Towards Instrumenting Collaborative Learning and Assessment in the Digital Ocean

Ilya Goldin, Center for Digital Data, Analytics, and Adaptive Learning, Pearson
Bruce M. McLaren, Human-Computer Interaction Institute, Carnegie Mellon University

Abstract

The necessary tool to enable effective and engaging learning is data-driven science, as articulated in the vision of the Digital Ocean (DiCerbo and Behrens, 2012). Collaboration activities are an important class of learning activities that may promote learning. However, because it is not trivial to instrument collaboration activities for comprehensive data collection, instructors and researchers may be deprived of insight into the learning process. We outline some relevant issues in data gathering in collaborative learning activities.

Position

Our goal is to teach the whole learner and to assess the whole learner. Further, our goal is to do this at scale and consistently, and to improve over time. As far as learning requires both individual and collaborative activities, these ought to be related to each other and deployed flexibly in a complementary, comprehensive manner.

Both teaching and assessment are loaded, inadequate terms. When we say “teach”, what we mean, among other things, is to create a time and a place where instructional resources are offered to the learner, the learner is able to and is motivated to take up these resources, the resources are effective and appropriate to the goals of the institution, teacher and learner, where students may collaborate and co-create knowledge (Scardamalia & Bereiter, 1996), where students practice the language of the domain of study as they create an identity as members of a community of practice (Lave & Wenger, 1991), where the motivation is as intrinsic (Ryan and Deci, 2000) as practically possible, where the learner exercises agency and control over the learning process.

When we say “assess”, what we mean is that, because the success of the teaching enterprise is both paramount and difficult to ensure, the learner must be engaged in activities that reveal whether, in fact, the learning we hope to facilitate has taken place. In addition, when students work together to learn together, as they do with our instructional system, we want to ensure smooth and effective collaboration, e.g., participation of all students, students sharing complementary knowledge, etc. Ideally, assessment activities and the activities of teaching are one and the same (DiCerbo and Behrens, 2012), and no assessments are administered solely for measurement purposes.

We see technology as a necessary ingredient to accomplish this at scale and consistently. For instance, individualized tutoring to skill mastery is very effective for acquisition of procedural knowledge (VanLehn, 2011), however scaling this approach to large student populations requires automated tutoring systems. Similarly, automation is necessary for systematic facilitation moves that are conditional on particular learner behavior in collaborative activities. (Kumar et al, 2006)

Technology aids us in a second way, namely by enabling unobtrusive data collection on the *process* of learning, not merely the outcome. The Digital Ocean of data about a learner (DiCerbo and Behrens, 2012) is characterized by “ongoing, ubiquitous” data collection and “dramatically large amounts of data”, collected inside and outside the classroom. For instance, a learning activity may be instrumented to log not only whether a problem-solving attempt led to a correct response, but also specific problem-solving errors, help requests, interaction with manipulative objects and learning resources, interactions between students (McLaren, Scheuer, and Miksatko, 2010), and the time taken on each activity step. By gathering data on the process and connecting process to outcome, we can identify student strategies that are unsuccessful, can intervene with learners who use unsuccessful processes, and improve learning and teaching for current and future students.

Yet another benefit to the use of technology is that it lets us express our theories of collaborative learning in a falsifiable form. For instance, if we posit that particular behaviors in collaborative activities ought to correlate with a latent trait in a learner, then computational models can be used to estimate the correlational (or even causal) strength of that relationship. If the relationship is weak, that may put our theory into doubt. (A weak relationship could also be due to other factors, e.g., poor operationalization, insufficient data sets, low experimental power, etc., but theory falsification via modeling must be considered as an explanation.) Further, computational models may be revised as a way of theory refinement.

However, collaborative learning activities present a particular challenge to this vision, because collaborative learning activities are not necessarily digitally mediated, such as the case with face-to-face conversations. Activities that are not digitally mediated are not easily amenable to instrumentation, which complicates, if not precludes, comprehensive data collection, and therefore deprives instructors and researchers of insight into the learning process.

