CSCL and Learning Analytics: Opportunities to Support Social Interaction, Self-Regulation and Socially Shared Regulation

Alyssa Friend Wise (co-chair), Simon Fraser University, Canada, afw3@sfu.ca
Roger Azevedo (co-chair), North Carolina State University, USA, razeved@ncsu.edu
Karsten Stegmann, Ludwig-Maximilians-Universität, Germany, stegmann@lmu.de
Jonna Malmberg, University of Oulu, Finland, jonna.malmberg@oulu.fi
Carolyn Penstein Rosé, Carnegie Mellon University, USA, cprose@cs.cmu.edu
Nicholas Mudrick, North Carolina State University, USA, nvmudric@ncsu.edu
Michelle Taub, North Carolina State University, USA, mtaub@ncsu.edu
Seth A. Martin, North Carolina State University, USA, samarti7@ncsu.edu
Jesse Farnsworth, North Carolina State University, USA, jjfarnsw@ncsu.edu
Jin Mu, The University of Hong Kong, Hong Kong, jinmu@hku.hk
Hanna Järvenoja, University of Oulu, Finland, hanna.jarvenoja@oulu.fi
Sanna Järvelä, University of Oulu, Finland, sanna.jarvela@oulu.fi
Miaomiao Wen, Carnegie Mellon University, USA, mwen@cs.cmu.edu
Diyi Yang, Carnegie Mellon University, USA, diyiy@cs.cmu.edu

Frank Fischer (discussant), Ludwig-Maximilians-Universität, Germany, frank.fischer@psy.lmu.de

Abstract: Research has generated deep insights into computer-supported collaborative learning (CSCL), but the cycle of impact on practice is relatively lengthy and slow. In contrast, work in learning analytics attempts to leverage the collection and analysis of data to improve learning processes and outcomes in-situ. Developing learning analytics to support CSCL thus offers the opportunity to make our research actionable in an immediate way by using data collected on collaborative processes in-progress to inform their future trajectories. Efforts in this direction are specifically promising in support of students' self- and socially shared- regulation of their learning. Data on collaborative and metacognitive activities can inform collaborating groups and help them to improve future joint efforts. In this symposium we bring together a collection of five papers that are exploring the space of connection between CSCL, learning analytics and self-regulation to advance thinking around these issues.

Keywords: learning analytics, socially shared regulation of learning, prompting, scaffolding

Introduction

Research in CSCL has a recognized tradition of generating insights about how to support collaborative learning with both hard and soft technologies (Tang et al., 2014); yet the cycle of impact on practice is relatively lengthy and slow. Findings from research studies are disseminated through conferences and journals and may eventually be taken up by teachers or designers to productively inform the experience of future students in classrooms. But for those with whom the initial data was collected, the opportunity has passed. In some cases, advances are reified into technological artifacts that can be sent out into the world, but technologies too can be co-opted and still the change and adoption process is slow.

Against this backdrop, the recent emergence of learning analytics as the collection and analysis of data traces used to inform learning activities while they are still in process, offers a special opportunity to "close the loop" (Clow, 2012, p.134) and make CSCL research actionable in a new, more immediate, way. Importantly, moving from CSCL-as-research to CSCL-as-learning-analytics is different from a simple application of previous findings. Rather it requires the generation of previously unexplored kinds of insights into CSCL processes and offers a way to simultaneously achieve high practical impact and new theoretical advances in the field. Specifically, there is an opportunity to unite work in CSCL with emerging research in the area of self-regulated learning (SRL).

Contemporary research on SRL focuses on the collection and analysis of complex, temporally-unfolding data using various interdisciplinary methods. Researchers use a variety of multi-channel SRL data such as log-files, eye-tracking, physiological sensors, facial expressions, utterances, etc. to examine the role of cognitive, metacognitive, affective, motivational, and social processes engaged in by individuals both together and on their own, at times in conjunction with advanced learning technologies. These methods and techniques challenge current conceptions of SRL as a purely individual process while simultaneously addressing emerging conceptualizations, such as socially shared regulated learning (Molenaar & Järvelä, 2014; Hadwin & Järvelä, 2013). Such data on collaborative, metacognitive and other learning-related processes can inform student groups and help them to improve their future joint activity. This process of data-informed reflection and change can also be thought of as a form of socially shared regulation itself.

