Collaborative learning, interdependence, and dyadic data analyses: Building knowledge and community practices

Lenka Schnaubert, University of Duisburg-Essen, lenka.schnaubert@uni-due.de
Rachel Lam, ETH Zurich, rachel.lam@gess.ethz.ch
Cynthia D'Angelo, SRI International, cynthia.dangelo@sri.com
Anne Deiglmayr, ETH Zurich, anne.deiglmayr@ifv.gess.ethz.ch
Claudia Mazziotti, Ruhr-Universität Bochum, claudia.mazziotti@rub.de
Freydis Vogel, Technical University of Munich, freydis.vogel@tum.de

Abstract: There is dilemma in dealing with dyadic data from collaborative learning studies, in that there is an implied (inter)dependence between learner-partners and the assumption of *independence* of subjects when quantitatively analyzing outcomes with common statistical methods. The goals of this workshop are to connect researchers within the Learning Sciences that share the interest of addressing the challenges and importance of quantitatively analyzing dyadic data towards forming an international, interactive space to create better access to information, tools, mentorship, advisement, and discussions specific to dyadic data. The main activities in the proposed half-day workshop include: (1) a thematic introduction by the workshop organizers; (2) discussions in small groups to identify key issues regarding dyadic data analyses; (3) formation of task forces to develop action plans to tackle issues; and (4) plans to build a community that could serve the needs of learning scientists that are involved in dyadic data analysis.

Introduction, Relevance to the Field, and Goals

Collaborative learning can be characterized by learners interacting while mutually influencing each other's cognitive processes (e.g., Dillenbourg, 1999). When designing for and investigating collaborative learning, researchers typically intend for small groups of students to interactively build upon each other's contributions, transactively exchange ideas, and share a joint focus of attention (e.g., Barron, 2003; Chi, 2009). Thus, it is conceptually inherent for collaborative learning to create some form of *interdependence* between learners. Focusing on quantitative data in particular, throughout the last decade, researchers have developed a variety of research designs and techniques for analyzing students' collaborative learning processes and outcomes (Häkkinen, 2013; Strijbos, 2016). However, despite a growing tradition of analyzing data from collaborative learning contexts, statistically handling interdependence remains a considerable challenge (Cress & Kimmerle, 2017; Kenny, Kashy, & Cook, 2006).

With respect to determining differences between conditions or effects of instructional interventions, some of the most common statistical analyses used in the field carry the assumption of independent subjects. By design, collaborative learning environments theoretically violate the assumption of independency, especially if there is interest in examining individual outcomes after learners engage in collaboration. Thus, a core issue around this dilemma is in how to deal with interdependence in different layers of analysis; we must embrace the very dependence we design for, as well as analytically account for shared variance between partners when necessary. The dilemma between the interdependence amongst partners in collaborative groups that researchers intentionally design for can be particularly problematic when analyzing *dyadic* data. Dyadic data is common in research on (computer-supported) collaborative-learning, e.g. in studies on tutoring or peer-feedback dialogue, peer-assisted learning, inquiry, or argumentation (e.g., Asterhan & Schwarz, 2009; Coleman, 1998; King, 1994). Although two interacting learners provide the simplest scenario to study both independent and interdependent learning processes and outcomes, it is difficult to handle the data quantitatively (Kenny et al., 2006).

A troubling approach that researchers might take to resolve this dilemma is in viewing interdependence in data as a statistical nuisance. To offer a simple example, suppose a study's aim is to determine if one instructional collaborative approach is better for learning over another. Prior to the instructional intervention, learners take a knowledge pretest; then process data from student interactions is collected; and finally, learners take a knowledge posttest. It is a relatively straightforward approach to analyze the process data strictly at the dyad level, i.e., with the unit as the dyad. However, if we want to analyze knowledge gains or spoken utterances at the individual level, we are constrained by the common analytic methods available (e.g., ANOVAs and t-tests assume independent subjects; analyzing individual outcomes by averaging between partners at the dyad level reduces power and evens out within-dyad variance). If we conceive this as a "problem" from a purely statistical angle, we may calculate intraclass correlations hoping to find *non-interdependence* so that we can use traditional ANOVAs to make decisions about instructional effectiveness (Cress & Kimmerle, 2017). In other words, we hope to find statistical non-interdependence, ignoring the problematic theoretical implications of partners within a collaborative dyad functioning independently (Gonzales & Griffin, 2012).

