

How to Aggregate Lesson Observation Data into Learning Analytics Dataset?

Maka Eradze ^[1], María Jesús Rodríguez-Triana ^{[1][2]} and Mart Laanpere ^[1]

¹ Tallinn University, Tallinn, Estonia

² Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
maka@tlu.ee, mjrt@tlu.ee, mart.laanpere@tlu.ee

Abstract. The technological environment that supports the learning process tends to be the main data source for Learning Analytics. However, this trend leaves out those parts of the learning process that are not computer-mediated. To overcome this problem, involving additional data gathering techniques such as ambient sensors, audio and video recordings, or even observations could enrich datasets. This paper focuses on how the data extracted from the observations can be integrated with data coming from activity tracking, resulting in a multimodal dataset. The paper identifies the need for theoretical and pedagogical semantics in multimodal learning analytics, and examines the xAPI potential for the multimodal data gathering and aggregation. Finally, we propose an approach for pedagogy-driven observational data identification. As a proof of concept, we have applied the approach in two research works where observations had been used to enrich or triangulate the results obtained for traditional data sources. Through these examples, we illustrate some of the challenges that multimodal dataset may present when including observational data.

Keywords: Multimodal learning analytics, learning sciences, classroom observation,

1 Introduction

Learning analytics (LA) is an interdisciplinary field mainly based on data coming from digital traces and digital realms. In order to understand and optimize the learning process, researchers pay especial attention to what is happening in computer-mediated contexts. However, the evidence gathered might be incomplete in real-world learning activities where there face-to-face and digital spaces are frequently combined [1], [2]. Multimodal learning analytics (MMLA) may be a promising approach for this kind of contexts, since researchers in this area are trying to identify and collect also real-world learning data [1]. In addition to the data sources compiled by Blikstein & Worsley in their state of the art [9] (such as speech signals, text-based and graphic-based content, or gestures), we argue that classroom observations of real world teaching and learning processes could be a relevant data input. Moreover, observations that capture

teacher pedagogical intentions are highly relevant information that can become a core of the analysis.

Research has shown that triangulating pedagogically grounded LA with teachers' observational data can be effectively used for teacher orchestration and research purposes [3]. Although there are multiple tools that support the observation process like Kobo Toolbox¹, FieldNotes², Ethos³, Followthehashtag⁴, Storify⁵, and VideoAnt⁶, to the best of our knowledge, there is no one that enables the integration of the observations with other data sources for later analysis.

From the theoretical perspective, there is a need for frameworks that take into consideration the pedagogical semantics in the data collection, integration and analysis. In addition, from the technical point of view, questions remain open about how to model, collect, and integrate the evidence when heterogeneous data sources used. Therefore, we hypothesize that the LA community will benefit from having an integrated solution that aligns pedagogical semantics with xAPI statements.

This paper proposes, first, pedagogy-aware observational data identification approach. To assess its validity, we have chosen existing research that used observations in combination with LA for different purposes. In order to verify whether the approach could be suitable for these cases, we have applied the approach to the observations of such works. Through this proof of concept, we have identified a set of challenges to be overcome when integrating observations with other LA data.

2 Related Work

Learning Analytics and educational Action Research are two research areas with similar goals (while the former uses educational data to foster learning, the later aims to improve the teaching practice), but different methods (LA draws from automatically collected data, and Action Research from observations) [5]. Thus, the combination of both could contribute to improvement of LA research and practice [6], e.g., by mitigating the lack of proper theoretical and pedagogical foundations of existing LA solutions [4].

The alignment between LA and Action Research entails the integration of observations as part of data sources used in the analysis. This step could have a clear impact on the analytics accuracy and representativeness. In most of the cases, part of the teaching and learning processes are not supported by technology. As demonstrated by some authors [12], enriching the datasets with observational data could contribute to

¹ <http://www.kobotoolbox.org>

² <http://fieldnotesapp.info>

³ <https://beta2.ethosapp.com>

⁴ <http://www.followthehashtag.com>

⁵ <https://storify.com>

⁶ <https://ant.umn.edu>

obtain a more realistic view of the educational scenario. However, the implementation of such enrichment is not trivial at different levels:

- **Data gathering:** The lack of guidance in classroom observation applications leads to unstructured and pedagogically neutral data with no consistent format [13].
- **Data integration:** Most of the LA solutions involve a limited number and variety of data sources [2] [7] [8], mainly due to the heterogeneity of data models, formats and granularity [10].
- **Data analysis:** The process of manual coding usually followed by the analysis of the observations is time-consuming and ineffective [13].