In this symposium we bring together a collection of five papers that are exploring the space of connections between CSCL, learning analytics and self-regulation. The researchers will each present their work, articulating the central conceptual, theoretical, methodological, and analytical issues that have arisen. The discussant will address how the papers collectively advance thinking around CSCL, SRL and learning analytics.

Collective and individual discussion analytics: Connecting learning intentions, discourse patterns and responsive action

Alyssa Friend Wise

While advances in the availability of data and methods for processing it present exciting opportunities to provide real-time feedback to students on their collaborative processes, translating a CSCL research program into learning analytics is non-trivial. An additional novel knowledge base is needed to leverage CSCL methods and models to be useful in this context. This paper focuses on a critical set of sociotechnical issues related to the use of CSCL analytics by students and teachers. First, analyses of collaborative learning that are meaningful to researchers are often complex, deeply theoretical and involve epistemological entailments. Thus we need to carefully consider what kinds of analyses are appropriate to share with learners, in what form to present them, and how to support their interpretation. Second, the use of learning analytics is fundamentally a process of sense-making, decision-taking, and action; thus another key area for development and research is into interventions that support metacognitive and self-regulatory processes around collaboration. Such interventions are important to help learners and teachers ask useful, relevant and actionable questions of the data (Verbert et al., 2013) as well as to effectively incorporate the use of the analytics into the flow of collaborative activities. Specifically, pedagogical interventions to support student use of collaborative learning analytics can be framed around the principles of integration, agency, reference frame and dialogue and the processes of grounding, goal-setting, monitoring and reflection (Wise, 2014). The issues of what analytics to share with learners and how to do so are explored in more concrete form below in the context of the E-Listening Project (Wise et al., 2014).

One foundational issue in generating learning analytics for CSCL is to be clear about the epistemological unit of analysis and action. A core feature of CSCL is a focus on the process of interaction and negotiation among the collective *group* (Stahl et al., 2006). In contrast, the field of learning analytics has largely focused on the *individual* (at times within a group) as the "target" for analytic insight and resultant action. This tension can be addressed in several ways. First, analytics can be conducted at the level of the group and then presented back to the group *collectively* to inform their future collaborative activities. In the E-Listening Project a graphical representation of the collective discussion is presented as a "Starburst" that the group can use to monitor whether contributions are receiving replies and if threads are being abandoned or ignored (Wise et al., 2013). Of course, the sense of collective responsibility for the discussion that engenders such activity is something that must be deliberately fostered in the community (Scardamalia & Bereiter, 2006). Second, analytics can be conducted at the level of the group, but used to make claims about particular individuals with respect to the group. A classic example of this is social network learning analytics (Shum & Ferguson, 2012) when one takes an ego-centric rather than whole-network view to make inferences about an individual's position in the network (Haythornthwaite & De Laat, 2010). Finally there may be aspects of CSCL environments that can reasonably be analyzed from an individual perspective. Learners' online listening behaviors (the ways in which they attend to existing comments in asynchronous discussion) are one such construct. Various indices of the depth and breadth of individuals' online listening can be calculated, shown back to individuals, and used to inform the subsequent behaviors. We now move from the question of what analytics to how they can be used.

