Exploring Idiographic Learning Analytics in Master's Thesis Writing: A Transition Network Approach

Hibiki Ito[1] and Mohammed Saqr[2]

[1] Graduate School of Informatics, Kyoto University, Kyoto, Japan
hibiki.itoo@gmail.com
[2] School of Computing, University of Eastern Finland, Joensuu, Finland
mohammed.saqr@uef.fi

Abstract. Personalising learning environments has been a key research area in learning analytics (LA). Yet, traditional nomothetic, between-person approaches struggle to capture individual learning processes due to learners' heterogeneity. Idiographic, person-specific LA has emerged as a promising approach to address this issue by focusing on a single subject. However, analytical methods used in idiographic studies have been limited, and the potential of idiographic LA has been discussed in limited application areas. To fill these gaps, we employ a recently proposed transition network analysis (TNA) framework to explore idiographic LA by focusing on the context of master's thesis writing. By combining ecological momentary assessment and digital trace data, our proof-of-concept case study demonstrates that TNA can be effectively utilised in idiographic analysis of the master's thesis process.

Keywords: Idiographic · Learning Analytics · Transition Network Analysis

1 Introduction

The integration of digital tools into education has driven a rapid expansion of educational data and computational methods. As a result, personalising learning environments and experiences through big data analytics has been a key research priority in the field of learning analytics (LA) [1, 2]. However, despite the promise of LA, traditional statistical and machine learning approaches that aim to derive group-level insights struggle to scale and derive nuanced understanding of individual learning processes [3–7]. In other words, such one-size-fits-all approaches prove inadequate [8].

To address this problem, idiographic LA [9], or single-case LA [10], has recently been proposed as a promising approach to complement group-level, between-person analytics by focusing on a single learner. Unlike traditional between-person, nomothetic methods, idiographic LA collects data from a single person and aims to derive insights about learning processes specific to that person. Although this person-specific methodology has historically been scarce in the literature of education research, recently growing empirical studies have shown its potential effectiveness [6, 11–16].

Nonetheless, prior idiographic research employed limited, though evolving, variety of analytical methods as well as data collection strategies [11]. Additionally, potential application of idiographic LA has been discussed in limited areas such as collaborative learning [17], doctoral education [18, 19] and inclusive education [12]. Towards filling these gaps, in this paper we focus on the context of master's thesis writing, which is a highly idiosyncratic, heterogeneous educational activity potentially suitable for idiographic LA, yet remains relatively underexplored in the LA literature [20, 21]. We utilise recently proposed transition network analysis (TNA) [22] and combine traditional ecological momentary assessment (EMA) with digital trace data, exploring how these new analytical method and data collection strategy effectively operationalise idiographic LA in the wiring process of master's thesis. To promote reproducibility and the development of educational technologies, the preprocessed datasets used in this study are published under the informed consent of the participant.[1]

2 Background

2.1 Idiographic learning analytics

While nomothetic methods struggle to capture individual processes in LA, and behavioural sciences at large, the idiographic approach is a complementary methodology that focuses on a single person, in contrast to the group-level, nomothetic approach. In particular, idiographic methods employ repeated measures to capture how the individual processes change over time. This is not to be confused with *within-person* analysis that also explores temporal changes within individuals, but aggregates results over a group of subjects [23].

In the field of education, especially in LA, there has been growing empirical evidence that group-level conclusions derived from nomothetic analyses often differ from individual insights obtained through person-specific, idiographic analyses [5–7, 13, 24–26].

A key theoretical framework to understand these results is the theory of ergodicity [27, 28]. In essence, group-level variance of a variable matches intra-individual variance if and only if that variable satisfies two conditions:

[1] See the online supplementary material available at https://hibiki-i.github.io/idiographicTNA.

homogeneity (i.e. it doesn't matter whom to measure) and *stationarity* (i.e. it doesn't matter when to measure). However, educational activities are essentially heterogeneous, and learning process is often dynamic and complex. Therefore, from the perspective of ergodicity, the nomothetic approach inherently struggles to capture individual learning.

