# Observational Scaffolding for Learning Analytics: A Methodological Proposal

Jairo Rodríguez-Medina[1(✉)], María Jesús Rodríguez-Triana[2,3], Maka Eradze[2], and Sara García-Sastre[4]

[1] Center for Transdisciplinary Research in Education, University of Valladolid, Valladolid, Spain
jairo.rodriguez.medina@uva.es
[2] School of Digital Technologies, Tallinn University, Tallinn, Estonia
[3] School of Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[4] GSIC-EMIC Research Group, University of Valladolid, Valladolid, Spain

**Abstract.** Temporal analysis of learning data is attracting the interest of researchers, and a growing body of Learning Analytics (LA) research applies lag sequential analysis. However, lack of methodological frameworks that guide the data gathering and analysis poses multiple conceptual, methodological, analytical and technical challenges. While observation as a technique has been already used in LA, systematic observation methods and designs have not been applied so far, and parameters often used in the observational domain (such as order and duration) are still under-researched. In this paper we propose a methodological framework, and illustrate its potential by applying it in the analysis of a Knowledge Forum dataset. Results show the potential of the proposed method to uncover behavioral patterns prospectively (lag +1 to lag +5) or retrospectively (lag −1 to lag −5), and to reduce this information through *polar coordinate analysis*. Moreover, as illustrated in this paper, observational methods offer a rigorous framework for LA datasets, enabling the replicability, validity and reliability of the results.

**Keywords:** Learning Analytics · Observational methodology
Temporal analytics · Lag sequential analysis · Polar coordinate analysis

## 1 Introduction

Observational methods have been used in education for decades. While humans have traditionally mediated observations, current Learning Analytics (LA) solutions provide automatic means to assist the data collection and analysis process. We could consider LA as "modern" data gathering and analysis technique that support the observational process, either by reducing the workload (thanks to the automation of the process) or enriching the datasets with data coming from digital spaces. Thus, the combination of both human and computer-mediated observations could offer a complementary view of educational contexts [1, 2], which is needed when teaching and learning process happen across spaces [3]. Even though LA offer new insights in the educational domain, current


V. Pammer-Schindler et al. (Eds.): EC-TEL 2018, LNCS 11082, pp. 617–621, 2018.
https://doi.org/10.1007/978-3-319-98572-5_58

solutions often lack of systematic methodological frameworks, compromising the replicability, validity and reliability of the results [4]. In this paper, we hypothesize that systematic observational methods [5, 6] could contribute to alleviating these challenges.

As Ochoa et al. [8] mention, three methodological challenges threaten the development of LA: (1) the difficulty of rigorously assessing the research results; (2) the studies are rarely comparable; and (3) sub-optimal methodologies and tools are often applied when better alternatives exist. These issues only recently have come to the foreground of LA research challenges, and evaluation frameworks have been proposed to assess the users' subjective impression of using an LA system [9] and to provide a more diagnostic view of the performance of such systems [10]. Moreover, a few cases involve several data sources for triangulation, and even more rarely, data from physical spaces [3, 11]. Eradze et al. [2] analyze the difficulties of integrating observational records into multimodal datasets, highlighting the need for a systematic procedure that defines the nature of the data and the unit of analysis, so that the observational design and the parameters to be registered are adjusted accordingly. This need could be addressed through indirect observation, a concept recently coined in the area of observational methods and considers s*tudying both verbal behavior and textual material, whether in the form of transcripts or original material produced by the participants in a study* [12]. The approach already applied to both conventional and new forms of communication (WhatsApp, Twitter, blog posts messages) [13] involves the analysis of data generated in physical or digital settings.

## 2 Proof of Concept and Discussion

In this "proof of concept", to assess the added value of applying a systematic observational approach, we have chosen a publicly available dataset previously analyzed by other authors [7]. The dataset contains information about 1101 notes in 50 threads supported by Knowledge Forum (http://www.knowledgeforum.com). Since we work with a predefined dataset, we will tackle the same research question posed by Chen et al. [7] in their data analysis i.e.: "What are the underlying behavioral patterns that could distinguish productive knowledge-building dialogues – dialogues with apparent attempts to advance collective knowledge?". In order to answer the question, we will apply lag sequential analysis to better understand the temporal relations and patterns, while in a slightly different manner. Our proposal for data transformation is novel as it is supported by observational methods devoted to analyze participants' behavior in an authentic setting using ad hoc observation instruments [5]. This scientific procedure is suited for the analysis of social interaction and it temporal evolution [5, 15].

## 3 Data Analysis

*Lag sequential analysis* was used to investigate sequential relationships between discrete behaviors (events) and interactive states. Additionally, we apply *polar coordinate analysis* [16, 14], this technique allows for data reduction by using the Zsum statistic ($Zsum = \Sigma z/\sqrt{n}$), where Z represents the independent values obtained from the adjusted residuals found for the respective lags of −5 to −1 and 1 to 5, with n as the number of lags. To carry out this analysis we used SDIS-GSEQ software package v. 5.1 [6] and HOISAN v. 1.6 [17].

## 4 Results

To illustrate the potential of polar coordinate analysis for the data reduction, we show an example focused on the *Supporting Discussion* (SD) behavioural category. As Fig. 1 depicts, in non-productive (or improvable) there was a significant mutual inhibition between posts coded as Supporting Discussion (SD) and Questioning (Q) (see Quadrant III, radius = 3.74, $p < .01$), while in the productive threads Supporting Discussion (SD) significantly activated Questioning (Q) (see Quadrant IV, radius = 2.57, $p < .05$). Similarly, Supporting Discussion (SD) posts had a significant inhibition on Theorizing (T) in the non-productive threads while no impact was detected on productive threads. Finally, Working with Information (WI) posts significantly activated the focal behaviour (SD) in the productive threads while in the non-productive threads a non-significant inhibition was detected on the focal behaviour (SD).

