Workshop Report:
Building Interdisciplinary Capacity for Understanding and Supporting Computer Supported Collaborative Learning
Cindy E. Hmelo-Silver, Indiana University, chmelosi@indiana.edu

**Motivation and Purpose**
The workshop entitled “Building Interdisciplinary Capacity for Understanding and Supporting Computer Supported Collaborative Learning” began with a proposal for an Early-concept Grant for Exploratory Research (EAGER) call from the National Science Foundation (NSF) in 2015. The genesis of this workshop began as part of a January 2015 Cyberlearning PI meeting in Arlington VA which the PI, Dr. Cindy Hmelo-Silver was asked by the meeting organizer to co-lead a short workshop on “Learning Analytics: How can we bring adaptive and collaborative learning closer.” We saw the need to take an exploratory approach to look for actionable indicators in work on learning analytics and adaptive support for collaborative learning. Learning analytics work showed promise for informing collaborative learning but many of the indicators being used were shallow measures of participation and engagement. The next logical step appeared to be to delve deeper and extract actionable indicators from research on collaboration, to determine whether or not supported by technology, they could be used to develop new technologies that would build models of collaboration that would be amenable to learning analytics, and ultimately lead to better adaptive support for collaborative learning, whether in stand-alone systems or to help teachers on a just-in-time basis.

**Starting Points & Process**
Within the collaborative learning literature itself, there are several target models of productive collaboration (see Hmelo-Silver & Chinn, 2016). Examples of such models include productive disciplinary engagement (Engle & Conant, 2002), Knowledge Building (Chan, 2013), and group cognition (Stahl, 2006). There are also multiple designs for development of collaborative learning environments such as problem-based learning, scripted collaboration, and group investigation (Hmelo-Silver & Chinn, 2016). One key challenge for this workshop was to understand what aspects of collaboration may be general across many different types of settings, what aspects are particular as a function of age, context, discipline, collaboration modality and to better understand these contingent factors as part of constructing a model of collaborative learning. This was particularly challenging because although learning events are happening all the time, learning is a highly contextual phenomenon.

Much research on learning analytics (LA) has focused on using large datasets to discover patterns in activity and resource use and then connect those kinds of data to learning outcomes (Baker & Siemens, 2014). A goal of learning analytics is to empower learners and instructors in ways that are informative and actionable. Much of the focus in learning analytics has been on individuals and often used shallow measures such as clickstream activity, time spent on the system, specific resource use, quiz data, social network data, etc. (Verbert, Duval, Klerx, Govaerts, & Santos, 2013). These data are easy to collect with current online systems, tend to focus on individuals, and are generally used in online settings (c.f. recent issue of *Journal of the Learning Sciences* edited by Martin & Sherin, 2013 and Rosé et al, 2008 for some exceptions). Thus, in the workshop, it was decided to focus on the challenge of using the discourse from collaborative learning as the data to be focused on as a target for learning analytics and adaptive scaffolding. One advantage of LA is that it can make data useful for supporting learning in real time by providing timely data to educators and learners. To accomplish this, we would need to have models of high-quality collaborative learning, to know how what is manifested in discourse, and to know what needs to be pulled out of the discourse to look at more and less productive patterns of collaboration in real time so that that is can be actionable.

**Insights, Issues, & New Ideas**
This report focuses on the goals of two workshops intended to build community and capacity around bringing learning analytics and collaborative learning closer together. The workshop brought together experts from collaborative learning who shared relevant datasets with people in analytics who had technical expertise that might transform that data into useful information. Our expectation that we could engage in productive discussions across disciplines to deepen our understanding and build on one another’s works.
One key goal of this workshop was to articulate the grand challenges for the fields of learning sciences, learning analytics, and information technology for addressing automated measurement and analysis of collaboration. Through an open call for participation, researchers from multiple disciplines such as learning sciences, computer and information sciences, and education were asked to identify specific collaborative learning data sets and analytic tools such that this would be an opportunity for workshopings ideas and developing potential collaborations. Specifically, the group would attempt to answer questions about the kinds of data collaboration needs to support, the kinds of data that can be usefully extracted and analyzed from these collaborations, the needs of adaptive support for learners and aids for teachers, and what kinds of tools are mediating collaboration and what this community would ideally like to see built in the future. This group would work together to create research agendas that would help tackle the needs of the growing field of computer supported collaborative learning (CSCL).