The issue of instrumentation is not addressed comprehensively by the field of Computer-Supported Collaborative Learning. Because of the multitude of ways that collaboration, learning and assessment are conceptualized in CSCL and the potential lack of alignment between activities, behaviors, learning and assessment (Chan and Van Aalst, 2004), different researchers focus on different aspects of data collection. For instance, some focus on process data aside from outcome, or vice versa, or fail to collect both individual and collaborative process and outcomes; others collect both kinds of data, but do not preserve sequence or time on task, so a process-outcome connection is impossible. (Von Davier and Halpin, 2013) Further, because there are so many possible causal relationships (e.g., the individual contribution at some time point may depend both on the state of the group and the state of the individual at prior time points), large data sets are required to address research questions with robust computational models.

In sum, we propose that there is fertile ground for research in collaborative learning as defined by the following intersection:

- learning and assessment activities that are digitally instrumented for ongoing, comprehensive data collection
- logging of both individual and group moves, work products, and interactions, with time, sequence, and outcome data; tagging / categorization of contributions and sequences of contributions
- logging of traditionally non-captured data, e.g., gestures, expressions, non-digitally mediated discussion
- computational modeling both as a way of testing theories of learning and collaboration, and as a mechanism for adaptive, personalized learning.

A Few Specific Proposals

The issues described above stand in the way of addressing some research questions in collaborative learning. Below are some specific proposals to expand the universe of data instrumentation in collaborative learning. These are not silver bullets of data collection, but they may be helpful. We present these as guidelines or heuristics that can guide the design of a collaborative activity. Many are not originated by us, but we have not seen them collected together or called out as choices in a design space.

Direct vs. ambient instrumentation: When student moves, utterances or work products are created in a digital space, some instrumentation is feasible by capturing the student's interaction with the digital interface, such as logging chat messages when students communicate by typing. However, if students communicate by speaking face to face, as they might in a classroom, then direct instrumentation is not available. Nonetheless, having digital recording devices nearby enables ambient instrumentation. (Blikstein, 2013; Oviatt & Cohen, 2013) For example, if two students share a tablet, the tablet may record the audio of student voices, the position and movement of the tablet itself, and other such data. Careful time-stamping of ambient data will enable subsequent cross-referencing between ambient and other records. We might find salient learning differences between a group of students where one is designated as a scribe, and another group who actively pass a tablet back and forth and all engage in writing on the tablet. Further, prosodic features of dialog may enrich our understanding of collaboration, even aside from any speech recognition technology. Ambient instrumentation is relevant even if work product is digital, e.g., observing facial expressions. (McDaniel et al., 2007)

Open vs. closed vs. scaffolded input: In exactly the same way that a multiple-choice test item presents a closed set of choices, and a free-response test item is an open form, collaborative interactions may be seen as open and closed. A natural language chat is an open activity, and interpreting students' messages, even with direct instrumentation (i.e., with no need for speech recognition and natural language processing), is a challenge. (Rose et al., 2008) Crucially, a collaborative activity may contain both closed and open aspects. For example, a human-human dialog may be replaced with a human-human-virtual agent trialog, where the virtual agent occasionally poses yes/no questions. Further, dialog moves may be scaffolded, e.g., with sentence openers that enable subsequent categorization of utterances. (Baker and Lund, 1997; Scheuer et al, 2013; Soller, 2005)

Raw vs. transformed observations: As with any educational technology, collaborative learning activities emit observations that may be stored "raw", with no transformation, or they may be transformed, e.g., through a "scoring" process. (Almond, Steinberg, & Mislevy, 2002; Baker et al, 2007) A natural language chat may be stored as-is, or it may be scored in multiple ways, such as for whether or not it entails a concept or invokes a conversation strategy. This distinction is especially powerful in how it interacts with *validated* observations, below. Either raw or transformed observations can be collected in structured computational representations, i.e., student and activity models.

