Development of actionable insights for regulating students' collaborative writing of scientific texts

Christian Hoffmann[1[0000-0002-0620-3621]], Nadine Mandran[1[0000-0002-8660-3827]], Cédric d'Ham[1[0000-0002-7313-7097]], Sébastien Rebaudo[1] and Mohamed Anis Haddouche[2]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
[2] IMT Atlantique, LABSTICC, 29238 Brest, France
christian.hoffmann@univ-grenoble-alpes.fr

Abstract. We develop indicators for teachers to monitor and regulate students' collaborative writing on a web-based science learning environment. Visualizations of carefully selected indicators are proposed to teachers in order to facilitate the tracking, analysis and management of the students' collaborative work process over time. Our research method is based on a user-centered approach. Via focus groups and interviews, teachers have participated in the design of the indicators and visualizations. This communication presents (a) the mapping from collected data to educational constructs underlying our analytical approach for collaborative writing, (b) indicators and visualizations produced to provide actionable insights to teachers, and (c) lessons learned from our iterative human-centered design process. The results are transferable to other learning environments and design processes.

Keywords: Collaborative learning, Collaborative writing, Learning Analytics Dashboards, User centered design.

1 Introduction and Motivation

Supported by online learning environments (OLE), collaborative writing (CW) of scientific texts is nowadays a frequent task asked to students at high school and university level. The aim of our work, falling within the field of educational collaboration analytics, is to deliver actionable insights [1] to teachers via a learning analytics dashboard (LAD) [2], *i.e.* figures and visualizations that allow tracking and regulating students' group writing in order to improve collaboration and group learning. Many research papers in the field of computer supported collaborative writing (CSCW), *e.g.* [3–5], are situated in the context of well-defined collaboration scripts allowing the collection and combined analysis of various data (data traces, chat messages, in class observations, etc.). By collaboration script we mean here the specification of a sequence of activities structuring the interaction between learners [6]. We address in this paper a more general case: automated characterization of CW based solely on data traces, designated to OLEs that can support a wide variety of collaboration scripts of which little is known when designing the analytics on the platform.

Starting from the general problematic dealing with how to improve students' CW of scientific documents on OLEs, we investigate in this communication the following research question: how to measure the degree of collaboration and communicate it effectively to teachers in a LAD?

2 Theoretical Framework

Collaborative writing

Several authors tried to categorize CW establishing taxonomies of writing strategies and student roles [7, 8]. Onrubia *et al.* [4] observed five different strategies, differentiating in particular between *summative text construction, i.e.* each student adds his text without modifying the text of the others, the result being a juxtaposition of the individual contributions and an *integrative text construction, i.e.* one student proposes an initial version and the other students contribute successively making modifications on the existing. This joins the distinction between cooperative and collaborative work organization. The first is characterized by an explicit division of work between the team members, *i.e.* each student writes a part of the text, the second by a co-construction of the text, *i.e.* all team members take responsibility of the whole text aligning their viewpoints. Students do not necessarily follow one well defined strategy but often a mix of them [9].

Collaboration analytics for LADs

The challenge for designers of LADs is to provide teachers with actionable group insights defined by Jorno & Gynther [1] as "*data that allows a corrective procedure, or feedback loop, established for a set of actions".* Martinez-Maldonado *et al.* [10] elaborated a conceptual model of collaboration analytics where these actionable insights are the main output. They emphasize the role of a clear "*mapping from low-level data to higher-order constructs that are educationally meaningful, and that can be understood by educators and learners"* for the assessment of the validity of collaboration analytics. They proposed a generic five-steps mapping scheme: Data → Derived features → Behavioral markers → Sub-constructs → Higher order constructs.

In order to characterize students' writing strategies (an example of an educational higher order construct), the CSCW literature suggests different concepts. We outline here two of them that we mobilize in our analytics: symmetry of action and territorial functioning. "*Symmetry of action is the extent to which the same range of actions is allowed to each agent"* [11]. This is usually guaranteed in educational OLEs designed for students, but the question remains to what extent users really use their capabilities and are actually symmetric *in* their action. The second construct we mobilize, territorial functioning, that indicates if the authors write in separate document spaces or revisit the text written by others, was discussed in the context of CSCW of academic documents by Larsen-Ledet & Korsgaard [12]. In addition to the chronology of the text's revisions, they paid particular attention to their spatial position in the document. Territorial behaviors of authors have multiple origins, as for example affective and cultural aspects, social norms, but depends also strongly on the particular task design and work organization in the group.

