaborative skills are critical to success in the knowledge economy, but it is hard to improve training without quality evidence. Collaborative processes that take place in the real (as opposed to virtual) world are notoriously difficult to assess without disrupting the natural flow of interactions. However, beginning with work in research labs, a variety of multi-modal sensors (e.g., audio, video, spatial, biometric) have been used to capture collaborative interactions, with the potential to transform what is measurable. Some of these technologies are now affordable enough to bring out of the lab and into the learning environments themselves.

The use of multi-modal data analysis aims to fill a gap between ethnographic studies, which preserve naturalistic features but are difficult to scale up, and quantitative studies, which can be both disruptive—interrupting natural flows to have people respond to questionnaires at regular intervals—and narrowly focused. While still emerging, developments in sensing 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. The working group was thus an opportunity to bring such a collection of interdisciplinary experts together to explore the opportunities and boundaries of using audio, video, and other sensing technologies to better understand collaboration and group dynamics. Specifically, the group sought to answer questions to questions such as:

- What are the facets of interest in terms of collaboration and group dynamics for instrumented learning spaces?
- What kinds of evidence are most valuable, and what are informative data?
- How do these answers depend on the uses of evidence and on the type of stakeholders?

Understanding collaborative processes in an instrumented learning space calls on convergent knowledge and methodology from learning sciences, from psychology, from computer science, and from 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. The working group engaged in a series of planning meetings, conducted via teleconference, with one in-person gathering in New York City. We explored affordances, constraints, and data infrastructure requirements for high-fidelity capture of performance, learning, and collaboration in active spaces. Moreover, we explored these questions not only through communal discussion and planning (ideation), but by physically instrumenting and collecting data from a prototyping activity which we carried out ourselves at the NYU Tandon Makerspace. The makerspace activity simulated in-vivo data collection to explore the processes and methods of evidence identification which can be transferred to other sites. We collectively constructed a multi-modal data archive, to be released for public use, and reflected on the experience and the process. This mode-switching—the researchers becoming research subjects—was a distinctive feature of our working group.

Starting Points & Process

Planning remotely, by sub-committee

As the working group was fairly large, we began our planning work by self-organizing 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. Each subgroup had a designated coordinator, or subgroup leader, and these leaders were the principal conduits of communication between groups.

From the outset of our collective effort, 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?" In other words, 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 an authentic experiment—that we the researchers were not authentic subjects—we nevertheless 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? Did individuals assume particular roles in the collaboration?

A further constraint of the in-person data collection that affected potential inferences was limited time. 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, from 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.

The prototyping day and debrief

In February 2019, seventeen researchers from around the country convened in person at the NYU Tandon Makerspace, a state-of-the art, 10,000 square foot space 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 the following 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. In addition, 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 (eg. 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. The next day, we gathered to reflect and debrief, collect notes, and inspect the data to see which instruments worked as planned. Spoiler: not all of them. This post-joint-activity reflection time was, perhaps unsurprisingly, one of the most illuminating phases of the working group. Below we synthesize some of the most common insights that emerged in the group postmortem conversations:

Insights, Issues, & New Ideas

Multimodal sensing of collaborative activities is still technologically very 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.*

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 "underdetermines" 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 can be in competition with learning design. The technical challenges of using multi-modal sensing technologies can easily overwhelm the design efforts of instrumented learning spaces, sometimes 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 resolution of the available technology to capture the nuances of individual movements. Modeling individual hand positions is still much 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.*

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. It is simply hard to design computer-human interactions for groups that people like and that enhance collaboration.

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.*

Directions and Recommendations

Directions for future research. 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, while 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 through future simulations and the collection of new data. These questions include:

- **Extent of Capture**: How much of what happened throughout the day was captured? How much data went unnoticed and uncaptured? 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?
- **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"?

Recommendations to the field

Immediate: Instrumented learning spaces and multi-modal learning analytics are still in their infancy. We propose three immediate-term recommendations to the field:

- **Recommendation 1**: Research should prioritize and publish findings on the attributes of individual instruments with respect to overall challenge of measuring collaborative group work, to help the field move from individual instruments to well-chosen combinations. Design guidance is needed to understand the tradeoffs between accuracy, affordability, and intrusiveness for different educational use-cases.