Solutions to this problem include the conceptualization, assessment, and joint modelling of intra- and interpersonal learning processes and outcomes (e.g. see mediation models by Deiglmayr, Loibl, and Rummel [2015]). Statistically, multi-level modeling offers a solution where data from collaborative dyads are modeled with partners nested within dyads (Cress, 2008; Kenny et al., 2006; Lam & Mulder, 2017). Yet, within the learning sciences, we are far from established in such methods, practices, or traditions (Janssen, Erkens, Kirschner, & Kanselaar, 2011). Moreover, traditional hierarchical models and multi-level analyses are not ideal for dyadic data, in particular, because regression-based approaches are often not appropriate (Kenny & Kashy, 2011) and analytic requirements (e.g. sample size) can be additional obstacles (Cress, 2008). While there are approaches specifically designed to handle dyadic data, such as the Actor-Partner-Interdependence-Model (e.g., Kenny et al., 2006; for an application in educational intervention providing R-code for the model see Müller, Richter, Križan, Hecht, and Ennemoser [2016]), there is currently only a slim knowledge base around these methods within the (CS)CL community, while both the theoretical and practical challenges remain.

Consequently, the research community does not only face obstacles concerned with developing adequate statistical models and approaches, but also a large knowledge gap within the community around the theoretical issues concerned with dyadic data analyses and the implications for research design. Therefore, the aim of this workshop is not only to give short-term advice on dyadic data analytic approaches, but more importantly to stimulate efforts to empower the research community to begin to tackle these concerns. The particular goals of this workshop are as follows. First and foremost, it is a community building effort that aims to connect researchers that have interest in addressing the challenges and importance of quantitatively analyzing dyadic data within the field of (CS)CL. (We have initiated some efforts over the past year, which we describe below.) Specifically, as a collective group, we intend to more concretely identify key issues, both theoretical and practical; brainstorm strategies for tackling these issues; share our existing knowledge and resources around analytic options and workarounds; and collaboratively develop strategic goals towards growing the knowledge base around dyadic data analysis. Additionally, we hope the workshop can kick-start an international, interactive space for one another as a community of scholars, creating better access to information, tools, mentorship, advisement, and discussions specific to dyadic data from collaborative learning research. Before we outline the specific outcomes and contributions of the workshop, we first describe the workshop structure.

Workshop Structure

We envision a half-day workshop where we collaboratively surface and discuss theoretical and practical issues around dyadic data analyses, share resources that we have worked with to analyze our data, and develop action plans to address the most relevant issues.

- <u>Introduction - Kick-off Presentations from Organizers</u>: The organizers will briefly elaborate on the workshop themes and goals, share their backgrounds on dyadic data analyses, and engage the participants to share concerns and ideas. Afterwards, key contributors will briefly present on major issues in quantitatively analyzing dyadic data, including methodological concerns and examples of handling real data.
- <u>Identifying Issues - Small Group Activity and Group Discussion</u>: Building on the inputs of all participants, we will form small groups to identify the most pressing issues and concerns within the context of each of our experiences with dyadic data. The outcome will be a short list of main issues agreed upon by all workshop participants that need further attention.
- <u>Creating Task Forces - Small Group Activity (Focus Groups)</u>: We will then reorganize into small focus groups with each addressing one issue towards drafting a statement of practical solutions and clarity in theoretical considerations. We hope to encourage groups to discuss concrete ideas and timelines to set practical goals (e.g., available resources? funding opportunities? interested parties? concrete action steps?).
- <u>Discussion and Summary - Group Discussion (Issues and Dealings)</u>: Focus groups will share the results of their discussions with the larger group. We will provide adequate time for discussing each focus issue in order to create opportunities for elaboration, challenging, and supporting each group. The organizers will work on-the-spot to summarize and integrate these outcomes into key points for a statement to be shared (in a draft form) with the participants at the workshop. The shared statement will be finalized by the organizers after the workshop and sent to all workshop participants for comments and feedback. Through a process of engaging all participants, we will produce an agreed upon collective statement for our community.
- <u>Conclusion and Outlook - Task Forces and Shared Resources</u>: Finally, towards long-term sustainability, we will invite participants to be part of longer standing task forces that can push forward particular issues and begin to implement measures for improving our community resources. Examples could include: drafting