Nevertheless, the adoption of idiographic methods in education research has been very scarce [11], with only a few recent examples in LA research [9, 10, 17–19]. In this paper, we aim to extend this line of research by exploring idiographic LA in the context of master's thesis writing.

2.2 Transition network analysis

Transition network analysis (TNA) [22] is a recently proposed methodological framework to understand individual learning processes informed by theories of learning. It builds on stochastic process mining, which explores how different events occur over time, and probabilistic networks, which examines how a process transits different states over time [22, 29]. Unlike other applications of related methods in LA such as stochastic process mining [30] and epistemic network analysis [31], TNA draws on the canonical view that the learning process is marked as a series of transitional phases between learning events or occurrences (i.e. actions that learners take). In addition, TNA enables to explore probabilistic transitions between not only occurrences but also states of learning (e.g. different engagement states), as we demonstrate later in this paper.

2.3 Case study: master's thesis writing

The idiographic approach is particularly valuable in differentiated, idiosyncratic educational contexts where nomothetic methods lack applicability. Thesis writing is an example of such learning contexts which prior works sought to understand and support by the approach of LA [20, 21, 32–36]. That said, these studies mainly focus on the undergraduate level, second-cycle master's level remaining underexplored [36]. Compared to undergraduate education, master's programmes typically involve a research project or thesis that requires more advanced academic, professional knowledge and skills [37]. Consequently, master's thesis projects are inherently more longitudinal and field-specific, requiring far more personalised support. In this paper, we shed light on master's thesis writing process by the idiographic approach of LA.

As idiographic analysis is a special case of within-person analysis, prior research borrow data collection methods from within-person analyses [23]. Nevertheless, majority of recent idiographic LA studies rely on ecological momentary assessment (EMA) and experience sampling method (ESM) [9, 11, 12, 18]. Although these methods effectively capture learning processes as a dynamic complex system, they solely rely on participants' self-reporting measures, imposing additional burden on participants and being subject to missed assessments. Therefore, in our case study, we complement EMA by digital trace data through web tracking and smart watch data. A closely related work includes the idiographic machine learning approach by Aalbers et al. [14] which developed person-specific and between-person machine learning models from smartphone log data to predict self-reported momentary stress. We, in contrast, focus on a wider range of psychological constructs affecting learning processes.

To demonstrate the analytical power of TNA, in this case study we seek to answer the following research questions (RQs) by TAN approach:

- **RQ1**: *How does each EMA feature change within a day and between days?*
- **RQ2**: *What are learning states and how do they unfold within a day and between days?*
- **RQ3**: *How do the learning states unfold within a day and between days according to different daily activity profiles?*

First, RQ1 is particularly relevant when studying a specific learning-related construct. Additionally, investigations like RQ2 would be typical areas of the application of TNA, as the theoretical foundation of TNA is based on sequential transitions of different states. Third, RQ3 extends RQ2 by separately asking the same question as RQ2 based on auxiliary information, namely, daily activity profiles. Overall, we focus on inter- and intra-day changes, aiming to understand both short- and long-term learning processes.

3 Methods

3.1 Context and data collection

The participant (N=1) is a full-time master's degree student in computer science in Finland. Data collection was conducted during the period of 158 days, starting from topic selection phase and ending on the submission day. The student successfully completed the thesis on time, and the thesis was later accepted. As the participant did not have any other course works during this period, we assume that collected data exclusively reflect the process of thesis work.

EMA was implemented to notify the student to answer a set of questions regarding learning experience at that moment in Likert scale one to ten. The notifications were sent five times a day between 9 AM and 8 PM at random so that the participant submits responses immediately after each notification. Additionally, unobtrusive data collection was also implemented through a web tracking app and a smart watch. The former tracked time spent per day on relevant web applications: Overleaf for writing thesis, Notion for planning and organising the project, Paperpile for managing and reading references and ChatGPT for supporting various thesis-related tasks. We selected these apps based on participant's intention to use them for thesis work before starting the thesis project. In addition, the smart watch recorded time spent on exercises, step counts and heart rate as event logs.

