Pedagogical Data Production as the Generative Foundation of Meaningful Learning Analytics

Satu Aksovaara, Minna Silvennoinen
Jamk University of Applied Sciences, Jyväskylä, Finland
satu.aksovaara@jamk.fi, minna.silvennoinen@jamk.fi

DOI
https://doi.org/10.18690/um.fov.4.2026.6

ISBN
978-961-299-152-4

Learner-centred and distributed higher education increases the need for information that makes learning visible beyond system-generated traces. While learning analytics has advanced methods for analysing learning processes, less attention has been paid to how learning design generates data on which such analytics depend. Drawing on a multi-year, design-oriented and practice-based research process, this process paper introduces a model that conceptualises pedagogical data production as a configuration of interconnected learning design dimensions. It specifies how the learning-related data, defined by its interpretive function, is produced through the alignment of learning design elements, which structure how such data is generated, becomes visible, and is interpreted, making learners' engagement, reasoning, and competence development visible. The proposed model positions learning analytics as a practice emerging from pedagogically structured learning processes and defines learning design as the generative condition of pedagogically meaningful learning-related data, through which learner agency and pedagogical action can be supported.

Keywords:
learning analytics, pedagogical data production, learning-related data, learning design, learner agency

University of Maribor Press

1 Introduction

In learner-centred distributed higher education (HE), making learning processes visible beyond system-generated traces has become increasingly important. Distributed HE refers to learning across multiple settings (e.g., online, onsite, hybrid, and work-related), where learning involves group work and personalised pathways, and both learners and teachers require timely, context-sensitive information as it unfolds.

While learning analytics (LA) enables large-scale data collection, the pedagogical production of learning-related data remains insufficiently conceptualised within learning design (LD). In this paper, LD is understood as the pedagogical structuring of learning activities, tasks, and assessment. LA has been defined as a discipline for analysing educational data to support learner progress and pedagogical decision-making (Siemens, 2013; Gašević et al., 2015). In practice, LA often relies on indicators such as submissions, log-ins, and time-on-task. These traces enable monitoring at scale but provide only a partial representation of learning processes. This is particularly evident when considering cognitive, motivational, and self-regulatory processes, which are central to learning (Panadero, 2017; Zimmerman, 2002; Buckingham Shum et al., 2019). As a result, what becomes visible is shaped by what is easily measurable, and important dimensions of learning remain less visible (Motz et al., 2023). This indicates a key limitation in current LA approaches. The generation of pedagogically meaningful learning-related data within LD remains insufficiently conceptualised in prior research (Mangaroska & Giannakos, 2019).

The limitations of trace-based approaches become particularly evident in HE contexts characterised by collaborative, project-based, and individually paced learning. While system traces may indicate activity, such as assignment submission, they primarily reflect group-level outcomes and provide limited insight into individual contributions, progression, or understanding. A central challenge is that learning-related trace data does not adequately capture how learning unfolds in social and contextual processes, nor does it represent learners' experiences and sense-making, which require pedagogically designed opportunities for articulation and reflection.

Structural and pedagogical transformations in HE increase the demand for timely, pedagogically meaningful information about learning processes (Ifenthaler & Yau, 2020; Teräs et al., 2020). Growing learner diversity, larger groups, and limited resources reshape how learning takes place. At the same time, in HE, learning-related data is often fragmented across platforms and interactions. The key issue is not the amount of data, but whether it is pedagogically meaningful, aligned with learning goals, and structured to inform action. This points to a central principle: LA must be grounded in LD, where learning activities structure what becomes visible as data.

In this study, meaningful LA refers to the intentional generation of learning-related data that makes learners' understanding, reasoning, and competence development visible in ways that inform pedagogical action and support learner agency. To address this, the study introduces a LA–integrated model of pedagogical data production. The model is articulated through the following research questions:

> RQ1: What constitutes pedagogically meaningful learning-related data within distributed higher education contexts?
> RQ2: How can such data be intentionally produced through LA in ways that support learner agency and pedagogical action?

2 Conceptual Background

While recent scholarship has emphasised aligning LD with educational theory and evidence (Neelen & Kirschner, 2020), the pedagogical generation of meaningful learning-related data remains under-conceptualised. This section develops a conceptual foundation for pedagogical data production through learner agency, LD, and ethical conditions and addresses the current gap in understanding how this data is intentionally generated.