For learning analytics of CSCL processes to help facilitate productive changes in the ways learners interact, they need to have their use framed as an integral part of the collaborative learning activity tied to goals, expectations and a reflective cycle. *Integration* of analytics refers to creating a clear thread between the goals of collaborative activity, the patterns of interaction that support these goals, and the ways in which the analytics reflect such patterns. This can be done through a process of *grounding* in which the parameters of the collaborative activity are established with students a priori. In the E-Listening project this is done through the presentation of guidelines that describe the purpose of collaborative online discussions and what is expected in terms of broad, deep, integrated and reflective attention to the posts of others (Wise et al., 2013). *Agency* in analytics use refers to getting learners (individually or collectively) to be proactively engaged in managing their own collaborative learning process. From a self-regulated learning perspective students can be supported through cycles of *setting proximate goals* (in the contexts of the larger activity goals described earlier) and then monitoring and evaluating progress towards them through *reflective activity* (Winne & Hadwin, 2010). In the E-Listening project this is done via individual goal-setting/reflection journals and collective meta-discussion about the discussions (Wise et al., 2013). Finally in making sense of and taking action based on analytic information, it is also important to consider the *reference frame* for evaluating discourse patterns (e.g. a theoretical ideal, other individuals / groups, changes from the start of the activity) and create a space for *negotiation* in which decisions about changes to the collaborative interactions become objects of attention themselves.

A script theory of guidance perspective on learning analytics for CSCL

Karsten Stegmann, Carolyn Penstein Rosé, and Jin Mu

Approaches to computer-supported collaborative learning (CSCL) are mainly based on a triad of assumptions: (1) Collaborative learning outperforms (under particular circumstances, e.g. with specific support) other methods when it comes to learning outcomes. Collaborative activities like argumentation (e.g. Clark, D'Angelo, & Menekse, 2009), reciprocal teaching or transactive co-construction (e.g. Molinari et al., 2013) are regarded as effective learning mechanisms. (2) Computer support enables both certain learning activities (e.g. simulation-based inquiry learning; de Jong & van Joolingen, 1998) and more direct support for certain activities (e.g. scaffolds as an inherent, but adaptive, component of the learning environment; cf. Koschmann, 1994). (3) The combination of collaborative learning and technology can have positive interaction effects that go beyond the simple combination of main effects. On the one hand, the quality of collaborative learning processes is lifted through adaptive scaffolds that positively moderate the positive effects of collaborative learning. On the other hand, the effects of technology functions (like access to various resources) on learning outcomes are boosted through collaborative learning (cf. Weinberger et al. 2010).

The Script Theory of Guidance (SToG; Fischer et al., 2013) provides a theoretical account to instructional support of collaborative learning activities. A principle of the SToG is that internal scripts, which are comprised of the four components play, scene, role, and scriptlet, guide collaborative learning behaviors. In CSCL research, the internal scripts (including their components) are usually measured using discourse analysis (e.g., Mu et al. 2013). A further important principle is the *optimal external scripting level principle*. According to the theory, external scripts can guide collaborative activities similar to internal scripts. These external scripts work best, according to the principle, if the external script has an optimal fit with the internal script.

Learning analytics may be used to measure internal scripts to enable adaptive collaboration scripts. The measurement of internal scripts usually requires, however, a sophisticated analysis of collaborative processes. To adapt, for example, an external collaboration script that scaffolds argumentative knowledge construction, the actual quality of single arguments and argumentation sequences might be assessed. Using methods of natural language processing, it is possible to measure internal scripts automatically (cf. Rosé, et al., 2008). The quality of argumentation is, however, highly task and context depended. Therefore, methods of natural language processing struggle if task and/or context of collaborative learning changes.

The ACODEA framework (Mu, et al., 2012) showed that this problem could be overcome in part by using a multi-layer procedure that first pre-processes and normalizes data from discussions with different tasks and contexts. In a first step, meaningful attributes are extracted against the background of a certain task and/or context. To assess the quality of argumentation, for example, theoretical concepts important in this specific discussion need to be identified to extract warrants. All utterances with theoretical concepts will be translated into an unified term (e.g. "concept"). In the second layer, the pre-processed data is segmented. In the third layer, coding of the segments is performed using the pre-processed data instead of raw data. While layer two and three are the regular procedure of analysis discourse data, the main difference proposed in the ACODEA framework is a translation of context and task specific raw data into a general common "language" that partials out the concrete content of discussion. There is some evidence that this proposed procedure works across different contents of discussions (Mu, et al., 2012) as well as for different types of text-based communication (Mu et al., 2014), although adaptation to such contextual differences remains an active area of machine learning research. Against this background, the SToG can be used as theoretical foundation of learning analytics that aim to measure internal scripts of collaborative learners. The ACODEA framework, in addition, provides an approach how internal scripts could be measured using natural language processing despite the fact that discussions take place in different contexts and with different tasks.