In principle, EMA data was collected while the student was engaged in thesis work, and notifications were disregarded during non-working periods. Additionally, the participant was encouraged to wear the smart watch throughout the day on days spent working on thesis, but not on holidays. On the other hand, the web tracking app was active all the time during the entire period, only tracking time spent on relevant apps to respect participant's privacy. The total number of EMA records is 342 over 134 unique days.

3.2 Analysis procedures

The collected data were first processed to form two structured datasets, an EMA dataset and a daily app usage dataset. The former contains five EMA questionnaire variables (Anxiety, Attraction, Commitment, Regulation and Support) as well as smart watch log aggregates (time spent on exercise in the last two hours, the average heart rate in the last two hours and step counts in the last two hours).

For RQ1, each time series of EMA features from the questionnaire is analysed by the `tsn` package in R that is currently being developed[2]. This allows for discretising the time series into a few states based on k-means clustering [38] and easily visualising the series with the clustered states. By setting $k = 3$, we partition each EMA feature into three distinct states, namely, low, average and high, thereby enablying the use of TNA. We then estimate intra-day transition networks using the `tna` package in R [39].

Next, using all EMA features, we classify the records by k-means clustering [38]. By doing so, we identify distinct learning states of the thesis writing process. Subsequently, to answer RQ2, inter- and intra-day transition networks are estimated.

Finally, to answer RQ3, we utilise the web app usage datsaet to classify EMA records by daily activity profiles (i.e. what the student did on the day). Hierachical clustering is performed by Ward's method using the `hclust` package [40]. Finally, we model the transitions of learning states within a day and between days of each activity profiles. See the supplementary material for the detailed procedure and source code.

[2] https://github.com/santikka/tsn, last accessed 2025/08/03

4 Results

4.1 RQ1: Within-feature TNA

Figure 1 shows how, for instance, anxiety and support change over the entier period (i.e. between days) and within a day. In the transition networks, the size of each node is proportionate to the total count of the state, and the outer incomplete circle of each node indicates the probability of that node being the initial state.

For both these features, transitions between the low and high states are relatively scarce, suggesting that participant's learning process rarely makes sudden heaps for these aspects. In addition, it is evident that self-loop probabilities exceed the 0.33 random baseline, implying that the corresponding states possess inherent persistence.

Focusing on each feature, we obtain more specific insights. For example, the time series of support appears to exhibit a certain trend over the period, despite the linear detrending in the preprocessing step. The existence of a non-linear trend would imply that support was not consistent during the thesis project. The strong self-loop on the low state additionally confirms that lack of support likely persists throughout a day.

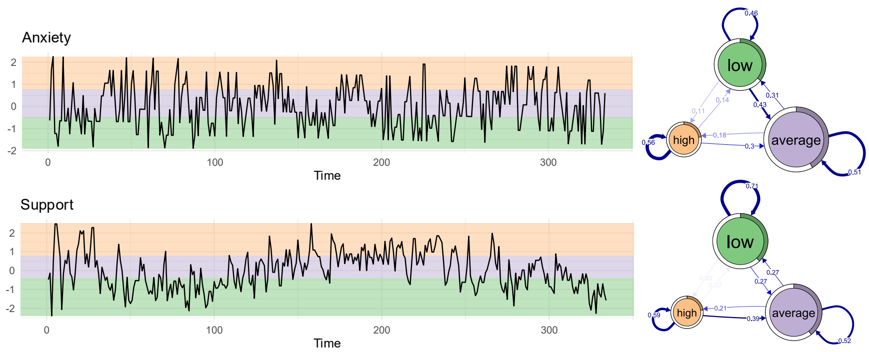

Figure 1. Time series analysis over the entire period (left column) and corresponding intra-day transition networks (right column)

4.2 RQ2: TNA over learning states

Next, we seek to identify distinct learning states and apply TNA on them. After detrending and standardising all EMA features, including the aggregates of smart watch data, we applied k-means clustering with different numbers of clusters. Figure 2 displays the elbow plot, namely, total within-cluster sums of squares for different numbers of clusters. Accordingly, k = 3 turns out to be feasible.