**Workshop 1**

In the first workshop, hosted at Indiana University in Bloomington, Indiana, social and behavioral scientists took the lead and set the scene, talking about analytic approaches to current collaborative data sets and collaborative models. They identified limitations and challenges in these manual analysis techniques. Additionally, they worked together to set goals and expectations for coding schemes they envisioned to really expand the collaborative possibilities in the future. To ground the discussion in state-of-the-art data collection techniques, four of the attendees brought data sets from previous research to share with other attendees for examination purposes. Included in this were the data themselves, descriptions and, where applicable, rubrics that were used in analysis. Of course, all due process was followed to cover any privacy concerns with the data. These data-sharers gave short presentations that introduced participants to different tools and techniques for collecting and analyzing data related to collaboration.

For the first workshop, we collected coding schemes from a group of researchers with diverse research backgrounds then re-coded them into emergent categories and generated an empirically grounded framework that reflects collaborative learning models to inform CSCL researchers and practitioners. By making this framework available to the CSCL community, we aimed to illuminate the field’s shared vision of the key dimensions of collaborative learning. This framework could also be a step toward providing a common language to promote sharing of analytic tools and enable synthesis across individual studies.

First in small groups, then later as a larger whole, participants examined the coding schemes that were contributed to examine the commonalities and distinctions. The coding schemes were designed for both face-to-face settings and online synchronous data. Indicators ranged from purely cognitive, such as references to knowledge, interpretation, metacognition, or reasoning to those indicators that were purely social such as facilitation, leadership, and interaction patterns. Monitoring and regulation codes spanned both the social and cognitive realm as collaboration places demands for regulation that involve monitoring both one's own cognition as well as the shared regulation within the group. There were also some researchers who used self-report data to look at collaboration.

The workshop discussions noted some concerns about what was missing and harder to understand than the line-by-line coding schemes indicated. First, there were some derivative analyses that would need to be done, but that could possibly be accomplished with coded data such as looking at equitable participation (or lack thereof). Another issue that participants raised was how we would determine progress in a group, whether it was in terms of advancement of ideas or in the social dynamics. Indicators of uptake of ideas, development of common ground, and group cohesion would demonstrate an aspect of this progress. One suggestion for an indicator of such progress would be awareness of disagreements and confusion and awareness of others as contributors. This also suggests a need for a way to capture the temporal dimensions of collaboration. Some of the groups raised the importance of needing to model the context in which collaboration was being examined. Another group also noted the importance of being able to detect shifts in engagement, such as from on-task to off-task and vice versa, changing roles, and change in the socioemotional climate. These seem particularly important in considering how such factors can impede or promote knowledge advancement. This is consistent with the concern by the group studying an online dataset, which also raised the issue of how to identify indicators of higher order thinking and reasoning, deep engagement with content, and collaboration dynamics.

Given the goal of the workshop, one question that was raised was whether there needed to be a way for identifying "scaffolding opportunities" in dialog if the goal was eventually to be able to provide adaptive support. Another question was raised as to what researchers could tell about collaboration based on artifacts that a group might produce. Additional issues related to coding and analytics included how collaboration data would be captured and processed (especially when part of the collaboration is face-to-face), segmentation, and grain
size. The issue of granularity of data will be an important consideration for being able to provide adaptive support as compared with what is needed for research purposes.

