Analytics of Social Processes in Learning Contexts: A Multi-Level Perspective

Carolyn P. Rosé (co-chair), Carnegie Mellon University, cprose@cs.cmu.edu
Dragan Gaesevic (co-chair), University of Edinburgh, dgasevic@acm.org
Pierre Dillenbourg, Swiss Federal Institute of Technology Lausanne, pierre.dillenbourg@epfl.ch

Yohan Jo, Gaurav Tomar, Oliver Ferschke
yohanj@cs.cmu.edu, gtomar@cs.cmu.edu, oliverf@cs.cmu.edu
Carnegie Mellon University

Gijsbert Erkens, Anouschka van Leeuwen, Jeroen Janssen, Mieke Brekelmans
g.erkens@fss.uu.nl, A.vanLeeuwen@uu.nl, J.J.H.M.Janssen@uu.nl, m.brekelmans@uu.nl
Utrecht University

Jennifer Tan, Elizabeth Koh, Imelda Caleon, Christin Jonathan, Simon Yang
jen.tan@nie.edu.sg, Elizabeth.koh@nie.edu.sg, imelda.caleon@nie.edu.sg, christin.jonathan@nie.edu.sg, simon.yang@nie.edu.sg
National Institute of Education, Nanyang Technological University, Singapore

Abstract: In the past two decades, the field of Machine Learning has not only greatly expanded in terms of the plethora of increasingly powerful modeling frameworks it has provided, but has also birthed the applied fields of Educational Data Mining and Learning Analytics. Learning Analytics has blossomed as an area in the Learning Sciences, promising impact for various stakeholders working at different educational levels, such as Instructional Designers, Students, Instructors, Policymakers and Administrators. This symposium offers a taste of cutting edge work across each of these levels, with a common emphasis on analytics applied to social processes.

Keywords: learning analytics, social analytics, discourse analytics

Introduction

Over the past two decades, the fields of Educational Data Mining and Learning Analytics have received growing prominence in research, policy and public literature. The fields evolved from the applied disciplines of Machine Learning, Intelligent Tutoring Systems and Data Mining, which in turn have their own roots in Applied Statistics, Cognitive Science, and Computational Linguistics. This invited symposium explores the niche domain Learning Analytics is establishing within the Learning Sciences and to offer a taste of the impact it is having at multiple levels within that sphere. The symposium will open new opportunities for bridge building and examine the commonalities for collaboration across related fields.

The earliest manifestations of Learning Analytics within the Learning Sciences community connected particularly with research into collaborative learning. A decade ago at CSCL 2005, attention was given to a vision for the field over next 10 years, and an important active ingredient in that formulation was the presence of machine learning. In particular, automating analysis of collaborative processes for making support adaptive and dynamic was one of the topics discussed. Nevertheless, in the same conference, none of the sessions were named in a way that acknowledged machine learning or data mining as constituting an area. Instead papers offering modeling technologies were "hidden" within other sessions on topics like Argumentation or Interactivity. From 2005 forward, a trend of increasing attention into machine learning was evident. There was a plenary keynote and two workshop keynotes on dynamic support for collaboration at the first Kaleidoscope CSCL Rendez-Vous held in Villars Switzerland. The trend of increasing attention continued at the 2007 conference, where papers in the area became more frequent. They appeared in sessions entitled "Tools & Interfaces" or "Methods of Scaffolding". In particular, "Methods of Scaffolding" gives evidence of beginnings of a vision that a new form of scaffolding for collaboration was becoming possible. In this session, Frank Fischer, leader in the area of scripted collaboration, presented a paper on the vision for fading of scripted support for collaboration. A major shift was apparent by the 2009 conference where there were two workshops with related work, one on Intelligent Support for CSCL and another on Interaction Analysis and Visualization, where automated and semi-automated analytic technologies were featured topics. In the main conference, a session was included on “Scripts & Adaptation” and another on “Data Mining and Process Analysis”.