Do collaborative groups benefit from a shared regulation tool? Sequential analysis of actualized regulation in social interaction

Jonna Malmberg, Hanna Järvenoja, and Sanna Järvelä

The field of self- and socially shared regulation of learning (SSRL) is increasingly interested in how temporal sequences of events (e.g., activating prior knowledge; constructing task perceptions and goals; using and adapting strategies) emerge in different stages of the learning process (Azevedo, 2014; Bannert & Sonnenberg, 2014; Volet et al. 2011; Molenaar & Chiu, 2014). Examining temporal sequences of events that incorporate phases of regulated learning can increase our understanding of the process in which students engage when learning alone or in groups. Earlier research considering sequential and temporal aspects of regulated learning focused on individual learning, but there is not much research focusing on capturing temporal sequences of regulation in collaborative groups in terms of how group members establish socially shared regulation in authentic group learning situations.

The problem is that group members often fail to recognize the target for SSRL and tend to use superficial strategies (Malmberg et al., 2014). A vast body of technological tools has been developed to support awareness of SSRL, but mostly this has happened at group level in on-line learning settings without giving guidelines of what SSRL strategy to use (Järvelä et al., 2015). Currently, there are no technological tools aiming to explicitly prompt SSRL in face-to-face collaborative learning. This study aims to capture patterns of social interaction through which collaborative groups actualize socially shared regulation. Furthermore, it aims to investigate whether groups benefit from the use of a tool designed to promote strategies for socially shared regulation of learning in face to face collaboration.

Second-year teacher-education students (N = 44, 36 females, 8 males, mean age 25 years) participated in a mathematics didactics course that lasted for two months. The math course comprised seven lectures, each involving a small collaborative group task, and one extensive collaborative course assignment where the groups were supposed to create a midterm plan for primary school dealing with a specific math topic. Both parts of the course were carried out simultaneously during two months in a class-like laboratory space, which made it possible to record all of the collaborative group work with a 360 degree video camera system. The students collaborated in groups of four students resulting in 11 groups. All together the data collection produced 88 hours of video recordings representing 41 videotaped collaborative group work situations. At the beginning of each collaborative learning session, the students used S-REG tool. S-REG is a visual iPad application focusing on group members' awareness of their cognitive, motivational and emotional states. Specifically S-REG promoted a) awareness of SSRL b) explication of SSRL and c) prompting strategies for SSRL.

The analysis of the video recordings proceeded by first identifying segments that included traceable task-focused cognitive and socio-emotional interaction (cf. Rogat & Linnenbrink Garcia, 2011). Second, these social interaction segments were classified to indicate situations that potentially call for individual and social level regulation. That is, the social interaction segments that included cognitive expressions such as disagreement, argumentation and agreement, or emotional or motivational expressions such as irritation, anxiety or lack of motivation were considered to have a possibility to include socially shared regulation of cognition, motivation or emotion. Finally, these selected segments were analyzed in more detail to capture patterns of social interaction through which collaborative groups actualize socially shared regulation.

The analysis of the S-REG tool was conducted in three phases. First, the duration of groups' discussion when using the S-REG tool was measured from each session. Second, the depth of the groups' discussion was rated on a scale from 0 to 1. The group scored 0 if prompts for SSRL were not elaborated and 1 if the prompts of SSRL were elaborated in the group's discussion. A Spearman's correlation coefficient was used to determine whether there is a relationship between the segments including SSRL and the duration and depth of discussion while using the S-REG tool. To capture social interaction patterns through which collaborative groups actualize socially shared regulation State Lag Sequential Analysis (LSA) was conducted. Since the ways groups collaborate is affected by previous experiences, the LSA was conducted only for those segments of the data that involved socially shared regulation of learning. Thus, from those segments each individual turn was coded until the socially shared regulation of learning was actualized.