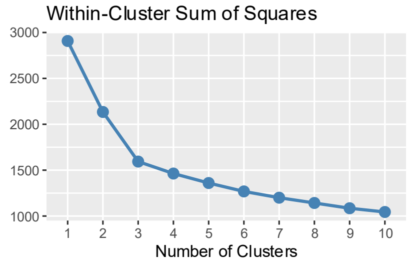

Figure 2. Elbow plot of k-means clustering of EMA records

Now that we have identified learning states, we answer the second part of RQ2 by TNA, namely, how these states unfold over time. Figure 3 shows the estimated inter- and intra-day transitions between the learning states. It is immediately observed that the networks are similar, implying certain fractality of the learning process. That is, transition patterns remain similar when the sizes of time windows are changed. Additionally, the self-loops on the engaged and struggling states are relatively strong, suggesting that these states tend to persist throughout a day. On the other hand, the thickest arrow going out from the active state bridges to the struggling state, indicating the student likely struggles with the thesis project after physical activity.

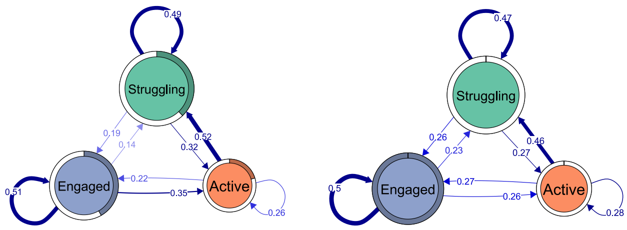

Figure 3. Intra-day (left) and inter-day (right) transition networks of learning states

At first glance, these findigns appear to be actionable insights. For example, since the struggling state often persists, timely interventions could be implemented when it is detected. However, it should be noted that the network is estimated over the entire period, reflecting the *average* transition probabilities. In logitudinal contexts, this aggregate approach could be misleading, as the dynamism of learning processes may differ across distinct phases. We now demonstrate this point by focusing on daily thesis-related activities.

4.3 RQ3: TNA by daily activity profiles

To determine daily activity profiles, we apply hierarchical clustering on the preprocessed web app usage dataset. Figure 4 shows the dendrogram of the clustering result, where three clusters (i.e. profiles) of web app usage are identified. These profiles are characterised as preparing, input- and output-oriented days (see the supplementary material).

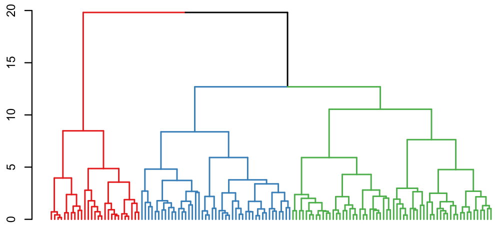

Figure 4. Dendrogram showing the clustering result of web app usage dataset

Partitioning the EMA dataset by these daily activity profiles, we now apply TNA to the distinct subsets of EMA records. Figure 5 displays the intra- and inter-day transition networks for different daily activity profiles. Notably, both the estimated intra- and inter-day networks for preparing days and those for input- and output-oriented days exhibit different patterns, while the latter two are rather similar. For the preparing days, the struggling state indicates a weaker transition to itself, implying that the state is less persistent. This is in contrast to the average transition network shown in the previous subsection, highlighting the need to partition the dataset for more nuanced and precise insights in such a logitudinal study.