3 Design and Research Method

We work according to the Design Based Research framework following the properties: "*anchored in the field, pragmatic, collaborative, integrative, iterative, flexible, traceability and generalization*" [13]. Indeed, our research is anchored in a real-life context: (i) we develop a web-based learning environment, called LabNbook, designed for supporting learners in the collaborative writing of scientific documents, which is used by more than 3500 students every year [14]; (ii) we work with all the stakeholders for designing the platform and evaluating it. We proceed in an iterative way so that the produced tools evolve throughout the implementation in the platform. The experimental process is fully documented [15].

To construct and evaluate our contributions, we pursue the following research agenda: a) define the indicators and visualizations allowing to characterize CW activities in terms of educational constructs, involving LAD experts and teachers, b) validate the produced artifacts with the users (acceptability, utility, usability) and c) evaluate the impact of the actual use of the artifacts. In this paper, we report results after going through the stages a) and b) several times in an iterative five step process:
Step 1: Construction of indicators and first visualization mockups by the designers of LabNbook who use it themselves in their teaching in experimental sciences.
Step 2: Semi-directive 40 to 60-minute interviews with three teachers (two experienced users of LabNbook and one novice). The exchanges covered (i) teachers' concerns when monitoring the collaborative work of their students, (ii) the usability of the indicators and their understanding, and (iii) the potential utility of the mockups.
Step 3: Production of a second version of the indicators and mockups by researchers.
Step 4: Two focus groups with developers of LADs and teachers using LabNbook. The first focus group (8 participants) was centered on the usability of the indicators, the second (4 participants) on the design and utility of the visualizations.
Step 5: New iteration regarding indicators and mockups. We present in section 4 and 5 the resulting versions at this stage.

4 Analyzing sequential collaborative writing of scientific texts

The field context: a web-based science learning environment
On the LabNbook environment, the teacher can structure the workspace shared by a team of learners according to his learning objectives, *e.g*. the writing of lab notebooks or scientific reports during laboratories, problem-based learning sessions or long-term projects. The scientific output produced by the team of learners, called “report”, is an ensemble of different documents, following the structure provided by the teacher. LabNbook operates in a “locked co-editing” mode [16], *i.e.* students can work simultaneously in the shared workspace but each document composing it can be edited only by one student at the time. Teachers can access learners' productions at any time to be informed of their progress and to send them feedback. For facilitating the monitoring of the learners, the environment provides a LAD for each report. The present work aims

at enhancing the existing LAD with a visualization that help teachers to situate students' writing strategies, *e.g.* to distinguish summative from integrative text construction [4].

Mapping "from clicks to constructs"

In Fig. 1, we present a mapping scheme, inspired from Martinez-Maldonado *et al.* [11], in order to explain our CW analytics on LabNbook. We split the one-dimensional diagram used in [11] into two parts to bring to light two main processes involved in educational collaboration analytics: (i) teachers' diagnosis and (ii) design choices.

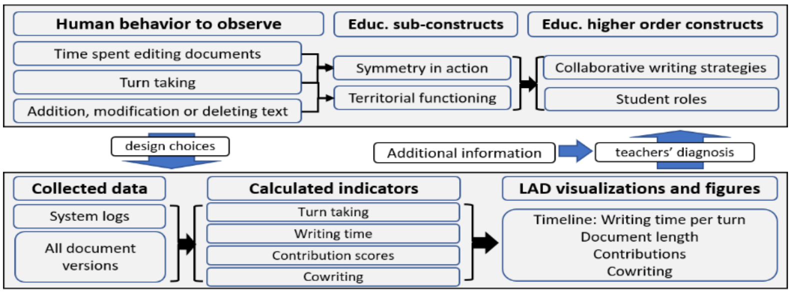

Fig. 1. Mapping of the implemented collaboration analytics

(i) Consistent with the goal of addressing OLEs that support a variety of collaboration scripts, we limit the analysis on LabNbook to the descriptive level and leave the diagnosis to teachers. They can most of the time combine their interpretation of the information given in the LAD with additional information *e.g.* in class observations, exchanges with students, *etc.*, in order to evaluate the ongoing collaboration process in terms of relevant educational constructs. (ii) During the design process, choices are made about which data to collect in order to describe the human behavior, which necessarily leads to an approximate representation.

The lower part of Fig. 1 traces the computer treatment from the collected data via the calculated indicators to the visualizations and figures communicated in the LAD. To capture students' behavior, we collect the following data: who edits a document (authorship), when (timestamp) and a version of the document each time the student validates his contribution.

Indicators to characterize collaborative writing of scientific texts

Our analytics are based on the educational sub-constructs described in section 2.2, symmetry in action and territorial functioning, for which we had to find a translation in terms of computationally calculable indicators.