In light of the center of gravity within the CSCL community, this symposium offers a taste of cutting edge work across each of these levels, with a common emphasis on analytics applied to social processes.

Could a model of educational design enhance learning analytics?

Pierre Dillenbourg

Many educational scenarios rely intensively on social processes. Therefore, if we want to use analytics to investigate learning processes, one first needs to model the social structure of the learning activities. This social structure is often implicit, but, if the designer would make it explicit, the collected data could be processed in a more accurate way. For instance, if the same lesson includes individual and team activities, the analytics need to map individuals to groups and vice-versa. We therefore developed a language for the purpose of modeling pedagogical scenarios, represented as directed geometrical graphs. Another rationale for this modeling language is that, by making the social structure explicit, rich learning activities that have been designed by the learning sciences community, generally for small classes, could be scaled up for use with thousands of participants.

The vertices of these graphs represent learning activities (in black on Fig. 1) and the edges capture the pedagogical relationship between activities (in red on Figure 1). The objective of modeling orchestration processes has emerged as a challenge in CSCL since typical learning scenarios have evolved towards multi-plane learning scenarios. An example of such a scenario is the group formation algorithm displayed between the second and third activity on Figure 1. The graph edges are associated with data operators that implement these algorithms. A graph of data operators constitutes a workflow, which enables the construction of scenarios that are more sophisticated than those currently implemented in MOOCs. The proposed modeling language is not only relevant in learning technologies, it also allows researchers in learning sciences to formally describe the structure of any lesson, from an elementary school lesson with 20 students to an online course with 20,000 participants. This modeling language formalizes the design of learning scenarios. With that in mind, we can then consider the analytics side.

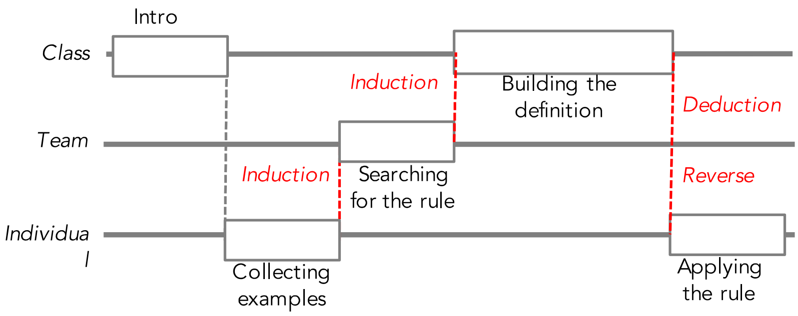


Figure 1. Example of an orchestration graph describing the following scenario: The teacher explains the goal—to find the rule that calculates the number of diagonals in a polygon from the number of edges. He assigns a number between 3 and 8 to each student. Each student draws a polygon with the number of edges assigned to him. Then, students form teams of 4 made up of students who drew a polygon with a different number of edges. They try to find the rule. After a while, each team presents its solution(s) to the class. The teacher compares the invented rules, proposes counter-examples that disprove some proposed rules, and ends up writing the formal rule. Finally, he asks students to apply the rule (G– Deduction edge label) in a reverse way (T– Reverse edge label) to calculate the number of edges of a polygon with 35 diagonals.

Let us consider that the learner state is detected at the end of each activity displayed in the graph (if finer grain analytics are required, any learning activity can be decomposed into sub-activities). In that way, the graph not only describes the pedagogical design of the scenario, but it also captures the sequence of leaner states. More precisely, the sequence captures one dimension of the learner modeling process, which in turn combines 3 sources of information (and is hence represented as a cube on Figure 2):

- Horizontally: since a learning path is a time series, a learners state can be inferred from his previous state. Let’s define $x_t(s)$ as the state of the student *s* at the end of the i^{th} activity (a_i). The state of the class at time *i* is a vector with the distribution of students in each state. When two successive activities have been completed, the two state vectors form a transition matrix M(i, i+1) in which the data in cell (r,c) is the percentage of students who have moved from $x_i(s)$=r to $X_{i+1}(s)$=c. The average entropy across the rows of M determines the weight of the edge between a_i and a_{i+1}: the lower this entropy, the better the current state of a learner can be predicted from her previous state.