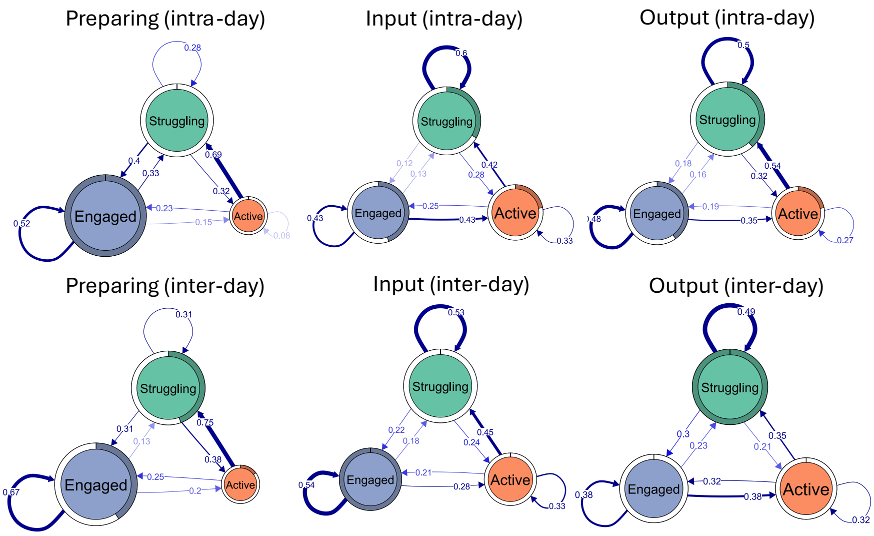

Figure 5. Intra-day and inter-day transition networks for preparing, input-oriented and output-oriented days.

5 Discussion and conclusion

5.1 Discussion

As personalising learning environments is increasingly important, the field of LA has adopted various analytical techniques to utilsie educational big data. As we discussed before, personalisation in highly idiosyncratic and heterogeneous contexts necessiates person-specific, idiographic LA. In particular, we have demonstrated how idiographic LA through the TNA approach could fit to master's thesis writing, where the LA research has been limited.

First, within-feature TNA was conducted, which provided insights on how different aspects of the thesis writing process change over time. Second, we demonstrated how TNA reveals the sequential transitions of learning states within a day and between days. Specifically, we identified engaged, struggling and active states and found that the first two tend to persist on average for both within a day and between days. Third, since the dataset is longitudinal, we partitioned it into three subsets and applied TNA separately. As a result, it was found that the struggling state was not consistently persistent on preparing days in contrast to the average-case observation.

Overall, our methodological case study demonstrates that TNA, through the intergration of EMA and trace data, can effectively capture the dynamic and complex process of master's thesis writing.

5.2 Limitations and future work

Our approach is, nonetheless, subject to several limitations. As noted by prior work [22], TNA relies on first-order Markov models. For example, in the RQ2 analysis, we observed a higher likelihood for the active-to-struggling transition, yet it remains unclear whether this transition always follows an engaged-to-active progression or instead reflects a return to struggling. In addition, repeated measurements through EMA imposes burden on learners and inherently suffers from self-report bias. Although trace data potentially mitigate this issue, formatting data and feature engineering would require analyst's efforts as demonstrated in this paper and be challenging to automate the procedure.

Realising the power of idiographic LA in smart learning environments would require more work on extending the TNA framework and data collection beyond EMA. At the same time, to translate TNA from proof-of-concept to practice, future work should involve technological advancements to operationalise TNA in real-world deployments.

5.3 Conclusion

By complementing the issues inherent in nomothetic LA, the idiographic approach has the potential to truly peronalise learning environments. We explored the TNA framework in the context of master's thesis writing. Our analytical methods and data collection strategies effectively revealed the dynamic complex process of thesis writing. To implement our approach in the real-world deployments, more work is needed to extend the TNA framework and go beyond the EMA data collection.

Acknowledgements. This work was supported by JSPS KAKENHI Grant Number 25KJ1515.