2.1 Pedagogically meaningful learning-related data and learner agency

LA research in higher education has frequently focused on instructional decision-making and institutional monitoring, particularly performance prediction and administrative uses (Leitner et al., 2017). In these approaches, learners often appear as data subjects rather than active participants in interpreting learning-related data (Buckingham Shum et al., 2019; West et al., 2019). While such applications support

large-scale coordination, they do not clarify what constitutes pedagogically meaningful learning-related data within learning processes at the course level.

From a learner-centred perspective, data is defined not by its technical availability but by its relevance to learners' learning processes. Such data becomes pedagogically meaningful for learners when it supports their understanding, regulation, and informed decision-making. Access to learning-related data enables learners to plan, monitor, and adjust their learning, thereby supporting their agency. In this sense, agency refers to learners' capacity to make informed decisions about their learning.

For teachers, pedagogically meaningful learning-related data enhances pedagogical awareness and enables adaptive guidance and timely support within ongoing learning processes (Banihashem et al., 2022, Teräs et al., 2020). In collaborative and distributed contexts, this applies to both individual and shared learning processes. Such data supports teachers in recognising learners' emerging understandings and areas of difficulty, thereby informing pedagogical action while sustaining learner agency.

2.2 LD as a mechanism for meaningful data production

While substantial research has focused on analytic techniques and visualisations, fewer studies have theorised the pedagogical generation of learning-related data within everyday teaching practice. A substantial proportion of studies do not directly measure learning outcomes or implement interventions in learning environments (Motz et al., 2023), and evidence of sustained pedagogical impact remains limited (Viberg et al., 2018). These observations have prompted calls to connect LA more explicitly with learning theory and learners' regulatory processes (Ifenthaler & Yau, 2020; Wong et al., 2019). Access to dashboards or tracking metrics alone does not ensure learner agency, particularly when representations focus primarily on behavioural traces detached from pedagogical context (Bodily & Verbert, 2017).

If pedagogically meaningful learning-related data is defined by its relevance to learning processes, the remaining question is how such data is produced and made visible in practice. Through LD, teachers actually influence which aspects of learning processes may become articulated, documented, and rendered interpretable as learning-related information (Mangaroska & Giannakos, 2019). This perspective

differs from approaches that treat LD primarily as a context for interpreting analytic outputs; instead, LD is understood here as the condition under which learning-related data is generated. What becomes visible as data is therefore shaped by how learning activities are configured and how opportunities for articulation, reflection, and data use are embedded within the learning process.

In competence-based HE, LD commonly follows an alignment logic connecting learning outcomes, assessment, and activities (Biggs, 1996). Beyond this alignment perspective, LD can be understood as a configuration of interrelated pedagogical elements forming a structured and coherent system (van den Akker, 2004) which shapes both learning processes and what becomes visible. From this perspective, LA is not external to LD but embedded within it, as part of the pedagogical system that shapes the production of learning-related data.

While prior research has explored connections between LD and LA, the focus has largely been on analysing learning processes or supporting instructional and design-related decision-making (e.g., Gašević et al., 2015; Er et al., 2019; Lim et al., 2023; Sajja et al., 2025). The perspective proposed here shifts attention to how pedagogical design shapes the data that is subsequently analysed. The central question thus moves from analysing available traces to configuring the pedagogical conditions under which meaningful learning-related data can emerge.

2.3 Responsible and Sustainable Pedagogical Data Production

Pedagogical data production is not value-neutral. LA ethics goes beyond legal compliance to require transparency, intentionality, and proportionality in data practices (Drachsler & Greller, 2016). If the generation of data is not grounded in pedagogically designed activities, or if its use is not transparent, attention may shift towards forms of monitoring that do not directly support learning and may limit learners' opportunities to act as active agents in interpreting and using such information. At the course level, teachers' pedagogical design decisions should determine what data is produced, for whom and how it becomes visible, and how it informs learning. In practice, however, this is not yet consistently recognised or implemented as part of the design phase.

Sustainable pedagogical data production requires that data collection, management, and storage are conducted in an ethical, transparent, and institutionally aligned manner (Drachsler & Greller, 2016). This entails alignment with institutional systems, clear data management practices, and well-defined responsibilities for data handling. At the course level, LA should therefore be grounded in consistent data practices. However, when separate tools outside the Learning Management System (LMS) are used to address emerging information needs, data may become fragmented, increasing management complexity and shifting responsibility for data use—including access rights and security—to individual teachers (Prinsloo & Slade, 2017; Drachsler & Greller, 2016). This fragmentation may weaken the connection between learning activities and the data, for example when learning-related data is distributed across simulation environments or learning tools, making it difficult to use as part of a coherent learning process.