- Vertically, the state of student *a* - $x_i(s_a)$ - is inferred from the state of another learner *b* at the same activity $x_i(s_b)$, for instance because s_a and s_b had very similar learning paths so far. The state of student may also be inferred from the state of the whole class, for instance if the student has always been one standard deviation above the class average. Simply stated, if the system has no information about the state of Mike in the third activity but knows that 90% of other learners failed it, it may infer that Mike has a good chance to fail as well.
- The third axis represents the cognitive diagnosis process, i.e., inferring the state of a learner from her behaviour denoted $b_i(s_a)$. This behavior representation refers to her answers, video player actions, assignments, gaze path, blood pressure, etc. In HMMs, behavioural variables are the observable variables and the learners' states are hidden states. This inference process can be a simple mapping between the answers in a quiz and a set of states or it can include a more complex interpretation processes, for instance when analysing gaze traces. Here we have also applied the notion of entropy to describe the probability that the inferred state matches the actual cognitive state of the learners.

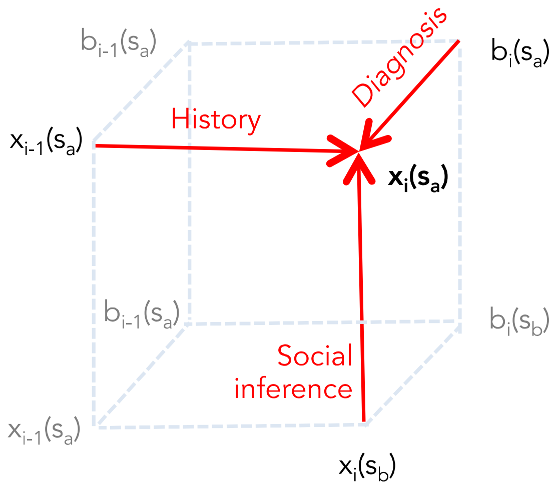


Figure 2. The Learning Analytics cube.

The connection between the graph structure and the cube corresponds to the hypothesis expressed by the title of this talk: I expect that a formal description of the designed educational structure should empower learning analytics applied to this structure. This approach has the ability to use data more insightfully than a naïve approach that ignores the underlying instructional design. Imagine a set of sensors placed in a car: they would generate better deductions if they were associated with a functional model of the car structure that explains the relationship between data points. The same should be true for education. For instance, the work on open learning analytics might lead us to aggregate data across MOOCs, which is a fantastic opportunity for learning sciences, but would produce meaningless results without caution. There is a risk to place in the same data set quizzes in which the learner reasons for 5 seconds (a memory question) alongside quizzes where they work 50 minutes (e.g. computing the noise ratio in a complex system and choosing among one of the proposed values). There is a risk to crunch as if they were similar scores from individual exercises, and even more an issue if team task scores are included as well. The need to describe structures can be addressed with existing taxonomies from researchers such as Bloom, Gagné, and D'Hainaut,using metadata standards (e.g. IMS LD) or with the proposed modeling language.

Learning analytics to support students: Enabling automated interventions

Carolyn P. Rosé, Yohan Jo, Gaurav Tomar, and Oliver Ferschke

An important research problem in learning analytics is to expedite the cycle of data leading to the analysis of student needs and the improvement of student support. On the basis of the importance of social interaction in learning, this work proposes a pipeline that includes data infrastructure for a common representation of social interaction data from multiple platforms; a probabilistic sequence model to analyze the effects of social connections on students' learning paths; and a social recommender system to support students for acquiring positive social capital.

