Personalization in Educational Data Mining and Learning Analytics: A Systematic Review (2015 - 2025)

Sylvio Rüdian
Humboldt-Universität zu Berlin
Department of Computer Science & Education
Berlin, Germany
ruediasy@hu-berlin.de

ABSTRACT

Personalization is widely recognized as a path toward more effective digital learning; however, its conceptual and methodological treatment varies across different research traditions. This study systematically synthesizes a decade of work (2015 - 2025) in the Educational Data Mining (EDM) and Learning Analytics & Knowledge (LAK) communities employing the PRISMA framework to clarify how personalization is enacted and evaluated. We compile and screen 503 papers to isolate those that implement learner-level adaptation, coding what is personalized, the indicators that drive it, and how studies assess its impact. A mixed quantitative-qualitative analysis compares the two communities, highlighting distinct emphases, recurring methodological challenges, and complementary strengths. The review provides an integrated map of personalization research, identifies persistent gaps in validity, adoption, and causal evidence, and outlines opportunities for cross-fertilization between technically focused EDM models and human-centered LAK practices.

Keywords

Adaptive Systems, Intelligent Tutoring System, Review

1. INTRODUCTION

Personalization has long been one of the most compelling promises of digitally mediated learning: the idea that learning material, feedback, and support can be adapted to meet the needs of individual learners rather than delivered as one-size-fits-all instruction. Over the past decade, this promise has been intensively investigated by two closely related research communities: Educational Data Mining (EDM) and Learning Analytics and Knowledge (LAK). They share many data sources and venues but often differ in theoretical commitments, methodological preferences, and the artifacts they produce [13]. As algorithmic techniques for sequential decision making and representation learning matured, and as institutions deployed analytics tools to support teachers and students, the term "personalization" became diverse. Yet it is used to describe a wide variety of interventions, from reinforcement-learning policies that select the "next best" activity to dashboards that help instructors tailor feedback. Without a careful, comparative map of how EDM and LAK operationalize personalization, it is challenging to synthesize evidence, identify gaps, or chart a coherent course of action.

The similarities and differences between LAK and EDM have been extensively examined in the scholarly literature. Both fields operate within the educational domain, leveraging data collection and analysis to improve learning processes and student outcomes. They share a commitment to data-driven approaches for understanding and enhancing education, frequently utilizing overlapping datasets and tasks such as predicting student performance or identifying at-risk learners [22]. Both LAK and EDM exploit large-scale educational data, employ statistical and machine learning techniques, and aim to transform raw data into actionable insights for educators, students, and administrators. Core methodologies such as classification, clustering, and prediction are widely used in both areas, reflecting their shared origins in the computational analysis of educational environments [32].

Despite these commonalities, LAK and EDM differ in methodological orientation and epistemological emphasis. EDM has traditionally focused on the development of novel computational techniques to interrogate educational data, often emphasizing model-driven approaches such as association rule mining, clustering, and predictive modeling. In contrast, LAK integrates a broader range of methods, including visualization, social network analysis, and statistical modeling, while maintaining a strong emphasis on the practical application of analytics to enhance learning and teaching. LAK scholarship frequently highlights the human and organizational dimensions of education. The two fields are therefore proximate yet complementary: EDM has historically concentrated on model- and policy-centric questions - such as how to represent latent knowledge, predict academic success, or optimize instructional decisions under uncertainty - whereas LAK has invested heavily in human-centered analytics, focusing on making data interpretable and actionable for learners, teachers, and institutions through feedback, visualization, and participatory design. Both perspectives are indispensable for advancing credible personalization at scale. Algorithmic policies lacking usability, transparency, or alignment with pedagogical practice are unlikely to succeed in authentic educational contexts, while well-designed

Sylvio Rüdian. Personalization in Educational Data Mining and Learning Analytics: A Systematic Review (2015 - 2025). In Anthony Botelho, Maria Mercedes T. Rodrigo, Adish Singla, Hiroaki Ogata, Hyojeong So, and Young Hoan Cho (eds.) Proceedings of the 19th International Conference on Educational Data Mining, Seoul, Republic of Korea, June, 2026, pp. 229–242. International Educational Data Mining Society (2026).

https://doi.org/10.5281/zenodo.21039837

dashboards that lack empirically validated adaptive mechanisms risk becoming well-intentioned but ineffective advice systems. A comparative perspective thus illuminates areas of conceptual convergence, points of productive methodological divergence, and opportunities for cross-fertilization that could accelerate progress in the field. Recent bibliometric studies further indicate that LAK typically pursues research with the explicit intent of improving learning actions and environments, whereas EDM is more strongly oriented toward understanding learning behaviors through algorithmic analysis [32]. However, a focused comparison of how the two communities conceptualize and operationalize personalization remains absent.