References

1. Romero, C., Ventura, S.: Educational data mining and learning analytics: An updated survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 10, e1355 (2020).
2. Society for Learning Analytics Research (SoLAR): Reimagining Learning Analytics, https://www.solaresearch.org/wp-content/uploads/2025/06/Reimagining-Learning-Analytics-V3-002.pdf, last accessed 2025/07/22.
3. Wilson, A., Watson, C., Thompson, T.L., Drew, V., Doyle, S.: Learning analytics: challenges and limitations. Teach. High. Educ. 22, 991–1007 (2017).
4. Yudelson, M., Fancsali, S., Ritter, S., Berman, S., Nixon, T., Joshi, A.: Better data beats big data. In: Educational data mining 2014. Citeseer, London (2014).
5. Fisher, A.J., Medaglia, J.D., Jeronimus, B.F.: Lack of group-to-individual generalizability is a threat to human subjects research. Proc. Natl. Acad. Sci. U. S. A. 115, E6106–E6115 (2018).
6. Saqr, M.: Group-level analysis of engagement poorly reflects individual students' processes: Why we need idiographic learning analytics. Comput. Human Behav. 150, 107991 (2024).
7. Saqr, M., López-Pernas, S.: Why explainable AI may not be enough: predictions and mispredictions in decision making in education. Smart Learn. Environ. 11, 1–16 (2024).
8. Gašević, D., Dawson, S., Rogers, T., Gasevic, D.: Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. Internet High. Educ. 28, 68–84 (2016).
9. Saqr, M., Lopez-Pernas, S.: Idiographic learning analytics: A definition and a case study. In: 2021 International Conference on Advanced Learning Technologies (ICALT). pp. 163–165. IEEE (2021).
10. Prieto, L.P., Rodríguez-Triana, M.J., Ley, T., Eagan, B.: The value of epistemic network analysis in single-case learning analytics: A case study in lifelong learning.

In: Communications in Computer and Information Science. pp. 202–217. Springer International Publishing, Cham (2021).
11. Ito, H., López-Pemas, S., Saqr, M.: A scoping review of idiographic research in education: Too little, but not too late. In: 2024 IEEE International Conference on Advanced Learning Technologies (ICALT). pp. 10–12. IEEE (2024).
12. Saqr, M., Cheng, R., López-Pernas, S., Beck, E.D.: Idiographic artificial intelligence to explain students' self-regulation: Toward precision education. Learn. Individ. Differ. 114, 102499 (2024).
13. Saqr, M., López-Pernas, S.: Momentary emotions emerge and evolve differently, yet are surprisingly stable within students. In: 2024 IEEE International Conference on Advanced Learning Technologies (ICALT). pp. 86–90. IEEE (2024).
14. Aalbers, G., Hendrickson, A.T., Vanden Abeele, M.M., Keijsers, L.: Smartphone-tracked digital markers of momentary subjective stress in college students: Idiographic machine learning analysis. JMIR MHealth UHealth. 11, e37469 (2023).
15. Peng, H., Lowie, W., Jager, S.: Unravelling the idiosyncrasy and commonality in L2 developmental processes: A time-series clustering methodology. Appl. Linguist. 43, 891–911 (2022).
16. Saqr, M., Vogelsmeier, L.V.D.E., López-Pernas, S.: Capturing where the learning process takes place: A person-specific and person-centered primer. Learn. Individ. Differ. 113, 102492 (2024).
17. Malmberg, J., Saqr, M., Järvenoja, H., Järvelä, S.: How the monitoring events of individual students are associated with phases of regulation: A network analysis approach. J. Learn. Anal. 9, 77–92 (2022).
18. Prieto, L.P., Pishtari, G., Dimitriadis, Y., Rodríguez-Triana, M.J., Ley, T., Odriozola-González, P.: Single-case learning analytics: Feasibility of a human-centered analytics approach to support doctoral education. J. Univers. Comput. Sci. 29, 1033–1068 (2023).
19. Prieto, L.P., Jovanovic, J., Odriozola-González, P., Rodríguez-Triana, M.J., Díaz-Chavarría, H.B., Dimitriadis, Y.: Disentangling doctoral well-being support in progress-focused workshops: Combining qualitative and quantitative data in single-case learning analytics. Learn. Individ. Differ. 121, 102705 (2025).
20. Larsson, K.: Managing Thesis Projects in Higher Education - Through Learning Analytics, https://www.diva-portal.org/smash/get/diva2:1429860/FULLTEXT01, (2020).
21. Nouri, J., Larsson, K., Saqr, M.: Identifying factors for master thesis completion and non-completion through learning analytics and machine learning. In: Lecture Notes in Computer Science. pp. 28–39. Springer International Publishing, Cham (2019).
22. Saqr, M., López-Pernas, S., Törmänen, T., Kaliisa, R., Misiejuk, K., Tikka, S.: Transition network analysis: A novel framework for modeling, visualizing, and identifying the temporal patterns of learners and learning processes. In: Proceedings of the 15th International Learning Analytics and Knowledge Conference. pp. 351–361. ACM, New York, NY, USA (2025).
23. Saqr, M., Ito, H., López-Pernas, S.: Individualized Analytics: Within-Person and Idiographic Analysis. In: Saqr, M. and López-Pernas, S. (eds) Advanced Learning Analytics Methods: AI, Precision and Complexity. p. (in press). Springer (2025).