Together, these considerations highlight that pedagogically meaningful data derives its value, for both learners and teachers, from how it is generated through LD and the ethical, transparent, and institutionally aligned conditions under which it is produced, managed, and interpreted. These dimensions also align with prior research on LA and feedback (e.g., Banihashem et al., 2022), which highlights similar forms of learning-related data. Figure 1 visualises these shared forms of learning-related data and their distinct interpretive functions in supporting learner agency and pedagogical action.

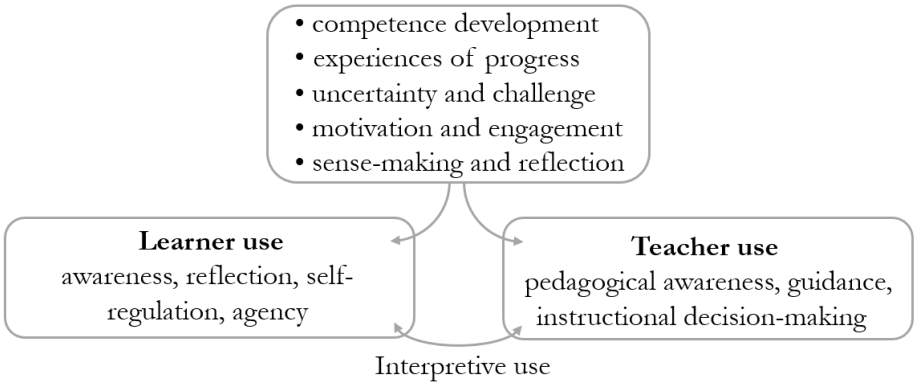

Figure 1: Shared forms of pedagogically meaningful learning-related data and their differentiated interpretation by learners and teachers

As illustrated in Figure 1, these forms of data are shared between learners and teachers but differ in how they are interpreted and used within the learning process. In response to RQ1, pedagogically meaningful learning-related data in distributed HE contexts consists of intentionally configured and pedagogically aligned information defined by its interpretive function rather than its technical origin. Such data makes learners' progress, challenges, motivation, and sense-making processes visible in ways that support learner agency and teacher decision-making under transparent and sustainable conditions.

3 Methodological Approach

This study draws on design-oriented and practice-based research (McKenney & Reeves, 2018; Sandoval, 2014), in which pedagogical interventions are iteratively developed, enacted, and analysed to identify transferable design principles. It presents a process-informed conceptual model developed through a multi-year research and development process (2018–2025) which also builds on five years of doctoral research. The aim is to abstract a conceptual model from sustained design-based work, rather than to report findings from a single empirical intervention.

The empirical development centred on two structurally distinct courses (2 and 10 ECTS), involving approximately 1,500 students annually and resulting in several thousand student participations across diverse cohorts and institutional contexts. The courses were selected to represent the organisational and pedagogical diversity. Together, they capture key variations typical of distributed HE, including differences in pedagogical scope, delivery modes (face-to-face, hybrid, and online), and learning approaches such as project-based, collaborative, small-group, work-integrated, and self-directed study.

Importantly, this variation was inherent to the courses themselves, as they were delivered to diverse learner cohorts and implemented by approximately 50 teachers annually. Participants included both national and international students from HE with selected extensions to other educational settings such as vocational upper secondary education, resulting in both structural and demographic diversity. This enabled examination of whether the emerging design mechanisms remained stable across differing pedagogical and institutional conditions.

Course development proceeded through multiple iterative cycles conducted annually, involving pedagogical co-design, implementation, reflection, and revision of learning tasks and assessment structures. The aim was to configure learning activities to intentionally generate learner-produced, competence-oriented data embedded within the LD.

The generation of learning-related data was pedagogically structured through embedded reflection and self-assessment processes positioned at key points in the learning process. These included diagnostic self-assessment at the beginning of the course to identify learners' initial state, ongoing reflective checkpoints to support monitoring and sense-making, and summative self-assessment at the end to support evaluation of competence development and learning experience. Reflection tasks included prompts related to learners' emotions, perceived progress, activity, development of competence in relation to learning objectives, experiences of teamwork and collaboration, and feedback on teaching and learning materials, capturing both performance-related and experiential aspects of the learning process.