The foundation of computational analytic work is representation of data. Much of our published work in analytics of collaboration in discussion has been focused on either chat data (Howley et al., 2013) or transcribed face-to-face discussion (Ai et al., 2010; Clarke et al., 2013). These can both be represented in a simple, uniform, flat sequence of text segments, each contributed by one speaker. However, when expanding to learning in MOOCs or learning in other online contexts such as opensource communities, the form that the discussions may take becomes more diverse as they are embedded in a variety of platforms. They may even occur simultaneously through multiple separate streams. To that end we offer a publically available data infrastructure we call DiscourseDB (https://discoursedb.github.io/), which enables translation of data from multiple streams into a common, integrated representation. The interface level representation is translated down into Discourses, with embedded Discourse Parts consisting of Contributions, which may be related to one another through Relations, and which are associated with content that can be associated with Annotations. This common representation enables combining data across communication streams and applying common modeling technologies.

As a concrete example, consider connectivist Massive Open Online Courses (cMOOCs) that include environments like the competency-based learning platform ProSolo, featured in a recent edX MOOC called Data, Analytics, and Learning (DALMOOC) (Rosé et al., 2015). In these environments data is rich and heterogeneous. In ProSolo, for example, student behaviors formally within the environment include follower-followee relations, posting wall notes including updates and goal notes, and commenting on notes. Students also engage in threaded discussions, blog and comment on blog posts, and tweet. These behaviors occur within accounts in other linked online community spaces. In a proof-of-concept using data from DALMOOC, we have transformed data from wall post comments, blogs and blog comments, and Twitter into DiscourseDB, and applied probabilistic graphical modeling techniques to identify typical student learning trajectories that could be supported through social recommendation.

Once the data of interest has been represented in a way that is generalizable across sources, the next step is to model student trajectories, especially as they relate to their observed social connections. This analytic approach enables us to identify opportunities where interventions can positively impact student trajectories. We propose a model that automatically extracts student learning paths composed of discussions across multiple platforms and active social engagement. This model aims to detect the pattern of students' learning paths conditioned characteristics of their social connections in a follower-followee network and thereby inform us of the influence of different configurations within the social space on student behavior. We define a student state in terms of the discussed topics and the document types used for discussions (e.g. forum, Twitter, blog), and identify these states in a bottom-up fashion through an integration of graphical probabilistic modeling techniques. Given data, the model infers a set of meaningful states along with the topics and document types for each state. The learned topics may be interpreted as informing us about students' interests and inclinations and how they evolve over time. The learned states provide insight into the ways students adopt different social interaction practices at different times.

Interpretation of the learned model from DALMOOC revealed an important problem. On the positive side, we saw that students who used the ProSolo affordances for setting learning goals persisted longer in the course, did more hands on practice, and spent more of their time in the environment doing course relevant work. Furthermore, students who chose to follow other students who had set concrete goals for the course, became increasingly engaged in course relevant activity. However, students in both of these categories were few and far between. Most students were found not to take advantage either of goal setting affordances or follower-followee affordances.

Nevertheless, the situation is not hopeless. Data mining technologies again applied over the common representation obtained through DiscourseDB enables a potential solution. A popular method for recommender systems is matrix factorization, which identifies a latent representation that connects recommendations to those the recommendations are made to (Yang et al., 2014). A feature-aware matrix factorization approach is able to combine data about preferences with an arbitrary feature representation that augments the latent state representation with information deemed to be potentially valuable in making the recommendations. In particular, we designed a context-aware matrix factorization model that uses features extracted from students' goal setting behavior as additional features. A corpus based experiment shows that our system can find appropriate followees who are not only qualified as positive role models but also relevant based on a model of affinity learned by means of a feature aware matrix factorization approach. Thus, our solution involves an analytics enabled intervention made possible through the DiscourseDB data infrastructure.