Scrutable vs. presented vs. concealed learner and activity models: Student and activity models may be *mirrored* to some or all of the collaborators (Jermann, 2004), or presented to the instructor or another third party (McLaren, Scheuer, and Miksatko, 2010). Because "mirroring" connotes displaying something exactly as it is, we prefer the term "presenting" to allow for displaying transformed observations. Presentation may vary in salience, e.g., from unobtrusive to prominent. Further, presentation may be *scrutable* (Kay, 2006) such that participants can examine, interrogate and possibly modify the models. Whatever is not presented is, by definition, concealed; concealment may be a rational choice, e.g., when presenting would distract from the task at hand.

Validated vs. unacknowledged observations: The most prominent form of presenting an observation is to require a learner to affirm its validity. Consider the difference between presenting two student dialogs, a typed one and a spoken one, that are displayed to the dialog participants in real time. A typed chat is conventionally displayed exactly as and when it is entered, whereas a spoken chat must undergo speech recognition before it can be presented. Because automated speech recognition is error prone, the system may present this dialog in a way that allows the participants to approve the transcript, i.e., giving the machine feedback on its success in transforming the raw observation. There is no comparable value in asking the students to validate a typed chat transcript, unless perhaps this functions as "metacognitive mirroring". (Jermann & Dillenbourg, 2008)

Data accumulation: Data accumulated in a digitally instrumented environment may be used to characterize students and activities in a variety of ways. Suppose that students are organized into groups in the first meeting of a semester-long class, and each groups engages in collaborative activities over the course of a semester. We might be able to somehow chart the performance of each group on each task over time. However, it would be difficult to decouple the contributions of individual students from the activity in each group; if a student in a group participates in few discussions, is that because the student is quiet, or because the remainder of the group does not value the student's contribution? An alternative activity structure that does provide some data on each student is to regroup students on a regular basis; we can then know if a student remains participates little across many groups. Further, having good data could help us create effective groups. (Crespo García, Pardo, & Delgado Kloos, 2004) This is akin to exploiting the birthday paradox for purposes of experimental and statistical inference.

Activity alignment: To exploit data accumulation for rich characterization of the objects of interest, activities may be aligned with the object of interest as the axis. Some examples: a student who participates in collaborative and non-collaborative learning activities can be characterized in terms of capacity for learning overall, or separately in both types of activities. Students may be asked to study a domain concept / knowledge component in collaborative and non-collaborative settings; if success rates change differently over time across the two settings, that may suggest that the knowledge components in the activities are distinct, i.e., that the domain model is poorly structured.

Conclusions

We have called out some distinctions in the design of collaborative learning activities for the purpose of digital instrumentation and data collection.

The space of activities in collaborative learning that we can present to students is, practically speaking, infinite. For instance, in the activity of peer review (Goldin, Ashley and Schunn, 2012), it is conceivable to vary the product/output of the activity, how the peer groups are constituted (e.g., purposefully or randomly), the size of the group, the roles in a group, the connection to other forms of assessment, and many other factors. Moreover, the context and goals of the activity will also affect how students relate to and therefore act in the collaborative process. In light of this variability, how should we construct collaborative activities? We propose that an important aspect of the construction of the activity is our ability to instrument the activity and capture high-fidelity digital data on process and outcome.

Intentionally, this is not a comprehensive view. The focus on data instrumentation in collaborative activities leaves unaddressed issues such as which activities promote learning or collaborative skills. The design space is also not a framework, e.g., unlike Mirroring to Guiding (Soller, Martinez, Jermann, & Muehlenbrock, 2005).

Nonetheless, we hope that this small catalog of data instrumentation issues may aid in the design of collaborative learning activities, and to advance research and practice in collaborative learning.

References

Almond, R., Steinberg, L., & Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. The Journal of Technology, Learning and Assessment, 1(5). Retrieved from https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1671

Baker, M., Andriessen, J., Lund, K., Amelsvoort, M., & Quignard, M. (2007). Rainbow: A framework for analysing computer-mediated pedagogical debates. *International Journal of Computer-Supported Collaborative Learning*, 2(2-3), 315–357. doi:10.1007/s11412-007-9022-4

Baker, M., & Lund, K. (1997). Promoting reflective interactions in a computer-supported collaborative learning environment. *Journal of Computer Assisted Learning*, 13, 175–193.