Prior to the second workshop, participants prepared in two ways. First, annotated anonymized data sets were made available for participants to review. Second, participants with more computational expertise shared relevant papers for more behaviorally oriented participants to review. At the second workshop - many months later - computer scientists examined the best methods for achieving the desired goals and presented ideas for automating the indicators of productive collaboration and how these might be actionable for promoting adaptive scaffolding. Together, both groups used these new schemes to inform improved collaborative models that will be easier to facilitate in a wide range of settings to a broad group of users.

Workshop 2
In the second workshop, hosted at Carnegie Mellon University in Pittsburgh, Pennsylvania, attendees focused on narrowing in on specifics from the first meeting, as most attendees had attended both workshops. The participants began on the first day by examining previously selected papers on a variety of topics. Day Two saw the group perform a “data jam,” in which they discussed particular data sets. Participants then broke into groups to tackle three different themes. These groups were self-selected based on interest.

Group 1 - Visualization techniques for CSCL Data
This group started with basic questions and assumptions about why visualization is important when examining data. It was decided that data in CSCL necessitates visualization and is key as a first step in analysis due to its natural complexity. This allows researchers to find the right grain size, timing, and ways to aggregate and re-code. None of these are possible without first visualizing the data.

This group then examined a specific data set and the coder’s first pass at visualization of that data. The first visualizations they examined (utterances by students in this case) required a great deal of restructuring and analysis by this group actually led to the creation of a new variable. Next, they looked at a scatter plot analyzing the length of utterance. This visualization allowed them to compare different groups or days to provide for quick comparisons. Next, they looked at a comparison of groups where colors represented duration of talk by each student. These were color coded by quantity of talk. This visualization allowed them to see who was talking over each other and allowed seeing how collaboration quality changed over time. This group covered topics such as:

- Student dependency on technology.
- Is technology necessary for good collaboration?
- Adaptive support versus other kinds of support technology such as scripts and augmentation.
- Functions of embedded assessments and the need for built-in validity evidence

The process from after transcription to restructuring/cleaning was discussed, and suggested steps were suggested. Finally, the group concluded that visualization is used to inform analysis rather than as the analysis itself. Visualization should be used as an iterative process throughout the analysis. Additionally, visualizations can give a feel for the data-providing nuance that some statistical analyses may miss.

Group 2 - Coding schemes for different levels of analysis
This group began with a data set and chose two teams from that set to analyze in particular. During the initial look, participants noted that when the data is captured in a flat Excel document, it is hard to capture some nuance of the conversation. The question arose that if this the same sort of format that students see when they type online, then it seems that students are typing simultaneously and might not have opportunity to read the responses of their peers prior to typing their own thoughts. This is a potential limitation of this data format.

This group focused on visions for adaptive support for CSCL and concluded that lines between adaptive support for collaborative learning and learning analytics may be becoming more blurred. Whom does the support go to: students? teachers? both? A different conception may involve supporting different kinds of adaptive support for collaboration of different kinds/groups of learners. Making sure that common ground and shared frameworks of understanding and regulation are occurring in the development of adaptive support for collaboration is key. Moving forward, then, may require the development of “big ideas” for adaptively supporting collaboration.

In discussing coding schemes, it was clear that what we code is indicative of what we think is important. This can help shape and steer analysis. Discussants pondered what some actionable items to automate coding schemes are and what affordances and constraints there are. It was decided that it might be impossible to automate the open-ended, inductive sorts of analysis, but that there are categories of analysis that we can
automate and agree upon to operationalize analysis. This would vary depending on research questions in each case and we would need to know how each level of our coding relates to other schemes and questions.

Another key question regarding timing or immediacy, that provide analysis months or years later, or can we do ‘quick skims’ that will inform practice tomorrow? For adaptive systems, we should be able to provide feedback or iteration in real time. For other forms of data (e.g. Wiki responses, blog posts, etc.), how do we make actionable data that can be used right away? In addition, importantly, when are researchers, TAs, etc. in the way of the genuine practice of students? How do we make this sort of research sustainable? Can teachers easily access the data analysis? A list of possible indicators was constructed:

• What questions are students asking? What resources are referenced? Can self-reported data from students inform teaching practices?
• Disagreements & similarities as indicators of phenomena. When and where are these things important?