In the following section, we propose an approach that tackles the aforementioned problems from a theoretical point of view. Afterwards, the approach is applied to two research studies in order to verify whether it could support the data gathering, collection and integration of the observational data.

3 Theoretical Inquiry: Towards a Solution

Three dimensions were taken into consideration in the design of our approach, namely:

- The philosophical and research approach that frames the purpose of the LA study;
- The educational theory and the pedagogical background that sustains the learning scenario;
- The technological and architectural aspects that condition the data gathering and integration of multiple and heterogeneous data sources;

This section introduces each dimension, reflecting on those areas when the different dimensions overlap. Afterwards, we describe how this approach affects the data gathering, integration and analysis.

3.1 The Approach

Philosophical approach. Current data gathering and analysis proposals can be classified in two main coarse-grained sets. Data-driven generate indicators in a bottom-up fashion, based on available data. Conversely, model-driven approaches need pre-specified models that guide the data gathering and analysis in a top-down process. No matter which approach is followed, the selection and definition of the unit of analysis plays an essential role. Indeed, the **unit of analysis** is used a critical instrument to dismiss one approach or another [14]. Since the unit of analysis has to also be manageable [15] and appropriate for its purpose [14], it is therefore important to have a consistent unit of analysis for multimodal learning analytics.

Technological context. Research in this field has suggested that it is possible to organize several heterogeneous data sources in the form of the xAPI statements and

analyze them with a specific framework in mind [11]. xAPI has a logic and syntax - *actor-verb-object*- that closely follows grammatical categories of most of languages as *subject-verb-object* (in a context).

Educational Theory. In our research, from philosophical point of view, we follow a constructivist approach. Thus, the goal is to enable learners to become actively engaged constructors of their own experience and knowledge. This motivation triggers our interest for understanding the learning activity. In order to track constructivist learning activities, xAPI is ideally suited [18]. While *actor* and *verb* concepts are straightforward in xAPI statements, the *object* has led some researchers to think that is necessarily a Vygotskian *activity system* [18][19] unit. In fact, this sentence-like specification is quite neutral in its essence, since the object is simply an object and not an “object of activity” [20] as claimed previously. This does not mean that, if we want to use activity theory for data collection and analysis, the object cannot become an “object of activity”. This leads us to argue that xAPI statements are not pedagogically biased. Indeed, they can be used to aggregate data with different semantics that are aligned with the pedagogical intentions.

Figure 1 shows how the three different dimensions of our approach intersect. The **unit of analysis** is modeled to be pedagogically neutral, semantically open (vocabulary is interchangeable), and system-independent. In this way, this unit of analysis will allow us collect data with different pedagogical semantics and integrate it, later on, with other data sources. In this approach, the learning event is the unit of analysis [16], which is expressed using xAPI statements.

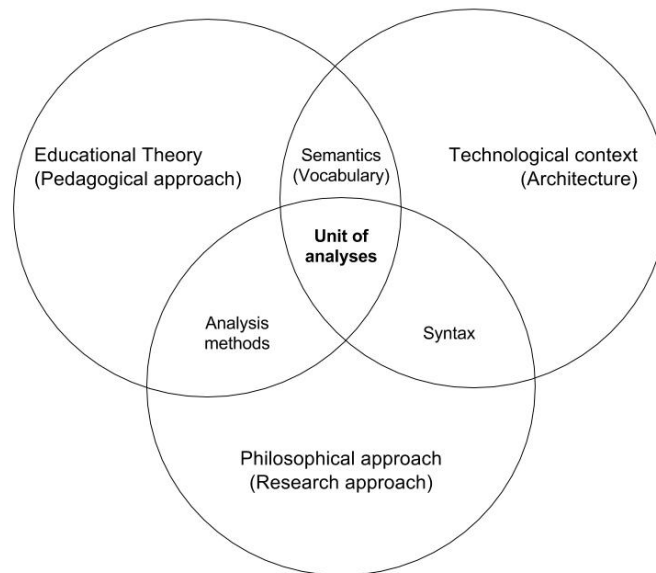


Fig. 1. The approach explained

3.2 The Approach in Action

The observation process is supposed to be carried out by an ad-hoc observer or any participant of the scenario, especially the teachers. The process will be supported by a classroom observation application implemented according to the approach presented in the previous section.