The results of this study indicate that even though the situation is calling for socially shared regulation, the groups do not always engage in any type of regulation even when it can be considered as a prerequisite for successful collaboration and learning. Also former research has shown evidence indicating that learners do not always optimally regulate their learning process when opportunities arise. There is a clear need to create ways to support groups in their efforts to regulate the learning process together. Former research has shown promising results, when individuals' regulation processes are supported with various technological applications (Johnson et al., 2011). Yet when groups are supported with their SSRL, high performing groups tend to benefit from the support the most (Malmberg et al., 2014). Therefore, it is important to identify whether the S-REG tool can be beneficial for SSRL. It is also important to identify social interaction patterns that have potential to support SSRL and improve the quality and depth of collaborative learning. By recognising such interaction patterns, it is possible not only to promote SSRL via technological tools, but also the interaction patterns that make it happen.

Do learners benefit from socially-regulated learning provided by artificial pedagogical agents? Implications for data analytics in supporting social interactions during complex learning

Roger Azevedo, Nicholas Mudrick, Michelle Taub, Seth A. Martin, and Jesse Farnsworth

Social interactions between humans and artificial pedagogical agents involve a multitude of temporally unfolding self- and other- cognitive, affective, metacognitive, and motivational (CAMM) regulatory processes during learning with advanced learning technologies (Azevedo et al., 2013). However, contemporary trace methodologies and analytical approaches to measuring SRL processes, capturing the real-time deployment of SRL processes (Azevedo & Aleven, 2013; Molenaar & Järvelä, 2014), still pose several challenges that somewhat impede our understanding of the social processes of learning. These challenges include the conceptualization of SRL versus other types of externally-regulated learning (e.g., CoRL, SSRL), embodiment of these different conceptualizations in artificial agents, the temporal alignment of multi-channel data (e.g., affective responses to the artificial agents' prompting and scaffolding, increased arousal based on the complexity of science diagrams, misconceptions revealed during human-artificial agent dialogue, the impact of the external regulation on learners' monitoring of cognitive and affective processes), accuracy of inferences about the impact of artificial agents' social processes on human learning and self-regulatory skills, and determining which multi-channel data (e.g., dialogue moves, utterances, log-files, eye-tracking, physiological indices) should serve as the basis for data analytics to determine the qualitative and quantitative nature of self- and socially shared regulated learning (SSRL). These challenges are fundamental to our community as interdisciplinary researchers in the field of CSCL grapple with (1) the burgeoning landscape of learning theories and models of instruction that focus on social interactions in different authentic contexts (e.g., human–artificial agents and SRL), (2) massive amounts of rich, multi-modal data for data analytics, and (3) the accuracy of inferences about complex social processes stemming from (2). By addressing them, we can make advanced learning technologies more CAMM-sensitive using externally regulating artificial social agents.

Recent advances in the study of self-regulated learning processes as *events* that temporally unfold in real time during learning and problem solving are transforming the fields of metacognition and self-regulated learning. New methods for detecting, tracking, collecting, and analyzing SRL data as events that have specific non-static attributes, such as frequency of use, duration, time-dependent patterns of use, and dynamics, including feedback mechanisms, offer novel ways to examine and understand the role of these processes across learning contexts, age groups, tasks, learning activities, etc. (Azevedo et al. 2010; 2013). These novel methods can reveal important patterns of SRL events, based on the use of various types of multi-modal data (e.g., eye-tracking, facial expression, utterances, conversational turns, log-files), that can significantly enhance our current understanding of the sequential and temporal nature of self- and socially-regulated learning (Azevedo, 2014; Hadwin & Järvelä, 2013; Molenaar & Järvelä, 2014). Therefore, these new methods, despite being exploratory in nature, have the potential to transform current conceptions of SRL by augmenting our models and theories of SRL by delineating micro-level processes (e.g., specific metacognitive processes, such as judgments of learning [JOL]) and contributing to existing theories and models that are either too abstract or focus on macro-level processes (e.g., monitoring), and by generating testable hypotheses based on the types of process data used and the evidenced results (Winne & Azevedo 2014; Zimmerman 2008). The focus of our paper is to present the issues and challenges associated with capturing, analyzing, and inferring CAMM SRL processes during learner-artificial pedagogical agent dialogue during learning with an intelligent tutoring system and the challenges they pose for learning analytics.