Learning analytics to support teachers: Regulating teaching practices through analytics in CSCL

Gijsbert Erkens, Anouschka van Leeuwen, Jeroen Janssen, & Mieke Brekelmans

As explained in the introduction to this symposium, Learning Analytics (LA) can be beneficial for multiple stakeholders. In contrast to approaches that directly target students (such as the work by Rosé et al.), in this contribution we consider LA aimed at supporting *teachers* during the phases of diagnosing, intervening, and evaluating students' activities during computer-supported collaborative learning (CSCL).

In collaborative learning, knowledge is constructed through discourse with other students by sharing and discussing resources and jointly building on task products. Teachers play an important role during the problem solving activities of groups of students. They do so by stimulating meaningful interaction between group members and offering support when needed. The support can be needed when groups experience problems with the task content or with the regulation of the task (cognitive or meta-cognitive level). Support may also be needed with regard to the process of collaboration or to the regulation of collaboration (social or meta-social level). Based on a diagnosis of the progress and quality of a group's' activities, the teacher has to decide whom the intervention will be aimed at and what type of intervention is most suited (direct instruction, hints, supporting questions, etc.). Lastly, the teacher has to evaluate, again by diagnostic observation, whether the intervention achieved its intended effect. This is a very demanding task because of the number of collaborative groups in a classroom, the time pressure, and the multitude and multidimensionality of information that is needed in this cycle of diagnosis, intervention and evaluation. Our hypothesis is therefore that LA could be supportive to teachers.

Within a CSCL environment, all students' actions are typically logged automatically and in real time. In most CSCL environments these actions represent clicks on task resources, interaction with task products and to communication within groups or more broadly within the class. These logs may serve as input for learning analytics tools and provide teachers with information about groups of students regarding both task and collaborative activities. LA can be used for real-time assessment to support teachers' moment-to-moment decision making in diagnosis, intervention, and evaluation for multiple groups simultaneously within a class (or even multiple classes), thereby possibly supporting adaptive teaching. This means that teachers, instead of having to monitor all activities of all groups separately, can receive a summary or overview of the situation regarding multiple groups more easily, which would aid the teacher in providing timely assessment and support.

Another function LA has the potential to fulfill is to analyze and report characteristics of collaboration that otherwise require diagnosis at multiple time points. Aggregating aspects of collaboration that cannot be reduced to a single event into a visible summary means information about such processes is more easily accessible to the teacher. For example, during computer-supported collaborative assignments, because group communication is logged, LA could provide the teacher with up-to-date reports about collaborative processes that would otherwise be hard to keep track of. When LAs are used for the purpose of supporting teaching practices, choices must be made. In this contribution we consider two of those choices, namely 1) focus and granularity of presented information, and 2) distribution of decision making between the teacher and the learning analytics system. Both choices will ultimately affect the teaching practices of the teacher.

Focus and granularity of information

The first choice we consider is which type of information to use as input to LA, and thus the focus of diagnosis presented to teachers. As mentioned, a common distinction in the field of CSCL is the difference between cognitive, task related, and social aspects of collaboration. A further question is at what level of granularity the information is presented. Information can either be given in *real time* as the collaboration unfolds, or as later as an aggregated measure. Table 1 represents the LA tools we have investigated in our research, categorized by focus and granularity of information.

Concerning real time measures, the Shared Space tool (top left) shows the level of agreement and disagreement within a group discussion in a graph alongside the chat, with right representing agreement and left representing disagreement. Furthermore, horizontal placement of chat utterances similarly indicates a status related to agreement and disagreement as computed based on the presence of discourse markers in the utterances themselves. The Concept Trail (top right) shows the occurrences of task relevant concepts (and synonyms) in utterances of the students in a graph on top of the chat window.