- **Recommendation 2**: The field should expand to bring in scholars in learning science, organizational behavior, and learning analytics to help unpack the interaction between contextual factors, instrumentation technologies, and inferences about learning.
- **Recommendation 3**: Researchers should pursue scholarship that identifies the contributions of context to the measurable properties of collaboration. Outcome measures and inferences about collaboration constructs are deeply entangled with the learning environment, the task design, and the technologies used for observation.

Near-term (3-5 years):

- **Recommendation 1:** The field should develop teacher-facing interfaces to help educators turn multi-modal data into useful feedback in real-time.
- **Recommendation 2:** Researchers should explore richer socio-techno-cognitive models to better understand how technology, environment, and cognition interact in learning spaces.

Longer term:

- **Recommendation 1:** Researchers and educators should make sure they are setting up to count what should count and using technologies to assess hard-to-measure constructs. What we measure affects what we do (in an instrumented learning space).
- **Recommendation 2:** 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.

References

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, 19(2), 59-92.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In *Assessment and teaching of 21st century skills* (pp. 37-56). Springer, Dordrecht.

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.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence‐centered design. *ETS Research Report Series*, *2003*(1), i-29.

Appendix A: Meeting Agenda

<table>
<tr><td>8:45 - 9:30</td><td colspan="2">Welcome & Opening Remarks
Overview and Safety Training at Tandon Makerspace

Form Teams (with 4 members in each team)
Team one: scripts, roles
Team two: scripts, no roles
Team three: no scripts, roles
Team four: no scripts, no roles
[To form teams of the day, we will first have the current working group members stay together. We will ask each member in the groups to draw a number between 1-4. The people drawing the same numbers will be members of the new teams of the day]</td></tr>
<tr><td>9:30 - 10:00</td><td colspan="2">Marshmallow Ice Breaker</td></tr>
<tr><td>10:00 - 10:10</td><td colspan="2">Introduce the Task: Your goal is to make the best cup with saucer you can.</td></tr>
<tr><td>10:10 – 10:20</td><td colspan="2">Break</td></tr>
<tr><td>10:20 – 10:50</td><td>Brainstorming</td><td>Work on Task</td></tr>
<tr><td>10:50 - 11:20</td><td>Sketching</td><td>Work on Task</td></tr>
<tr><td>11:20 - 12:00</td><td>Prototyping 1</td><td>Work on Task</td></tr>
<tr><td>12:00 - 12:30</td><td>Testing and debugging 1</td><td>Work on Task</td></tr>
<tr><td>12:30 - 13:30</td><td colspan="2">Lunch</td></tr>
<tr><td>13:30 - 14:00</td><td>Feedback from the other scripted group</td><td>Feedback from the other non-scripted group</td></tr>
<tr><td>14:00 - 15:00</td><td>Prototyping 2</td><td>Work on Task</td></tr>
</table>

<table>
<tr><td>15:00 - 15:30</td><td>Testing and debugging 2</td><td>Work on Task</td></tr>
<tr><td>15:30 - 16:00</td><td colspan="2">Refinement and Preparations for Presentations</td></tr>
<tr><td>16:00 - 16:30</td><td colspan="2">Presentations
All the participants will score the product(s) of each team based on the following criteria (rate from 1 to 5, 1 being lowest and 5 being the highest score):
● Functionality
● Creativity
● Aesthetics
● Affordability
● Environmental friendliness</td></tr>
<tr><td>16:30 - 17:00</td><td colspan="2">Closing Discussion</td></tr>
</table>

Appendix B: List of Participants

Yoav Bergner
Anne-Laure Fayard
Win Burleson
Gustavo Almeida
Xavier Ochoa
Kayla DesPortes
Mike Tissenbaum
Veronica Newhardt
Lin Lin
Matthew Berland
Marcelo Worsley
Tom Moher
Bertrand Schneider
Yoon-Jeon Kim
Caitlin Martin
Peter Wardrip
Ingrid Erickson
Dani Herro
Paulo Blikstein