**Group 3 - Text Mining Approaches**

This group of scholars began with the following overarching discussion topics:

• What are indicators of good collaboration?
• What are some tools for text mining
• What are the limits of text mining? What can and can’t it do well?
• What are the different levels of analysis and problems associated with identifying patterns between speakers?

The team discussed the meaning of “quantitative” and “qualitative.” This led to discussions of content analysis and quantifying of qualitative data. It was decided that computational methods are different because they are not manual. For example, computational approaches can help segment the data and determine what to do discourse analysis on. It then seemed valuable to forget for a moment how we analyzed computationally, and instead think about what metric(s) we might want, which would capture what good collaboration is. Suggestions were that good collaborative discourse is as informal as possible. Another was that quality of collaboration focuses on two things, synthesis and negotiation. Synthesis, meaning to what extent are people noticing information from a variety of people.

This group began with a discussion of the need to do close discourse analysis and methods to analyze collaboration if researchers want students to connect discourse with disciplinary aspects of the course. Questions of how to analyze data better when the data is scaled up arose. When arriving at desirable and undesirable indicators of CL, the team determined that the following would be useful prompts to bear in mind:

• Consider disciplinary engagement.
• Think about social/emotional factors.
• Account for cognitive factors.
• Goal is to identify things that would be detectable by a machine.

The group discussed that when our dialogues are all task-oriented, such as problem solving, computer science learning and debugging, we look for someone not just to give a suggestion in a vacuum but to say how the suggestion relates either to what they did or to other people’s posts (interactivity). In addition, participants agreed that this level of collaboration was more likely to be afforded by ill-structured problems or complex tasks.

Although there was some discussion of whether developers would develop the kinds of tools that CSCL researchers need, the group noted that companies developing dialog systems tend not to consider conversation from a sociocultural perspective. However, a discussion of existing text mining technologies ensued and discussed several tools:

• Lightside (Mayfield, Adamson, & Rosé, 2013): Features extraction from a set of textual items (Looks for commonalities in those items in different categories (if the categories are coded). It can use linguistic features and things like length of thread, number of upvotes, and number of times it had been read.
• Coh-Metrix (McNamara, Louwerse, McCarthy, & Graesser, 2010): Tools that work in an unsupervised way, the work originated with student essays, not designed for collaboration.
• Tools in R: Handles sentiment analysis, but need to know either exactly how they work or do manual validation of them.

At the end of Workshop 2, it became clear to participants that further discussion was merited. A small group met again at CSCL 2015 in order to discuss the three major themes discussed above. Some of that discussion has been weaved into the notes above. Based on these questions and ideas, a smaller team would hone in on six central themes for further discussion in a third workshop.
Group 4 Collaborative Problem Solving
The themes for discussion were:
• Why is collaboration important?
• What are the types of interactions that foster the desired outcomes?
• Is it important that collaboration be seen as a 21st century competency that could be infused in NGSS, common core, ISTE, etc.?
• Can analytics act as a mediating object between students and teachers?
• How hard is it to assess Collaborative Problem-Solving (CPS) skills?
• What is the difference between developing processes to support collaborative learning and how to develop the best support content learning?

Conclusion: Outcomes
After a brief third summative workshop meeting, it became clear that collaborative processes are becoming increasingly necessary for successful science practice, complex problem solving, and business innovation. Because of the complexity associated with supporting collaborative learning activities, it is not surprising teachers with less teaching experience may depend more on traditional instructivist teaching practices that focus on individual learner. Technology can play an important role in helping to address this problem, but only if we carefully design technology to help foster support for collaborative processes and students’ abilities to eventually manage collaborative processes without technological assistance.