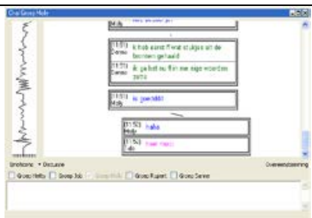
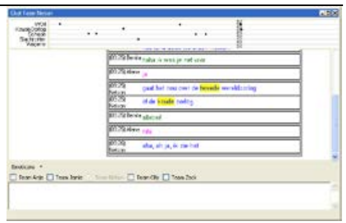
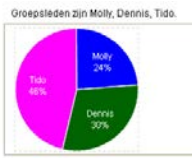
Concerning aggregated characteristics, Participation statistics (bottom left) offers a diagram of proportional participation of group members in chat and productivity tools. Task progress statistics (bottom right) show the relative task progress between the groups in the class as well as their average progress.

In two prior studies, we investigated the effects of LA concerning social activities and cognitive activities on teacher diagnosis and interventions (van Leeuwen, Janssen, Erkens, & Brekelmans, 2014; 2015). When provided with the LA tools the focus of diagnoses or interventions of the teachers were directed by the type of information shown: cognitive or social. Effects of granularity were less clear.

Distribution of decision making between teacher and learning analytics system

LA can support multiple stakeholders. In our research, we have provided LA to *teachers* that represent information on degree of agreement, equal participation, use of task concepts and task progress, but do not give advice or recommendations. In principle, advice and recommendations could automatically be provided to *students* by the learning analytics system as well, as was suggested in the contribution by Rosé et al. A combination of targeting teachers and students is also possible, as demonstrated by the Tan et al. contribution.

Table 1: Focus and granularity of learning analytic tools

	Focus of LA tools			
Granularity	**Social activities**		**Cognitive activities**	
Continuous characteristics of discussion	Shared Space: Occurrence of agreement and disagreement		Concept Trail: Use of task relevant keywords	
Cumulative quantity of activities	Participation Statistics		Task progress Statistics	

Thus, we see that LA tools can be placed on a continuum of how much control and decision making is left to the teacher. On the one end of the continuum are tools that are solely aimed at supporting the teacher in the phase of diagnosis, and the interpretation of the information shown by the tools remains the teacher's purview, as are the intervention and evaluation thereof. Further on the other end of the continuum are LA that are used in scripted scenarios in which tools automatically give alerts or even deliver intervening actions to signaled problems. In the end the question is whether LA tools are used by teachers to regulate their teaching practices or whether LA are used to regulate the teaching practices of teachers.

Learning analytics to support policy: Identifying and fostering 21st century collaborative, critical and connective literacies among diverse learners

Jennifer Tan, Elizabeth Koh, Imelda Caleon, Christin Jonathan, Simon Yang

A major global educational challenge today lies in the question of how schooling systems, policies and practices can more effectively foster 21st century literacies and enhance educational equity among diverse learners—not only at the individual level but more importantly at the collective level. This is with particular sensitivity to the highly-networked, technology-mediated social and learning contexts of modern life. Alongside improved understandings of the dynamic and non-linear nature of 21st century skills and their constitutive socio-interactional processes, policymakers and administrators worldwide are increasingly cognizant of the limitations of conventional modalities of assessment and pedagogic designs. Consequently, many are actively partnering with design-based learning scientists and researchers to explore the affordances that contemporary social learning analytics can bring to bear on this educational imperative of our time.

This contribution provides an exemplification of one such exploratory multi-stakeholder effort in the K-12 schooling sector in Singapore. To this end, we showcase *Wi*READ, a web-based collaborative reading and social learning analytics environment aimed at fostering senior school students' critical reading skills. An important goal of this work is deepening reading engagement levels, and promoting self-regulated and collaborative knowledge construction in the literacy domain, during and beyond formal English class time. Its primary objective is that of motivating and scaffolding students to develop richer dialogue and quality interactions with peers around multimodal texts, thereby enhancing their personal connection to and appreciation for reading as a highly relevant, generative and meaningful social practice.