Pedagogical design, course enactment, and research activities were closely intertwined with the configuration of LA within the institutional LMS, which functioned both as the pedagogical environment and as a documentary infrastructure for the development process. Documentation included archived course structures, design artefacts, analytics configurations, learner-produced materials, research notes, and documented teacher–researcher discussions.

This work was conducted in close collaboration between teachers and researchers, with pedagogical decisions and analytic interpretations jointly examined and refined in regularly organised meetings. Students contributed by producing learning-related data through pedagogically designed reflection points and delayed course feedback.

Across iterative cycles, course implementations were compared using a qualitative approach to identify relationships between LD decisions, learner articulation and feedback, data visibility, and pedagogical action. Teacher–researcher meeting discussions were documented to ensure traceability. In one course iteration, the focus was on understanding team functioning and how collaboration unfolded, using student reflections on roles, challenges, and decision-making as the primary data source.

These data informed subsequent design decisions and revealed recurring patterns in how pedagogical design shaped the visibility of learning-related data. Over successive iterations, these patterns were progressively abstracted through systematic comparison into relationships linking design decisions to the production of interpretable evidence, forming the basis of the conceptual model presented in the following section.

4 LD–Integrated Data Production

The model conceptualises pedagogical data production as a **configuration of interconnected LD dimensions** as outlined in Section 2. Figure 2 illustrates how these interconnected dimensions enable the intentional production of pedagogically meaningful data to support learner agency and pedagogical action, particularly in distributed HE contexts.

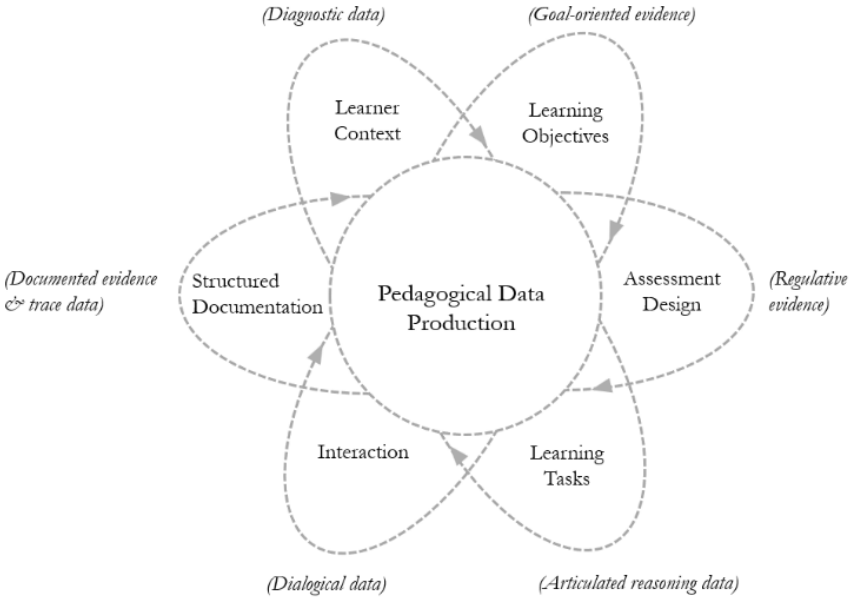

Figure 2: Pedagogical data production model where each dimension structures and contributes to the generation of learning-related data within ongoing learning process

Learner context shapes what becomes visible about learners' starting points, experiences, and constraints. Through structured onboarding tasks and reflection activities embedded within the course design, this configuration generates *diagnostic data* related to prior knowledge, expectations, and emerging uncertainties.

Learning objectives define what counts as learning and thereby shape the nature of the evidence to be produced as *goal-oriented evidence*. Within the course design, they guide the configuration of assessment events and learning tasks that prompt learners to articulate reasoning, demonstrate applied understanding, justify decisions, and reflect in relation to stated objectives.

Assessment design structures when and how *regulative learning-related evidence* is generated by integrating diagnostic, formative, and summative elements into an ongoing assessment process. These assessment moments across the learning process organise the production of data, such as self- and peer-feedback, and evidence of competence development. Assessment design thus organises the timing, form, and use of learning-related data.