Collaborative learning activities are generally complex and require collective sense-making. A broad definition defines collaborative learning as “small groups of learners working together towards a common goal, whether it is one that they have set themselves or one that has been externally set (e.g., by a teacher); the tasks are intended to promote learning through the interactions” (Hmelo-Silver & Chinn, p.849). When collaborating, group members depend on others to share what they know to synthesize collective information, create shared understanding, negotiate what the group knows, develop a shared understanding, and use the group’s collective cognitive efforts to create new knowledge or solve complex problems (Roschelle & Teasley, 1995; Borge, Ong Shiu, & Rosé, 2018).

There are specific qualities that are associated with productive collaborative learning (Hmelo-Silver & Chinn, 2016). These include (1) socioemotional and disciplinary engagement, (2) positive interdependence, (3) mutual respect, (4) high-quality social and cognitive strategy use, i.e., uptake of peers’ ideas, and communication of alternative perspectives, (5) balanced task and verbal participation, and (6) co-regulation and shared regulation.

The core practices for productive collaboration in task management and regulation include:
• Balanced task participation: Group members take on equal shares of the work and responsibility for its completion, ensuring that all have equal access to important materials.
• High quality cognitive strategy use: Group members use different strategies to ensure that desired domain methods and techniques are carried out.
• Positive interdependence: Group members depend on each other for developing and refining ideas for the project and that the task not split up and completed by isolated individuals.

Collective meaning-making
Collective meaning-making is one of the most fundamental practices in collaborative activity (Stahl, 2006; Suthers, 2006). Group members share and negotiate meaning to create shared understanding for the purpose of extending knowledge. The group must work to make sure that they all understand shared ideas from the perspective of the person who shared them.

The core practices for productive collaboration in collective meaning-making include:
• Disciplinary engagement: Group members collectively think about and discuss important domain concepts.
• Balanced verbal participation: Group members contribute equally to collective meaning making process such that no member or members are dominating verbal discourse or are largely inactive during conversation.
• High quality socio-cognitive strategy use: Group members use different strategies to ensure that ideas are shared, elaborated, analyzed, and checked for understanding.
Socio-emotional Interactions
Socio-emotional engagement refers to the extent to which members are equally invested in quality of the group outcomes, including the group’s and individual member’s well-being, as well as what is produced by the group. The core practices for productive collaboration in Socio-emotional interactions include:

- Emotional engagement: Group members are invested in the well-being of the group, its members, and the project.
- Mutual respect: Group members treat each other, their ideas, and their efforts with esteem.
- High quality social strategy use: Group members use different strategies to ensure that all members feel that their ideas are valued, that they can admit when they do not understand something, and their mistakes will be seen opportunities to learn.

Synthesizing coding schemes
Additionally, a paper on coding schemes as lenses on collaborative learning was shared with the group. Twenty-one coding schemes, varying from face-to-face to online settings to small and large-scale data sets, were collected from nine researchers from a range of disciplinary backgrounds who participated in the conferences and meetings. We also examined coding schemes for collaboration or collaborative learning from iJCSCL journal articles from March 2015 to March 2017, which included 11 additional coding schemes from 11 articles. In total, we analyzed 387 codes from 32 coding schemes. A study that grew out of the first workshop produced 13 final categories across the cognitive, integrated, and social domains. The integrated domain included codes that had both cognitive and social elements. We conducted frequency counts in each category and ran frequencies of coding schemes that included at least one code in that category. In total, there were 68 original codes in the cognitive domain, 193 original codes in the integrated domain, and 107 original codes in the social domain, with 19 in the “other” category. Examining the distribution of codes shows that the integrated domain was the most prevalent, accounting for nearly 50% of the total codes and all but one coding scheme including at least one code from the integrated domain. Within the integrated domain, elaboration was the most common and was included in an overwhelming majority of the coding schemes (81%). The cognitive domain codes accounted for approximately 18% of the total codes and at least one cognitive domain-related code was included in 69% of the total coding schemes. The study synthesized the existing understanding of collaborative learning among researchers and identified the big themes that researchers have explored.