24. Saqr, M.: Modelling within - person idiographic variance could help explain and individualize learning. Br. J. Educ. Technol. 54, 1077–1094 (2023).
25. Saqr, M., López-Pernas, S.: Changes in online engagement at the within-person level, profiles, dynamics and association with achievement. Internet High. Educ. 67, 101031 (2025).
26. Saqr, M., López-Pernas, S.: Mapping the self in self - regulation using complex dynamic systems approach. Br. J. Educ. Technol. 55, 1376-1397 (2024).
27. Molenaar, P.C.M.: A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. Measurement (Mahwah NJ). 2, 201–218 (2004).
28. Piccirillo, M.L., Rodebaugh, T.L.: Foundations of idiographic methods in psychology and applications for psychotherapy. Clin. Psychol. Rev. 71, 90–100 (2019).
29. Saqr, M., López-Pernas, S., Tikka, S.: Mapping Relational Dynamics with Transition Network Analysis: A Primer and Tutorial. In: Saqr, M. and López-Pernas, S. (eds) Advanced Learning Analytics Methods: AI, Precision and Complexity. Springer (2025 in press).
30. Saint, J., Gašević, D., Matcha, W., Uzir, N.A., Pardo, A.: Combining analytic methods to unlock sequential and temporal patterns of self-regulated learning. In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge. ACM, New York, NY, USA (2020). https://doi.org/10.1145/3375462.3375487.
31. Shaffer, D.W., Hatfield, D., Svarovsky, G.N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A.A., Mislevy, R.: Epistemic network analysis: A prototype for 21st-century assessment of learning. Int. J. Learn. Media. 1, 33–53 (2009).
32. Nouri, J., Larsson, K., Saqr, M.: Bachelor thesis analytics: Using machine learning to predict dropout and identify performance factors. iJAI. 1, 116 (2019).
33. Hansen, P., Hansson, H.: Optimizing student and supervisor interaction during the SciPro thesis process – concepts and design. In: Advances in Web-Based Learning -- ICWL 2015. pp. 245–250. Springer International Publishing, Cham (2015).
34. Rapp, C., Kruse, O.: Thesis writer (TW): Tapping scale effects in academic writing instruction. In: Proceedings of the Third (2016) ACM Conference on Learning @ Scale. pp. 325–326. ACM, New York, NY, USA (2016).
35. Rapp, C., Kauf, P.: Scaling academic writing instruction: Evaluation of a scaffolding tool (thesis writer). Int. J. Artif. Intell. Educ. 28, 590–615 (2018).
36. Law, T.-J., Ting, C.-Y., Ng, H., Goh, H.-N., Albert, Q.: A review of graduate on time prediction. Int. J. Adv. Sci. Eng. Inf. Technol. 14, 1957–1966 (2024).
37. UNESCO Institute for Statistics: International Standard Classification of Education (ISCED) 2011. UNESCO Institute for Statistics (2012).
38. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: A K-means clustering algorithm. J. R. Stat. Soc. Ser. C. Appl. Stat. 28, 100 (1979).
39. Tikka, S., López-Pernas, S., Saqr, M.: Tna: An R package for transition network analysis. Appl. Psychol. Meas. 01466216251348840 (2025).
40. Murtagh, F., Legendre, P.: Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J. Classification. 31, 274–295 (2014).