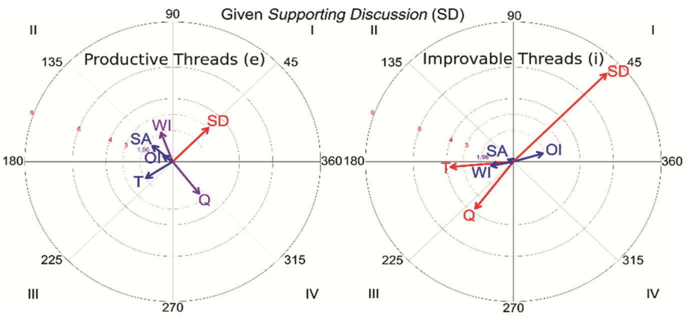


**Fig. 1.** Polar coordinate analysis results for Supporting Discussion (SD) as the focal behavior. Significant relationships between focal and conditional behaviors marked in red ($p < .01$) and purple ($p < .05$) colors (Color figure online)

## 5 Discussion

This paper proposes the application of systematic observational methods [5, 6] as a way to alleviate the methodological and analytical challenges of the Learning Analytic community identified before [8]. The example also provides a proof of concept for the informative potential that polar coordinate analysis may have for data reduction in the field of LA. The application of a rigorous observational design allowed us to uncover behavioral patterns prospectively (lag +1 to lag +5) or retrospectively (lag −1 to lag 5), and to reduce this information through polar coordinate analysis [16]. Thus, this technique may have a remarkable potential in order to interpret the analysis of big datasets, which is common in LA. Moreover, aligned with the open-source movement and the existence of public datasets in LA, the application of open-source software widely adopted by the community of observational methods (e.g., SDIS-GSEQ, HOISAN or THEME) could contribute to address the assessment and comparison challenge reported by Ochoa et al. [8].

## References

1. Macfadyen, L.P., Dawson, S.: Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. J. Educ. Technol. Soc. **15**, 149–163 (2012)
2. Eradze, M., Rodríguez-Triana, M.J., Laanpere, M.: Semantically annotated lesson observation data in learning analytics datasets: a reference model. Interact. Des. Archit. J. **33**, 75–91 (2017)
3. Rodríguez-Triana, M.J., Vozniuk, A., Holzer, A., Gillet, D., Prieto, L.P., Boroujeni, M.S., Schwendimann, B.A.: Monitoring, awareness and reflection in blended technology enhanced learning: a systematic review. Int. J. Technol. Enhanc. Learn. **9**, 126–150 (2017)
4. Blikstein, P., Worsley, M.: Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. J. Learn. Anal. **3**, 220–238 (2016)
5. Anguera, M.T.: Observational methods (General). In: Fernández-Ballesteros, R. (ed.) Encyclopedia of Psychological Assessment, vol. 2, pp. 632–637. Sage, London (2003)
6. Bakeman, R., Quera, V.: Sequential Analysis and Observational Methods for the Behavioral Sciences. Cambridge University Press, Cambridge (2011)
7. Chen, B., Resendes, M., Chai, C.S., Hong, H.Y.: Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. Interact. Learn. Environ. **25**, 162–175 (2017)
8. Ochoa, X., Hershkovitz, A., Wise, A., Knight, S.: Towards a convergent development of learning analytics. J. Learn. Anal. **4**, 1–6 (2017)
9. Scheffel, M., Drachsler, H., Toisoul, C., Ternier, S., Specht, M.: The proof of the pudding: examining validity and reliability of the evaluation framework for learning analytics. In: Lavoué, É., Drachsler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) EC-TEL 2017. LNCS, vol. 10474, pp. 194–208. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66610-5_15
10. Rodríguez-Triana, M.J., Prieto, L.P., Martínez-Monés, A., Asensio-Pérez, J.I., Dimitriadis, Y.: The teacher in the loop: customizing multimodal learning analytics for blended learning. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp. 417–426. ACM (2018)
11. Schwendimann, B.A., Rodriguez-Triana, M.J., Vozniuk, A., Prieto, L.P., Boroujeni, M.S., Holzer, A., Gillet, D., Dillenbourg, P.: Perceiving learning at a glance: a systematic literature review of learning dashboard research. IEEE Trans. Learn. Technol. **10**, 30–41 (2017)
12. Anguera, M.T., Portell, M., Chacón-Moscoso, S.: Indirect observation in everyday contexts: concepts and methodological guidelines within a mixed methods framework. Front. Psychol. **9**, 1–20 (2018). https://doi.org/10.18608/jla.2017.43.1
13. Radzikowski, J., Stefanidis, A., Jacobsen, K.H., Croitoru, A., Crooks, A., Delamater, P.L.: The measles vaccination narrative in Twitter: a quantitative analysis. JMIR Pub. Heal. Surveill. **2** (2016)
14. Sackett, G.P.: Lag sequential analysis as a data reduction technique in social interaction research. Except. Infant. Psychosoc. **4**, 300–340 (1980)
15. Bakeman, R., Gottman, J.M.: Observing Interaction. Cambridge University Press, Cambridge (1997)

16. Anguera, M.T.: From prospective patterns in behavior to joint analysis with a retrospective perspective. In: Colloque sur invitation "Méthodologie d'analyse des interactions sociales" Université de la Sorbonne, Paris (1997)
17. Mendo, A.H., López, J.A.L., Paulis, J.C., Sánchez, V.M., Brincones, J.L.P.: Hoisan 1.2: Programa informático para uso en metodología observacional. Cuad. Psicol. del Deport **12**, 55–78 (2012)