Learning task configuration determines what learners are required to articulate, shaping the generation of *articulated reasoning data*. Tasks that involve comparison, explanation, justification, or reflection prompt learners to make their thinking explicit. Such tasks extend beyond behavioural indicators and produce traces of reasoning through which learners produce data about their thinking and experiences. Positioned at specific moments in the learning process, learning tasks also function as points at which learners are activated to produce data.

Interaction structures determine how learners engage in collaborative phases of the learning processes, shaping the generation of *dialogical data*. Discussions, peer and teacher interaction, and guided interaction prompt learners to articulate perspectives and negotiate meaning, producing traces of participation, collaboration, and shared understanding. Through interaction structures learners produce data about social and dialogical dimensions of learning.

Documentation structures, supported by the LMS, provide the infrastructure for recording and making learning-related data available for use within the course. An LMS enables the documentation of pedagogically structured artefacts. This may

include data originating from different learning contexts (e.g. external tools), which is brought into the LMS through learning activities or tasks. Centralising data within institutional LMS environments supports consistent, transparent, and secure handling of it. While the LMS organises this *documented evidence and trace data*, it does not determine its pedagogical meaning.

The model conceptualises LD as a data-generating mechanism in which pedagogical elements such as objectives, tasks, assessment, interaction, and documentation structures function as conditions for the generation, visibility, and interpretability of learning-related data within ongoing learning processes. Together, these elements form a coherent pedagogical configuration in which data is intentionally produced, made visible, and can be used as part of the learning process. When data production is embedded in ongoing learning processes, learners become active producers of such data, while both learners and teachers act as its interpreters and users. In this way, LA can support learner-centred understanding and foster learner agency (e.g., Hooshyar et al., 2023), while also providing a basis for data-informed pedagogical decision-making.

5 Implications for LA as pedagogical practice

Experiential and competence-related dimensions of learning become analytically visible through pedagogical structures that invite learners to articulate, justify, document, and reflect on their activity. This repositions learners not only as recipients of analytics but as active contributors to the production of interpretable evidence within ongoing learning processes (see also e.g. Hooshyar et al., 2023; Viberg et al. 2018).

Rather than treating LA as a technological capacity emerging from trace infrastructures (e.g., Gašević et al., 2015; Drachsler & Greller, 2016), the model conceptualises analytics as a property of pedagogical structure. What becomes analytically visible is shaped by design decisions guided by anticipated informational needs, rather than by system capability alone. Pedagogical dimensions generate distinct forms of meaningful learning-related data-diagnostic, goal-oriented, regulative, dialogical, and reasoning-based-thereby addressing RQ1, while specifying how such data can be intentionally produced through learning design (RQ2). In distributed HE contexts, the key challenge is how learning-related data is

pedagogically produced and made meaningful across learning processes that span multiple environments. This shifts the focus of learning analytics from system-generated traces to pedagogically structured learning processes, enabling the model to extend beyond digitally mediated environments to diverse learning contexts. Repositioning LA as a pedagogical property does not remove institutional constraints, but clarifies that meaningful analytics depends on pedagogical conditions under which learning becomes visible, interpretable, and actionable.

In this perspective, LA primarily supports learning during the process by informing self-regulation and pedagogical guidance, rather than retrospective monitoring. By defining learning-related data structurally, the model extends beyond system-specific implementations and can operate across diverse learning contexts. This reconceptualisation aligns LA with a data-informed perspective, where pedagogically configured evidence complements professional judgment and positions teachers as designers of the conditions under which learning-related data can emerge.

As AI-supported and data-informed education expands, its value depends on the availability of pedagogically meaningful learning-related data (e.g., Yan et al., 2024; Sajja et al., 2025). This study emphasises that this value ultimately depends on how such data is pedagogically produced and made meaningful within learning processes.

6 Conclusion

This study demonstrates that learning-related data becomes interpretable through pedagogical configuration. LA is best understood as a pedagogically embedded practice that emerges from intentionally designed learning processes producing interpretable evidence. By conceptualising LD as the generative condition of learning-related data, the study shows how such data can be intentionally produced within ongoing learning processes to support learner agency and pedagogical action. In this perspective, learners are not only recipients of analytics but active contributors to the production of learning-related evidence.

End notes

Use of Generative AI in the Writing Process. Generative AI was used solely for language editing. All research and interpretation are the authors' own, and they take full responsibility for the manuscript.