A study was conducted using 150 college students who took part in a 2-day experiment with MetaTutor to learn about the human circulatory system. Participants were randomly assigned to either the adaptive or non-adaptive condition. In the adaptive condition, participants were prompted to use several key SRL processes during their learning (e.g., activating relevant prior knowledge, assessing their emerging understanding [JOL], using effective learning strategies) by the pedagogical agents (PAs) embedded in MetaTutor. While those in the non-adaptive condition had the opportunity to engage in these processes, they were not instructed to do so by the PAs. During the 2-hour lesson session with MetaTutor, we collected the following data from each participant:

eye-tracking, video recording of the face (for affect detection and classification), log-files (e.g., quiz results, summaries and metacognitive judgments, learn-agent dialogue), notes and drawings, and physiological data (e.g., electro-dermal activity). We also collected pretest and posttest data and several self-report measures on emotions, motivation, agent likeability and metacognitive knowledge about specific SRL processes. Our results will focus on describing, using multiple-level trace (process) data, participants' self-regulatory behaviors and how they are related to learning outcomes. For example, *micro-level data* provides information on: (1) fluctuations in affective states (e.g., frequencies, duration, and transitions of basic and learning-centered emotions), (2) eye-tracking processes (e.g., fixations and gaze behaviors on specific areas of interest [AOIs] such as the pedagogical agents, SRL palette, multimedia content, learning goals), and (3) log-file data, which details the duration and sequencing of specific behaviors (e.g., frequency and time spent on relevant vs. irrelevant pages and diagrams). *Mid-level data* (1) represents learners' accuracy in making metacognitive judgments (related to calibration and overconfidence in mastery of multimedia content related to a particular learning goal); (2) provides information on the deployment of cognitive and metacognitive processes based on the frequency use of the SRL palette (embedded in the system interface); (3) illustrates their emotion generation and regulation during different sub-goals; (4) provides information on their regulatory processes associated with adaptive (and non-adaptive) changes during the learning session; (5) reveals their knowledge integration across representations of information; (6) exemplifies changes in their self-regulatory processes based on learner-agent dialogue moves; and, (7) provides evidence of how the deployment of CAMM processes is associated with knowledge construction activities (e.g., taking notes, summarizing) and is predictive of quiz scores. *Macro-level data* provides information on changes in students' learning based on their pretest-posttest scores. Data sources and analyses presented in this paper will provide evidence that has the potential to address challenges in learning analytics (e.g., which data, timing of inferences, within- and between- channel aids to understanding SSRL).

Discourse analytics to support persistent participation in MOOCs

Carolyn Penstein Rosé, Miaomiao Wen, and Diyi Yang

Recent research in the field of CSCL has produced technology for automating analysis of collaborative processes in real time (Rosé et al., 2008; Gweon et al., 2013) and using this analysis to trigger in-process support that increased the effectiveness of collaboration and student learning (Kumar & Rosé, 2011; Dyke et al., 2013; Adamson et al., 2014). The time is now ripe address a more challenging problem. With the rise of massive open online courses (MOOCs), we have an opportunity to extend this technology for the purpose of supporting collaborative interactions that could create thriving online learning communities to create a learning experience that increases learner autonomy (Cotteral, 2000), motivation and goal-setting (Pintrich, 2000), as well as self-regulation (Zimmerman, 2008). We are exploring a new form of automated, just-in-time support for effective online learning, powered through analytics applied to data from discussion forum posts.