To achieve this, the techno-pedagogical design of *Wi*READ focuses on 2 key learning affordances: online peer interactions around reading, and the social learning analytics dashboards for students and teachers.

First, the web-based social reading and discussion tool was underpinned by Vygotskian socio-constructivist theories (Stahl, Koschmann & Suthers, 2006) and a Multiliteracies pedagogical framework (Tan & McWilliam, 2009). The Multiliteracies framework comprises 4 essential dimensions of effective 'new literacies' enculturation in learners: situated practice, overt instruction, critical framing, and transformed practice. The micro-level of pedagogical scaffolding scripts comprise 7 critical lenses (message, purpose, audience, assumption, point of view, inference, impact of language/visuals) and 5 critical talk types (I *think* that, I think so *because*, I *agree*, I *disagree*, I need to *ask*), in turn informed by Paul-Elder's (2001) 'wheel of critical reasoning' and our own work on dialogic indicators of collective creativity and criticality (Tan, Caleon, Jonathan and Koh, 2014). These frameworks have served as a meta-schema for guiding students' collaborative critique of texts on *Wi*READ in that students were required to tag each of their comments/replies with 1 critical lens and 1 critical talk type (Figure 3). Each critical lens and critical talk type tag contained a 'popover' that provided students with question prompts and sentence starters, thereby providing students with a constant reference illustrating how each tag could be used to critique texts more deeply.

Second, the social learning analytics affordance of *Wi*READ was designed with the aspiration of providing rich, meaningful and timely formative feedback to students and teachers, so as to help monitor varying levels of socio-interactional reading engagement and progress. In this way, adaptive modifications can be made to learning strategies by students and pedagogical practices by teachers to improve process related learning outcomes. To achieve this, the individual student and consolidated class-level teacher dashboards comprise a range of social learning analytics visualizations that include social networks analytics, discourse and content analytics, dispositional analytics, and achievement analytics (Figure 4).

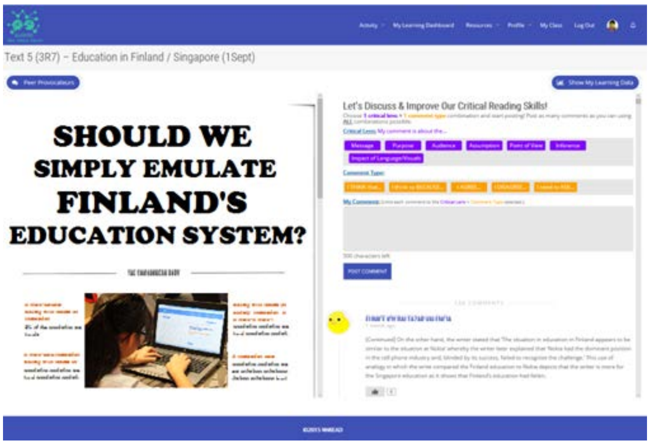


Figure 3. Texts, critical lenses, talk types, and popovers.

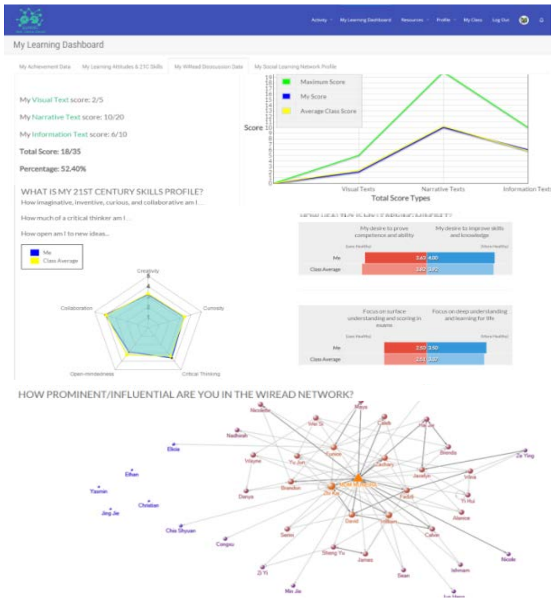


Figure 4. Dashboard visualizations (selected only).