References

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review, 37*, 100489. https://doi.org/10.1016/j.edurev.2022.100489

Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32(3), 347–364. https://doi.org/10.1007/BF00138871

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics*, *6*(2), 1–9. https://doi.org/10.18608/jla.2019.62.1

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue: A checklist for trusted learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 89–98). ACM. https://doi.org/10.1145/2883851.2883893

Er, E., Gómez-Sánchez, E., Dimitriadis, Y., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., & Alvarez-Alvarez, S. (2019). Aligning learning design and learning analytics through instructor involvement: A MOOC case study. *Interactive Learning Environments*, 27(5–6), 685–698. https://doi.org/10.1080/10494820.2019.1610455

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Hooshyar, D., Tammets, K., Ley, T., Aus, K., & Kollom, K. (2023). Learning analytics in supporting student agency: A systematic review. *Sustainability, 15*(18), 13662. https://doi.org/10.3390/su151813662

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success: Reflections on current practice and future directions. Educational Technology Research and Development, 68(4), 1963–1988. https://doi.org/10.1007/s11423-020-09788-z

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education - A literature review. *Learning Analytics: Fundaments, Applications, and Trends* (pp. 1-23). Springer, Cham. https://doi.org/10.1007/978-3-319-52977-6_1

Lim, L. A., Atif, A., Heggart, K., & Sutton, N. (2023). In Search of Alignment between Learning Analytics and Learning Design: A Multiple Case Study in a Higher Education Institution. *Education Sciences*, 13(11), 1114. https://doi.org/10.3390/educsci13111114

Mangaroska, K., & Giannakos, M. (2019). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. IEEE Transactions on Learning Technologies, 12(4), 516–534. https://doi.org/10.1109/TLT.2018.2868673

McKenney, S., & Reeves, T. (2018). Conducting Educational Design Research (2nd ed.). Routledge. https://doi.org/10.4324/9781315105642

Motz, B. A., Bergner, Y., Brooks, C. A., Gladden, A., Gray, G., Lang, C., Li, W., Marmolejo-Ramos, F., & Quick, J. D. (2023). A LAK of Direction: Misalignment Between the Goals of Learning Analytics and its Research Scholarship. *Journal of Learning Analytics*, *10*(2), 1–13. https://doi.org/10.18608/jla.2023.7913

Neelen, M., & Kirschner, P. A. (2020). *Evidence-informed learning design: Creating training to improve performance*. Kogan Page.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422. https://doi.org/10.3389/fpsyg.2017.00422

Prinsloo, P., & Slade, S. (2017). *An elephant in the learning analytics room: The obligation to act.* In J. A. Larusson & B. White (Eds.), Learning analytics: From research to practice (pp. 269–282). Springer.

Sajja, R., Sermet, Y., Cwiertny, D., & Demir, I. (2025). Integrating AI and learning analytics for data-driven pedagogical decisions and personalized interventions in education. *Technology, Knowledge and Learning*, 1–31. https://doi.org/10.1007/s10758-025-09897-9

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences*, *23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Teräs, M., Silvola, A., Teräs, H., Hartikainen, S., Aksovaara, S., Hietaniemi, R., & Muukkonen, H. (2020). Learning analytics for students: Synthesis of two user needs studies in Finnish higher education. In *EdMedia + Innovate Learning* (pp. 455–463). Association for the Advancement of Computing in Education (AACE). https://urn.fi/URN:NBN:fi-fe2022021519072

van den Akker, J. (2004). Curriculum Perspectives: An Introduction. In *Curriculum Landscapes and Trends*. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-1205-7_1

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

West, D., Luzeckyj, A., Toohey, D., Vanderlelie, J., & Searle, B. (2019). Do academics and university administrators really know better? The ethics of positioning student perspectives in learning analytics. *Australasian Journal of Educational Technology, 35*(6), 27–41. https://doi.org/10.14742/ajet.4653

Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., Houben, G.-J., & Paas, F. (2019). Educational theories and learning analytics: From data to knowledge. In *Utilizing Learning Analytics to Support Study Success* (pp. 3–25). Springer, Cham. https://doi.org/10.1007/978-3-319-64792-0_1

Yan, L., Martinez-Maldonado, R., & Gašević, D. (2024). Generative artificial intelligence in learning analytics: Contextualising opportunities and challenges through the learning analytics cycle. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 101–111). ACM.https://doi.org/10.1145/3636555.3636856

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2