One great overarching challenge is to create a form of MOOC environment that effectively fosters community connections that provide the type of socially supportive environment to sustain the motivation of students to persist with instruction (Yang et al., 2014a). Even case studies of particularly dedicated MOOC instructors who work hard to keep up with needs as they emerge during the threaded discussions in MOOC environments ultimately discover that support needs far outnumber resources instructors are able to offer (Rosé et al., 2015). In our recent work, machine learning has been used in a MOOC context to identify factors displayed through linguistic choices encoded in MOOC discussion forum posts as a way of identifying students during times when they are particularly vulnerable to dropout. In this way, the hope is that scarce human resources could be channeled to where they are most needed, or augmented with automated forms of just-in-time support, that might enable students to persist in the course through times of elevated vulnerability. In all of this work, we have started with observations of attitudes, orientations, and dispositions that are visible in discussion forum posts and that are associated with learning or persistence in prior literature. We validated hypotheses about what factors would ultimately flag students at risk by utilizing a statistical analysis technique referred to as survival analysis, which has been used to gauge the impact of time variant factors on dropout in other types of online communities (Wang et al., 2013). Survival models are able to quantify the extent to which fluctuations in time variant factors predict relative probability of dropout at specific time points within a user's participation trajectory. Factors we have had success modeling through discourse analytics, which have been validated as significant predictors of dropout using survival modeling include confusion and disinterest (Yang et al., 2014a), motivation and cognitive engagement (Wen et al., 2014a), student attitudes towards course affordances and tools (Wen et al., 2014b), satisfaction with help received, and relationship formation and loss (Yang et al., 2014b). Of all of the factors explored so far, the most dramatic impact on attrition was related to relationship formation and relationship loss in the MOOC discussion forums, even though the students who participate in those forums are known to be among the most highly committed to the course to begin with. In these results we find support for the importance of community, and evidence of the potential positive impact of work towards greater integration in the community, and engagement in joint meaning making towards deeper engagement with the course materials.

References

Adamson, D., et al. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs, *International Journal of AI in Education, 24*(1), 91-121.

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning. *Metacognition and Learning*, *9*, 217–228.

Azevedo, R., et al., (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427–449). Amsterdam, The Netherlands: Springer.

Azevedo, R., et al. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: issues and challenges. *Educational Psychologist, 45*, 210-223.

Bannert, M., et al. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, *9*, 161–185.

Clark, D. B., et al. (2009). Initial structuring of online discussions to improve learning and argumentation: Incorporating students' own explanations as seed comments versus an augmented-preset approach to seeding discussions. *Journal of Science Education and Technology*, *18*, 321-333.

Clow, D. (2012). The learning analytics cycle: closing the loop effectively. In S. Buckingham Shum, D. Gasevic & R. Ferguson (Eds*.) Proceedings of LAK 2012* (pp. 134-138). Vancouver, Canada: ACM.

Cotterall, S. (2000). Promoting learner autonomy through the curriculum: Principles for designing language courses. *ELT Journal*, *54*(2), 109-117.

De Jong & van Joolingen (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, *68*, 179-201.

Dyke, G., et al. (2013). Enhancing Scientific Reasoning and Discussion with Conversational Agents, *IEEE Transactions on Learning Technologies, 6*(3), 240-247.

Fischer, F., et al. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48*(1), 56-66.

Gweon, G., et al. (2013). Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation, *International Journal of Computer Supported Collaborative Learning, 8*(2), 245-265.

Hadwin, A. F., et al. (2001). Context moderates students ' self-reports about how they study. *Journal of Educational Psychology*, *93*, 477–487.

Haythornthwaite, C., & De Laat, M. (2010). Social networks and learning networks: Using social network perspectives to understand social learning. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell & T. Ryberg (Eds.) *Proceedings of the 7th International Conference on Networked Learning* (pp. 183-190). Lancaster, UK: Lancaster University.

Järvelä, S., et al., (2015) Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development, 63*(1), 125-142.

Johnson, A. M., et al. (2011). The temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, *29*, 471–504.

Koschmann, T. D. (1994). Toward a theory of computer support for collaborative learning. *The Journal of the Learning Sciences*, *3*, 219-225.