best practice approaches for research designs, methods, and analytic techniques; collecting R, SAS, SPSS, MPlus syntax that can be made public for the (CS)CL community; developing reporting guidelines for analytic decisions regarding dyadic data analyses; engaging in community-building efforts; planning special issues, workshops, and other efforts for disseminating information; engaging with statistics experts (within and outside the field) who can provide assistance. Immediately following the workshop, participants will be invited to access a shared "dyadic data" wiki and join our group mailing list to stay updated and engaged.

Expected Outcomes and Contributions

Towards building a community around dyadic data analysis in (CS)CL, we aim to achieve the following:

- create a collective open document that addresses key issues, approaches, and challenges around analyses
- work the open document into a community vision statement to be approved by all participants
- initiate a shared online space for communication in the form of a wiki-based knowledge interface where we, as a community, share information, code, syntax, analytic strategies/tools and other related resources
- grow the community in a systematic way (which was informally started in 2017)
- set up task forces to address identified challenges, potential solutions, and actions steps to advance the knowledge base

The contributions will be both short- and long-term. For the short-term, we will provide a venue for collective practical guidance and assistance for those struggling with the issues around dyadic data analysis. For the future, we aim to engage members in activities that grow and sustain the community within the learning sciences.

Organizers' Background and Relations to other Events

The organizers are learning scientists that use primarily quantitative and mixed methods experimental research designs in the larger collaborative learning and CSCL community. We are early-to-mid career scholars and mature doctoral students. We offer brief bio's below:

- Rachel Lam is a senior scientist at ETH Zurich and examines how cognitive activities prepare students to learn from collaboration. She has worked with statisticians to analyze individual learning outcomes from dyadic collaborative learning data, using dyadic modeling and an effect size calculation technique for clustered data. She initiated an international group made up of 20 scholars interested in the issues around dyadic data and held the first informal meeting at EARLI 2017. The group has since continued to engage.
- Lenka Schnaubert is a final-year Ph.D. student in educational psychology at the University of Duisburg-Essen. She studies how providing group awareness information supports regulation processes within dyads learning collaboratively. In her studies, she found that providing group awareness information may foster statistical interdependence of related variables. Such punctual interdependencies and their effect on statistical analyses motivated her join the dyadic data analyses group at EARLI 2017.
- Cynthia D'Angelo is a senior researcher at SRI International, focusing on technology-enhanced learning environments and learning analytics. She is studying the use of multimodal student data to help understand and automatically measure the collaborative learning of small groups. She organized a workshop for CSCL 2017 focusing on adaptive supports and evaluation of collaborative learning. She has also organized a workshop for EC-TEL 2016 to foster an international community of TEL and cyberlearning researchers.
- Anne Deiglmayr is a postdoctoral researcher with a focus on dyadic learning in the STEM subjects. In her research, she explores ways of assessing and jointly modelling individual-level data, such as pre- and post-test, and dyad-level data, such as transactive discussions and collaborative inferences.
- Claudia Mazziotti just completed her Ph.D. at the Institute of Educational Research, Ruhr-University Bochum and is currently a fellow at SRI International. She investigates how different kinds of collaborative learning processes relate to students' conceptual knowledge in Productive Failure learning settings. She first met with Rachel and Freydis Vogel to discuss analytical issues of dyadic data at ICLS 2016.
- Freydis Vogel did her doctorate in Educational Science and Educational Psychology about CSCL scaffolded by scripts. She is especially interested in studying the effects of different scaffolds on collaborative learning processes and learning outcomes, and has developed instruments and coding schemes to reveal the most beneficial activities during learning. She is also a part of the informal dyadic data group.