To better understand how the approach would be applied, we describe it through the steps of the common protocol that guides the observation process:

Step 1. Be aware of the elements that belong to the learning context. To facilitate the data gathering (seen as an observer's task) and to enable the integration, it will be necessary to register in all the actors and objects in advance. In that way, the observer will be able to link the events to the corresponding actors and objects. A first implementation challenge will be to know in advance not only about the actors and objects but also to extract the corresponding identifiers which are necessary for later integration and analysis across data sources. To solve this issue, some authors proposed to use the learning design and its instantiation in the technological environment as description of the context [12]. However, this solution is not flexible enough for learning scenarios where new participants or objects may emerge during the activities.

Step 2. Define the areas of focus, the indicators to be obtained in order to illuminate such areas, and the specific events to be observed. We should not forget that we envision the observations as part of a multimodal dataset. Thus, it will be necessary to define, as a whole, how the different areas of interest are informed by the data sources available, and the trackable events. In the case of the observations, the application will be loaded with the vocabulary necessary to describe the events (xAPI verbs).

Step 3. Collect observable events. In this case, the observations will be recorded following the *subject-verb-object* structure, using the set of previously loaded subjects, verbs, and objects. These events will be presented as xAPI statements that will be timestamped and sent to a learning record storage together with the rest of the multimodal dataset. It should be noted that a first study was already carried out to ensure the whether it was feasible to register the observations following the aforementioned format [13].

Step 4. Analyze and interpret the results. The observations will be analyzed with the rest of the events tracked by the complementary data sources, extracting the indicators previously chosen for the different areas of focus.

To better support the integration with other data sources we expect to explore the definition of vocabularies and xAPI Recipes that help us to take also into account the context as suggested by Bakharia et al. [19]. Recipes are set of rules that govern how to use xAPI so that we can ensure, first consistent data to describe similar activities from different sources, and second interoperability across systems.

4 Proof of Concept

To illustrate the potential of our proposal, we have identified 2 research papers that make use of both observations and LA. This section provides a proof of concept of how such research could benefit from an application that implements the approach described in section 3.

Case 1: The first paper describes a study where the teacher reflects on the aspects to be evaluated in a learning scenario, selects the data sources that are relevant for each aspect, and finally chooses the events to be used in the LA process [2]. As part of the data sources, the teacher decided to include her own classroom observational data. The events registered by the teacher were specified in advance and covered: the students who attended the face-to-face sessions (which were mapped with activities), the students who had submitted the productions associated to each activity. The teacher registered the events manually using Google Spreadsheets and ad-hoc solution had to be implemented to retrieve the evidence, translate it into a machine-readable format, and integrate it with the rest of the data sources.

Case 2: The second research paper applied a multiple data gathering techniques for triangulation in a face-to-face course supported by technology (observations, questionnaires, logs, and learning outcomes in the form of text) [21]. An observer attended the course in order to register the face-to-face interaction. Concretely, the observer registered the communication process, indicating the speaker, the kind of action (e.g., lecture, question, answer) and the target audience.

In both cases, the processes followed and the unit of analysis is compliant with the proposal presented in this paper. Thus, the envisioned application could have contributed to automatize and simplify the data gathering and integration processes.

5 Conclusions and Future Work

In this paper, we have discussed the importance of observational data inclusion into MMLA dataset. Based on the literature review, we have proposed an approach and an observational data aggregation solution. The suggested approach is an integrated view that answers to challenges such as standards (xAPI), pedagogy (semantics) and data source (real world data). Based on the proof of concept, we envision that the presented approach could be suitable for pedagogy-aware real-world, observational data identification, and it could serve a basis for development of observational data collection solution in a form of classroom observation app.