Blikstein, P. (2013). Multimodal Learning Analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 102–106). New York, NY, USA: ACM. doi:10.1145/2460296.2460316

Chan, C. K., & Van Aalst, J. (2004). Learning, assessment and collaboration in computer-supported environments. In J.-W. Strijbos, P. A. Kirschner, & R. L. Martens (Eds.), What we know about CSCL (pp. 87–112). Springer.

Crespo García, R. M., Pardo, A., & Delgado Kloos, C. (2004). An adaptive strategy for peer review. In 34th Annual Frontiers in Education (pp. 765–771). Savannah, GA, USA. doi:10.1109/FIE.2004.1408636

DiCerbo, K. E., & Behrens, J. T. (2012). Implications of the digital ocean on current and future assessment. In Computers and their impact on state assessment: Recent history and predictions for the future (pp. 273–306). Charlotte, NC: Information Age.

Goldin, I. M., Ashley, K. D., & Schunn, C. D. (2012). Redesigning Educational Peer Review Interactions Using Computer Tools: An Introduction. *Journal of Writing Research*, 4(2), 111–119.

Kumar, R., Rosé, C., Aleven, V., Iglesias, A., & Robinson, A. (2006). Evaluating the effectiveness of tutorial dialogue instruction in an exploratory learning context. In *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 666–674). Springer.

Jermann, P. (2004). Computer support for interaction regulation in collaborative problem-solving (PhD Dissertation). University of Geneva, Switzerland. Retrieved from https://documents.epfl.ch/users/p/pj/pjermann/www/thesis/thesis-jermann.pdf

Jermann, P., & Dillenbourg, P. (2008). Group mirrors to support interaction regulation in collaborative problem solving. *Computers & Education*, 51(1), 279–296. doi:10.1016/j.compedu.2007.05.012

Kay, J. (2006). Scrutable Adaptation: Because We Can and Must. In V. P. Wade, H. Ashman, & B. Smyth (Eds.), Adaptive Hypermedia and Adaptive Web-Based Systems (pp. 11–19). Springer Berlin Heidelberg.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

McDaniel, B., D'Mello, S., King, B., Chipman, P., Tapp, K., & Graesser, A. (2007). Facial features for affective state detection in learning environments. In *Proceedings of the 29th Annual Cognitive Science Society* (pp. 467–472).

McLaren, B.M., Scheuer, O., & Mikšátko, J. (2010). Supporting collaborative learning and e-Discussions using artificial intelligence techniques. *International Journal of Artificial Intelligence in Education (IJAIED) 20*(1), 1-46.

Oviatt, S., & Cohen, A. (2013). Written and multimodal representations as predictors of expertise and problem-solving success in mathematics. In *Proceedings of the 15th ACM on International conference on multimodal interaction* (pp. 599–606). ACM.

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. doi:10.1007/s11412-007-9034-0

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67. doi:10.1006/ceps.1999.1020

Scardamalia, M., & Bereiter, C. (1996). Computer Support for Knowledge-Building Communities. In *CSCL: Theory and Practice of an Emerging Paradigm* (pp. 249–268). Mahwah, NJ: Erlbaum.

Scheuer, O., McLaren, B. M., Weinberger, A., & Niebuhr, S. (2013). Promoting critical, elaborative discussions through a collaboration script and argument maps. *Instructional Science*. doi:10.1007/s11251-013-9274-5.

Soller, A. (2004). Computational Modeling and Analysis of Knowledge Sharing in Collaborative Distance Learning. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 14(4), 351–381.

Soller, A., Martinez, A., Jermann, P., & Muehlenbrock, M. (2005). From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning. *International Journal of Artificial Intelligence in Education*, *15*(4), 261 – 290.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist, 46(4), 197–221. doi:10.1080/00461520.2011.611369

Von Davier, A. A., & Halpin, P. F. (2013). Collaborative Problem Solving and the Assessment of Cognitive Skills: Psychometric Considerations (ETS Research Report No. RR-13-41). Princeton, NJ. Retrieved from http://www.ets.org/Media/Research/pdf/RR-13-41.pdf