To evaluate the symmetry in action we construct three indicators, calculated at the level of each document composing a report: (i) turn taking, (ii) writing time and (iii) contribution scores. (i) Turn taking is the number of editor changes. Each time the contributor to the document changes, the indicator is incremented by one. (ii) Writing time is an approximation for the time spent by a student in modifying the document. The system checks for modifications every 30 seconds, so only 30 seconds periods when changes are actually made are added up. The writing time is therefore more significative as the usually measured connection time (timespan between login and logout) which contains a larger fraction of inactivity time. (iii) To convert the iterative text modifications in contribution scores for each student, we use the python library ‘difflib’: the score corresponds to the number of words that the student wrote.

Consistent with automated analytics at a descriptive level, our evaluation of territorial functioning is limited to the observation of the successive authors contributions to the shared document. To this end, we construct a cowriting indicator. The cowriting measures to what extend changes are made by one (or more) author(s) on a text passage produced by another author. The choice of the size of the text passage to consider is not evident. In the actual implementation, we chose to detect cowriting at the level of sentences because we consider them as semantic units where a joint intervention indicates the negotiation of ideas, characteristic of collaboration, in contrast to cooperation. Extending the size of the relevant text passages to paragraphs could be another sound choice. The cowriting score of a document is expressed as a percentage: 0% means that all sentences have been written by a single author; 100% means that all sentences of the document have been written collaboratively.

5 Visualization of the collaborative writing process

Fig. 2 shows the visualization that we designed to track the CW on LabNbook. In the example, a team of 3 students produced a report composed of 8 different documents.

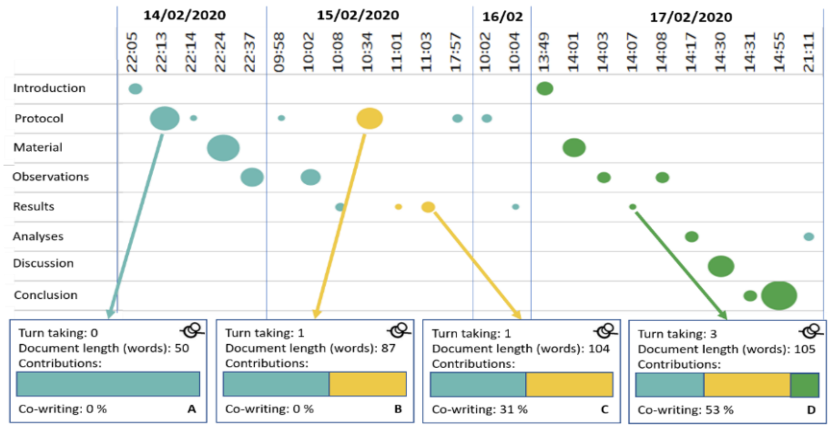

Fig. 2. Mockup for the visualization of CW processes

Each line corresponds to a document with its name in the first column. The x-axis is the timeline and each bubble represents the saving of a new version of the document. We have chosen an equal distance layout (same space between two savings) not a continuous time axis. Three elements of information are given directly in the timeline for each document version: who worked on the document (a different color per student), when it was edited (axis) and the writing time (area of the bubble). With a click on a bubble, the teacher can display additional information in a panel (A à D in Fig. 2) that can be pinned under the timeline: the number of turn takings up to the date, the length of the document (in # of words), the individual contributions (visualized by a stacked bar chart) and the Cowriting score of the document.

Visualization as in Fig. 2 allows a teacher to get a wealth of information about how the report was co-constructed, among others: work duration, work phases, student roles and type of collaboration. We discuss here only the latter, based on the two couples of example panels A/B and C/D. The construction of the "Protocol" (2nd line) is *summative*: student "blue" wrote the beginning (panel A), then student "yellow" added a second part, without revisiting the existing text (Cowriting stays at 0 in panel B). On the contrary, for the "Results" (5th line): student "blue" initiated the document, student "yellow" completed editing some of the existing text (Cowriting at 31% in panel C) and finally student "green" revisited the text, barely adding words (Cowriting increases to 53% in panel D). An *integrative* text construction seems to characterize the writing process of the "Results" document.

6 Lessons learned

Here we report briefly 4 lessons learned from our human centered design process:

1) **Take time for iterations:** the first mockup was created almost a year ago and 3 major iterations have been necessary so far to stabilize the indicators and visualizations. The design process requires time.
2) **Understanding precedes action:** ensure that teachers understand the indicators so that they can take appropriate action. Our experience suggests that teachers need a brief definition and the properties of each indicator while giving the detailed calculation is not necessary.
3) **Be careful with aggregation:** several complementary indicators describing the situation are more appreciated by the teachers than aggregated indicators, which are more difficult to interpret and may prevent action.
4) **Prefer simple visualizations and options:** the teachers in our interviews and focus groups preferred usual at-a-glance visualizations to more sophisticated representations. They asked for opportunities to obtain additional information on demand.