In our future work, both the approach and the architecture will establish the basis of the conceptual model/design of an app that will support the structured data gathering during the observation process, and enable xAPI compliant data export for its integration with other data sources. Design-based research methodology will be applied using scenario-based participatory design sessions that are aimed to validate the presented approach and the conceptual model of the app.

References

1. Ochoa, X., & Worsley, M. (2016). Editorial: Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics*, 3(2), 213-219.
2. Rodríguez-Triana, M. J., Prieto Santos, L. P., Vozniuk, A., Shirvani Boroujeni, M., Schwendimann, B. A., Holzer, A. C., & Gillet, D. (In press) *Monitoring, Awareness and Reflection in Blended Technology Enhanced Learning: a Systematic Review. International Journal of Technology Enhanced Learning*.
3. Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2013). Towards a script-aware monitoring process of computer-supported collaborative learning scenarios. *International Journal of Technology Enhanced Learning*, 5(2), 151-167.
4. Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64-71.
5. Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6), 318-331.
6. Dyckhoff, A., Lukarov, V., Muslim, A., Chatti, M. A.; Schroeder, U. (2013, April). Supporting action research with learning analytics. In: Proceedings of the Third International Conference on Learning Analytics & Knowledge. ACM New York, NY, USA, pp. 220-229
7. Nistor, N., Dertnl, M., Klamma, R. (2015). Learning Analytics: Trends and Issues of the Empirical Research of the Years 2011-2014. Design for Teaching and Learning in a Networked World. EC-TEL 2015. Lecture Notes in Computer Science, vol. 9307, pp. 453-459.
8. Schwendimann, B., Rodríguez-Triana, M., Vozniuk, A., Prieto, L., Boroujeni, M., Holzer, A. Gillet, D., Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, X(c), 1-1
9. Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220-238.
10. Ferguson, R. (2012) Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304-317.
11. Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., & Specht, M. (2016, April). Learning Pulse: Using Wearable Biosensors and Learning Analytics to Investigate and Predict Learning Success in Self-regulated Learning. In *CrossLAK* (pp. 34-39).

12. Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2013). Towards a script-aware monitoring process of computer-supported collaborative learning scenarios. *International Journal of Technology Enhanced Learning*, 5(2), 151-167.
13. Eradze, M., Väljataga, T., & Laanpere, M. (2014, August). Observing the use of e-textbooks in the classroom: towards “Offline” Learning Analytics. In *International Conference on Web-Based Learning* (pp. 254-263). Springer International Publishing.
14. Matusov, E. (2007). In Search of the Appropriate Unit of Analysis for Sociocultural Research. *Culture & Psychology*, 13(3), 307-333.
15. Lefstein, A., Snell, J., & Israeli, M. (2015). From moves to sequences: Expanding the unit of analysis in the study of classroom discourse. *British Educational Research Journal*, 41(5), 866-885.
16. Suthers, D., & Rosen, D. (2011, February). A unified framework for multi-level analysis of distributed learning. In *Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 64-74). ACM.
17. Vygotsky, L. S. (1987). *The collected works of LS Vygotsky: Volume 1: Thinking and speech* (N. Minick, Trans. Vol. 1).
18. Kevan, J. M., & Ryan, P. R. (2016). Experience API: Flexible, decentralized and activity-centric data collection. *Technology, Knowledge and Learning*, 21(1), 143-149.
19. Bakharia, A., Kitto, K., Pardo, A., Gašević, D., & Dawson, S. (2016, April). Recipe for success: lessons learnt from using xAPI within the connected learning analytics toolkit. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 378-382). ACM.
20. Kaptelinin, V. (2005). The object of activity: Making sense of the sense-maker. *Mind, Culture, and Activity*, 12(1), 4-18.
21. Rodríguez-Triana, M. J., Holzer, A., Prieto, L. P., & Gillet, D. (2016, September). Examining the effects of social media in co-located classrooms: A case study based on SpeakUp. In *European Conference on Technology Enhanced Learning* (pp. 247-262). Springer International Publishing.