Participation Solicitation

We will solicit participation from the dyadic data analysis group that met at EARLI 2017 and from a workshop that took place at CSCL 2017 via group emails and individual invitation. From there, we will ask our existing group members to share recruitment notices with their relevant networks. Each organizer will also send out notices through each of our connected networks (e.g., EARLI and relevant EARLI SIGs, ISLS, CSCL, AERA).

We aim to accept up to 30 participants. Workshop participants should have some experience with research using dyadic data or a particular interest in the topic; experience in specific statistical approaches is not mandatory. We explicitly encourage senior and junior community members to apply, as we aim for a diverse group of participants. If we receive over 30 applicants, we will implement a set of criteria for selecting the best fitting participants that ensures diversity with regard to participants' interests and academic backgrounds.

Informal Advisory Committee

Several senior researchers from the (CS)CL community support the organizing team in the need for knowledge- and community-building relevant to dyadic data analyses. These include: Daniel Bodemer, University of Duisburg-Essen (Germany); Cindy Hmelo-Silver, Indiana University (USA); Manu Kapur, ETH Zurich (Switzerland); and Nikol Rummel, Ruhr-Universität Bochum (Germany).

References

Asterhan, C.S. & Schwarz, B.B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive Science, 33*, 374-400. https://.doi.org/10.1111/j.1551-6709.2009.01017.x

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Coleman, E.B. (1998). Using explanatory knowledge during collaborative problem solving in science. *Journal of the Learning Sciences, 7*(3-4), 387-427. https://doi.org/10.1080/10508406.1998.9672059

Cress, U. (2008). The need for considering multilevel analysis in CSCL research—An appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning*, *3*(1), 69–84. https://doi.org/10.1007/s11412-007-9032-2

Cress, U., & Kimmerle, J. (2017). The Interrelations of Individual Learning and Collective Knowledge Construction: A Cognitive-Systemic Framework. In S. Schwan & U. Cress (Eds.), *The Psychology of Digital Learning: Constructing, Exchanging, and Acquiring Knowledge with Digital Media* (pp. 123–145). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-49077-9_7

Deiglmayr, A., Loibl, K., & Rummel, N. (2015). The mediating role of interactive learning activities in CSCL: An INPUT-PROCESS-OUTCOME model. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the Material Conditions of Learning: Computer Supported Collaborative Learning (CSCL) Conference 2015* (Vol. 2, pp. 518–522). ISLS.

Dillenbourg, P. (1999). What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1–19). Oxford, UK: Elsevier.

Gonzalez, R., & Griffin, D. (2012). Dyadic data analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA Handbook of Research Methods in Psychology, Vol 3: Data Analysis and Research Publication.* (pp. 439-450). Washington, DC, US: APA.

Häkkinen, P. (2013). Multiphase method for analysing online discussions. *Journal of Computer Assisted Learning*, *29*(6), 547–555. https://doi.org/10.1111/jcal.12015

Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2011). Multilevel Analysis in CSCL Research. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 187–205). Springer US. https://doi.org/10.1007/978-1-4419-7710-6_9

Kenny, D. A., & Kashy, D. A. (2011). Dyadic data analysis using multilevel modeling. In J. Hox & J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis* (S. 335-370). New York, Hove: Routledge.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic Data Analysis*. New York, NY: Guilford Press.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31*(2), 338-368. https://doi.org/10.2307/1163313

Lam, R., & Muldner, K. (2017). Manipulating cognitive engagement in preparation-to-collaborate tasks and the effects on learning. *Learning and Instruction*, *52*(Supplement C), 90–101. https://doi.org/10.1016/j.learninstruc.2017.05.002

Müller, B., Richter, T., Križan, A., Hecht, T. & Ennemoser, M. (2016). How to analyze individual and interpersonal effects in peer-tutored reading intervention. *The Journal of Experimental Education, 84,* 744-763. https://doi.org/10.1080/00220973.2015.1065219

Strijbos, J. W. (2016). Assessment of collaborative learning. In G. T. L. Brown, & L. R. Harris (Eds.), *Handbook of social and human conditions in assessment* (pp. 302-318). New York: Routledge.