Lessons 3 and 4 confirm similar observations made by Gibson & Martinez-Maldonado [17] and Michos *et al.* [18].

7 Conclusion and Future work

We propose indicators and visualizations that allow teachers to diagnose the CW activities of their students, adapted to all OLEs offering collaborative sequential editing of texts. They make it possible to distinguish different strategies, such as the following common examples in higher education: task sharing when writing team reports on a project (summative text construction); co-construction of a scientific argument (integrative text construction). Future research can concern semantic analysis of the produced texts in order to examine what kind of integrative writing is ongoing. We are also working on a LAD designated for students. Students should have access to information about their CW process, to enhance awareness, reflection and self-regulation.

References

1. Jørnø, R.L., Gynther, K.: What Constitutes an 'Actionable Insight' in Learning Analytics? Learning Analytics. 5, (2018)
2. Sedrakyan, G., Mannens, E., Verbert, K.: Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. Journal of Computer Languages. 50, 19-38 (2019)
3. Olson, J.S., Wang, D., Olson, G.M., Zhang, J.: How People Write Together Now: Beginning the Investigation with Advanced Undergraduates in a Project Course. ACM Trans. Comput.-Hum. Interact. 24, 4:1-4:40 (2017)
4. Onrubia, J., Engel, A.: Strategies for collaborative writing and phases of knowledge construction in CSCL environments. Computers & Education. 53, 1256-1265 (2009)
5. Sundgren, M., Jaldemark, J.: Visualizing online collaborative writing strategies in higher education group assignments. IJILT. 37, 351-373 (2020)
6. Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., Fischer, F.: Specifying computer-supported collaboration scripts. Computer Supported Learning. 2, 211-224 (2007)
7. Posner, I.R., Baecker, R.M.: How people write together (groupware). In: Proceedings of the 25th Hawaii International Conference on System Sciences. p. 127-138. IEEE (1992)
8. Lowry, P.B., Curtis, A., Lowry, M.R.: Building a Taxonomy and Nomenclature of Collaborative Writing to Improve Interdisciplinary Research and Practice. The Journal of Business Communication. 41, 66-99 (2004)
9. Limbu, L., Markauskaite, L.: How do learners experience joint writing: University students' conceptions of online collaborative writing tasks and environments. Computers & Education. 82, 393-408 (2015)
10. Martinez-Maldonado, R., Gaševic, D., Echeverria, V., Fernandez Nieto, G., Swiecki, Z., Buckingham Shum, S.: What Do You Mean by Collaboration Analytics? A Conceptual Model. Journal of Learning Analytics. 8, 126-153 (2021)
11. Dillenbourg, P.: What do you mean by collaborative learning? In: Dillenbourg (éd.) Collaborative-learning: Cognitive and Computational Approaches. p.1-19. Elsevier, Oxford (1999)
12. Larsen-Ledet, I., Korsgaard, H.: Territorial Functioning in Collaborative Writing. Comput Supported Coop Work. 28, 391-433 (2019)
13. Wang, F., Hannafin, M.J.: Design-based research and technology-enhanced learning environments. ETR&D. 53, 5-23 (2005).
14. d'Ham, C., Wajeman, C., Girault, I., Marzin Janvier, P.: Transposition des démarches expérimentales dans un environnement numérique de support. In: Actes des 11e rencontres scientifiques de l'ARDiST. p. 705-716. , Bruxelles, Belgique (2020)
15. Mandran, N.: Traceable human experiment design research. ISTE Ltd / John Wiley and Sons Inc, Hoboken, NJ (2017)
16. Wang, D., Tan, H., Lu, T.: Why Users Do Not Want to Write Together When They Are Writing Together: Users' Rationales for Today's Collaborative Writing Practices. Proc. ACM Hum.-Comput. Interact. 1, 107:1-107:18 (2017)
17. Gibson, A., Martinez-Maldonado, R.: That dashboard looks nice, but what does it mean?: towards making meaning explicit in learning analytics design. In: Proceedings of the 29th Australian Conference on CHI. p. 528-532. ACM, Brisbane Queensland Australia (2017)
18. Michos, K., Lang, C., Hernández-Leo, D., Price-Dennis, D.: Involving teachers in learning analytics design: lessons learned from two case studies. In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge. p. 94-99. Association for Computing Machinery, New York, NY, USA (2020)