CSCL and Equity
Additionally, as discussion continued the idea of educational equity took a prominent spot, resulting in a CSCL conference submission (discussed below). These discussions included whether or not CSCL should give up on educational change, a discussion of what an adaptive support best scenario might look like, and the political turn in the learning sciences broadly that seeks to understand what role learning research can play in improving lives and in social justice. The idea of educational equity came into play. Defined as equitable access to ambitious collaborative practices – the kinds valued by CSCL as promoting learner agency, dialogue, and identity development – as well as CSCL’s understanding of scripts and scaffolding has the potential for unique and useful contributions to educational equity projects.

As the thinking on Adaptive Collaborative Learning Support (ACLS) and equity came into focus, contributors explored the possibilities of challenges of ACLS for supporting student agency and educational equity projects (See Uttamchandani & Hmelo-Silver, 2019). Specifically, agency as linked to identity development defined broadly, where in equitable learning situations students can take on and take up disciplinary identities rather than be rejected by them (Bell, van Horne, & Cheng, 2017) was considered. Adaptive support came forward as a possible answer to these challenges as it can focus on collaborative practices. In this way, even when agnostic to content, support could help students it understands as frustrated by prompting collaborative learning strategies or could, based on various learning analytics, prompt teachers to visit groups that appear especially sullen or enthusiastic.

Finally, a white paper is in preparation (Borge, D’Angelo, Wise, Uttamchandani, Lester, & Hmelo Silver, in preparation). While the white paper will touch on all progress made in these workshops, the focus will be on the equity issues that culminated in the final workshop. In the paper, we argue that when educational equity is defined as equitable access to ambitious collaborative practices that promote learner agency, dialogue, and identity development, CSCL’s understanding of scripts and scaffolding has the potential for unique and useful contributions to educational equity projects. However, there is tension between viewing scripts as compromising learner agency versus viewing scripts as supporting new forms of learner agency (Wise &
Schwarz, 2017). This led to the reflection that one grand challenge for ACLS is being able to provide the structure that promotes learner agency for all learners.

Final Thoughts
One outcome of this project is a common, shared language for talking about different aspects of collaborative learning that can be helpful conducting, reporting, and comparing research on collaborative learning. It can help researchers situate or advance their work through using existing coding schemes as a shared way of talking about collaboration. The project synthesized how cross-disciplinary collaborative learning researchers conceptualized high-quality collaboration and indicators of lesser quality. Such efforts can help to increase understanding among these different research communities and help different stakeholders in the Computer Supported Collaborative Learning (CSCL) community design for (and study) collaborative learning. We believe that practitioners can make use of the core practices identified in this project for productive collaboration in task management and regulation, collective meaning-making and socio-emotional interactions. Additionally, our hope is that researchers can further explore the interactions between these core practices and their benefit to high quality, productive collaboration.

Finally, the project led to thinking of ACLS as an equity issue to provide access to collaborative learning to a broader range of students. Considering ACLS as an equity issue to provide access to ambitious forms of collaborative learning has potential to serve underserved students and support their teachers in bringing reformed models of teaching to learners who do not usually have those opportunities. We conclude that if educational equity is defined as equitable access to ambitious collaborative practices that CSCL’s understanding of scripts and scaffolding must explicitly attend to and design for educational equity.

References


**Acknowledgements**

These workshops were supported by NSF grant # DGE #1547731. We are thankful to all the participants in the workshop series and in particular those who participated in workshop 3 (Marcela Borge, Alyssa Wise, Cynthia D’Angelo, and James Lester) and Carolyn Rosé for hosting Workshop 2. We thank Ed Gentry for coordinating the workshop and for his extensive editorial assistance.

**Appendix**

Participant List: [http://go.iu.edu/2bHb](http://go.iu.edu/2bHb)