Drawing on a combination of 4 critical stakeholder perspectives—a facilitator-researcher, 3 English teachers, a policymaker and participant students (N=114)—the possibilities of *Wi*READ's social learning analytics will be highlighted in this contribution. This includes visualizations for 1) making visible and motivating students' agentic development of 21st century collaborative, critical and connective literacies, and 2) early identification of and support for disengaged and at-risk learners. This will be complemented by a critical discussion of the pedagogical paradoxes and complexities that accompany these possibilities. We situate this discussion against the backdrop of a higher call by policymakers and funding stakeholders to engender broader adoption, translation and diffusion of 21st century technology-enhanced social learning and analytics innovations such as *Wi*READ beyond one 'seed innovation' school within a relatively centralized and high-performing Singapore education system. In doing so, we foreground the educational promises and problems that can arise as the 'rubber' of well-intentioned learning innovations 'hits the road' of entrenched socio-institutional beliefs and practices in mainstream schooling.

Conclusions

Considering work in this young, emerging area at all levels, we see many common concerns, with multiple stakeholders influencing and benefitting from analytics in each. We see common interests in data representation, modeling technologies, and feedback to end users. We see promise of impact, but far to go in terms of serious deployment and adoption. This innovative work promises to challenge and extent what is possible both in the field of Learning Analytics and in the field of Learning Sciences through bridge building. In the discussion we will reflect on the current state of this work and discuss next steps for productive synergy between fields.

Acknowledgements

This work was funded in part by NSF grant ACI-1443068.

References

Ai, H., Sionti, M., Wang, Y. C., Rosé, C. P. (2010). Finding Transactive Contributions in Whole Group Classroom Discussions, in *Proceedings of the 9th International Conference of the Learning Sciences, Volume 1: Full Papers*, pp 976-983

Clarke, S., Chen, G., Stainton, K., Katz, S., Greeno, J., Resnick, L., Dyke, G., Howley, H., Adamson, D., Rosé, C. P. (2013). The Impact of CSCL Beyond the Online Environment, *Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference 2013, Volume 1*

Howley, I., Mayfield, E. & Rosé, C. P. (2013). Linguistic Analysis Methods for Studying Small Groups, in Cindy Hmelo-Silver, Angela O'Donnell, Carol Chan, & Clark Chin (Eds.) *International Handbook of Collaborative Learning*, Taylor and Francis, Inc.

Paul, R., & Elder, L. (2001). Critical Thinking: Tools for taking charge of your learning and your life. Prentice Hall, Upper Saddle River, NJ.

Rosé, C. P., Ferschke, O., Tomar, G., Yang, D., Howley, I., Aleven, V., Siemens, G., Crosslin, M., Gasevic, D. (2015). Challenges and Opportunities of Dual-Layer MOOCs: Reflections from an edX Deployment Study, *Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 2*

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. Cambridge handbook of the learning sciences, 2006, 409-426.

Tan, J. P-L., Caleon, I. S., Jonathan, C. R., & Koh, E. (2014). A dialogic framework for assessing collective creativity in computer-supported collaborative problem-solving tasks. Research and Practice in Technology Enhanced Learning, 9, 3, 411-437.

Tan, J. P-L., & McWilliam, E. (2009). From Literacy to Multiliteracies: Diverse Learners and Pedagogical Practice. Pedagogies: An International Journal, 4, 3, 213-225

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2014). Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. *Computers & Education*, *79*, 28–39.

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education*, *90*, 80–94.

Yang, D. & Rosé, C. P. (2014). Constrained Question Recommendation in MOOCs via Submodality, *Proceedings of the 2014 ACM International Conference on Information and Knowledge Management.*