Kumar, R. & Rosé, C. P. (2011). Architecture for building conversational agents that support collaborative learning, *IEEE Transactions on Learning, 4*, 21-34.

Malmberg, J., et al. (re-submitted). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior*.

Molenaar, I., & Chiu, M. M. (2014). Dissecting sequences of regulation and cognition: statistical discourse analysis of primary school children's collaborative learning. *Metacognition and Learning*, *9*, 137–160.

Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning, 9*, 75-85.

Molinari, G., et al. (2013). Emotion feedback during computer-mediated collaboration: Effects on self-reported emotions and perceived interaction. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar, S. (Eds.), *Proceedings of CSCL 2013* (pp. 336-343). ISLS.

Mu, J., et al. (2013). How collaboration scripts are internalized: A script theory of guidance perspective. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar, S. (Eds.), *Proceedings of CSCL 2013* (pp. 56-66). Madison WI: ISLS.
Mu, J., et al. (2012). The ACODEA framework: Developing segmentation and classification schemes for fully automatic analysis of online discussions. *International Journal of Computer-Supported Collaborative Learning*, *7*, 285–305.
Mu, J., et al. (2014). Automatic coding of questioning patterns in knowledge building discourse. *Proceedings of ICLS 2014* (pp 333-340). Boulder CO: ISLS.
Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92, 544.
Rosé, C. P., et al. (2015). Supportive technologies for group discussion in MOOCs. *Current Issues in Emerging eLearning, 2*(1), Article 5.
Rosé, C. P., et al. (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*, 237-271.
Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed). *The Cambridge handbook of the learning sciences* (pp 97-118). New York: Cambridge.
Shum, S. B., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society, 15*, 3-26.
Stahl, G., et al., (2006). CSCL: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*. (pp. 409-426). Cambridge, UK: Cambridge University Press.
Tang, K. Y., et al. (2014). Contemporary intellectual structure of CSCL research (2006–2013): A co-citation network analysis with an education focus. *International Journal of Computer-Supported Collaborative Learning, 9*, 335-363.
Verbert, K., et al. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, *57*, 1500-1509.
Volet, S., et al. (2012). Metacognitive regulation in collaborative learning. Interpersonal regulation of learning and motivation: methodological advances (pp. 67–101). In M. Vauras & S. Volet (Eds.) *Interpersonal regulation of learning and motivation: Methodological advances*. EARLI series: Routledge.
Wang, Yi-Chia, et al. (2012). To stay or leave?: The relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of CSCW 2012* (pp. 833-842). Seattle: ACM.
Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, *26*, 506-515.
Wen, M., et al. (2014a). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper et al. (Eds.) *Proceedings of EDM 2014* (pp. 130-137). London, UK: IEDMS.
Wen, M., et al. (2014b). Linguistic reflections of student engagement in Massive Open Online Courses. In *Proceedings of ICWSM 2014,* Ann Arbor, MI: AAAI.
Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Mahwah, NJ: Lawrence Erlbaum.
Winne, P. H., & Hadwin, A. F. (2010). Self-regulated learning and socio-cognitive theory. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed) (pp. 503-508). New York: Elsevier, Oxford.
Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In S. Teasley & A. Pardo (Eds.) *Proceedings LAK 2014* (pp. 203-211). Indianapolis, IN: ACM.
Wise, A. F., et al. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. In D. Suthers & K. Verbert (Eds.) *Proceedings of LAK 2013* (pp. 48-56). Leuven, Belgium: ACM
Wise, A. F., et al. (2014). Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics, 1*(2), 48-71.
Yang, D., et al. (2014a). Towards an integration of text and graph clustering methods as a lens for studying social interaction in MOOCs. *International Review of Research in Open and Distance Learning, 15*(5).
Yang, D., et al. (2014b). Peer influence on attrition in massively open online courses. In J. Stamper et al. (Eds.) *Proceedings of EDM 2014* (pp. 405-406). London, UK: IEDMS.
Zimmerman, B. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183.