Background for a PhD in Learning Analytics at Mediate research group

Increased use of digital technology in learning presents new opportunities to collect data about what the users/learners are doing. In principle, all the activities learners engage in with technology in different learning environments can be collected (e.g. text input, math test results, operation of interactive simulations, web search keywords, pointer movements and clicks), analysed, compared, and submitted to different stakeholders (e.g. learners, teachers, educational institutions). Collection and analysis of data of this type are commonly referred to as learning analytics (LA). A major goal of LA is to use this information to adapt the learning process i.e. to tailor the information presented to each individual or group to their level of proficiency and make it available in tasks, representations, simulations, and animations.

Our research on learning analytics in the Mediate group relies on specific theoretical premises. Theoretical perspectives from the learning sciences (socio-cultural and socio-cognitive) are used to guide the selection of research methods and to inform the instructional design and the design of human-computer interaction. Because of this theoretical orientation, our work on LA departs from purely technical approaches such as educational data mining (EDM) that use data as their sole resource for analysis (Baker & Siemens, 2013).

Siemens (2012, p. 3) defines learning analytics as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs". Given our theoretical perspective the aims of LA are threefold, as we elaborate below.

First, LA is a set of methods to collect information about user activity. To achieve this, learners must use digital environment that makes data collection feasible (i.e. allows data to be collected automatically and by informed consent). Constraining use to one particular application or learning environment, such as interactive math problems, essay writing, or a virtual science lab, limits the type of data to be collected but makes the task more manageable than data collection involving multiple applications and even multiple platforms. Internet based technologies that encourage mass collaboration (e.g. MOOCs) have an advantage of attracting many users and capturing large amounts of data, but pose a new challenge in terms of relevance, complexity, scale and privacy. In either case, using a standalone learning technology or a combination of several ones, will typically require special considerations regarding what type of data to collect and what to omit in the stream of actions and interactions mediated by the technology.

Second, the data collected needs to be processed (analysis). The central concern for

researchers in the learning sciences is to achieve interpretations of ”raw” data that makes sense to human readers, i.e. meaningful data. For example, if the data collected shows that a piece of text has been present on a screen for a certain amount of time, it may imply that it has been read thoroughly, but it could also mean that the reader became bored and engaged in something else. An central challenge in LA is to transform collected data into meaningful inferences and to study to what extent LA engines (sometimes referred to as "machine learning") succeed in this or needs to be supplemented with additional data, and if so, of what kind. At Mediate we are interested in including data at the “meaning level” (i.e. human interpretation) to ensure and enhance the validity of the computers’ interpretation. Another challenge is how usage data collected from one individual can be rendered meaningful across populations (e.g. peer groups) and possibly also across different contexts and cultural differences. A simple example for education is the use of algorithms to suggest courses to take based on relationships, such as study structures performed by more experienced students. One way to deal with these issues and make statistically significant predictions is to use large amounts of data, multiple samplings, and to search for recurrent patterns of relationships of actors and items, and to combine this with qualitative data from observations on how humans engage in interactive meaning-making.

Third, the analysis of data must be reported to the relevant stakeholders (e.g., learners, teachers, school leaders, and policy makers). Arguable the main goal of learning analytics is to provide the learner with what he/she needs at the right point in time to enhance the learning process (formative assessment) and relate to the learning outcome (summative assessment). This phase of the LA process allows for experimentation with adaptive feedback, i.e. feedback that varies based on individual differences in performance. Adaptation of this kind can be accomplished in different ways, from choosing among a fixed set of proficiency/competence levels for the individual learner (known from e.g. gaming environments organized in ascending levels of challenges) and to more refined approaches where the feedback can be based on (1) a dynamic model of what the student knows and computed from his/her learning activity, and (2) a model of how a specific subject domain can be learned (a conceptual model).