2. FROM ADAPTATION TO PERSONALIZATION

Personalization is widely invoked in digital learning, yet the term is used inconsistently across research traditions and is frequently conflated with adjacent concepts such as adaptation and customization. Because this paper compares personalization work across EDM and LAK, we require a definition that is conceptually defensible, broad enough to cover both communities' artifacts, and precise enough to support coherent interpretation of results.

Because our aim is to compare two communities whose personalization artifacts range from fully algorithmic decision policies (e.g., sequencing and recommendation) to teacher- and learner-facing tools (e.g., dashboards and feedback systems), we adopt a definition that is simultaneously inclusive and boundary-setting. Specifically, we require a framing that (a) is broad enough to cover both automated and human-in-the-loop instantiations of individualized support, (b) excludes changes that are applied uniformly at the cohort or course level (i.e., adaptation without differentiation between learners), (c) distinguishes system-driven personalization from user-initiated customization, and (d) is operationalizable for transparent screening and coding. This emphasis on codeability matters for a comparative review: without explicit decision rules, the term "personalization" risks collapsing into a rhetorical umbrella, making it difficult to interpret patterns across venues or to attribute differences to real methodological or conceptual divergence rather than to definitional drift.

We begin with the general notion of *adaptation*. In dictionary terms, adaptation denotes modifying an artefact, process, or environment so that it suits a new purpose or situation, and an *adaptive* system is one that can change in response to varying conditions [86]. These definitions are intentionally broad: adaptation can occur at multiple levels (individual, subgroup, course, or institution) and can be enacted by humans, algorithms, or both. In educational contexts, this breadth is reflected in uses of "adaptation" that range from optimizing the fit between learning requirements and course content [61] to broader forms of adjustment such as shifting instructional delivery modes or transferring courses across settings [34]. Such interventions involve change in response to context, but they are not necessarily individualized.

Personalization is a narrower, person-centered form of adaptation. Building on the Oxford framing, we use personalization to refer to tailoring content, processes, or interfaces to the needs of a particular individual [86]. The key differentiator is the *target of the adaptation*: personalization aims to differentiate the learning experience between learners rather than to adjust an experience uniformly for a cohort. This individualized focus is prominent across multiple strands of the literature. Work on adaptive hypermedia and web-based learning has long framed personalization as selecting and presenting information based on learner models [17, 18]. In online course settings, personalization is often instantiated as individualized sequencing or learning-path construction, where the system selects activities or resources based on learner prerequisites or performance [90, 8]. In the broader discourse of "personalized learning", personalization is frequently described as optimizing pace and approach to the learner, encompassing learning objectives, instructional strategies, and the sequencing of content [85, 8]. Related accounts emphasize motivational and competency dimensions [37] or characterize personalization along practical axes such as goals, path, time, place, and pace [10]. Although these formulations vary in granularity and pedagogical emphasis, they converge on the principle that instructional decisions should depend on learner-specific information rather than being uniform for all learners.

A second distinction concerns *locus of control*. Whereas personalization and adaptation are commonly enacted system-side (often implicitly through models or rules) *customization* refers to changes initiated by users based on explicitly specified preferences [86, 93]. Customization can contribute to individualized experiences, but it differs from personalization in who specifies the change and how it is triggered. This distinction is especially relevant for learning systems, where personalization is frequently implemented without explicit learner requests (e.g., through adaptive sequencing, recommendations, or tailored feedback).

This paper uses *personalization* to denote **individual-level, person-centered adaptation** of the learning experience, distinguished from broader adaptation by its learner-specific target and from customization by its system-side locus of control. This definition supports a consistent interpretation of the diverse mechanisms labeled as personalization in EDM and LAK, ranging from sequencing and recommendation to dashboards and feedback systems, while avoiding conflation with non-individualized course-level adjustments.

3. THIS REVIEW

This review provides a structured, side-by-side account of personalization research in EDM and LAK from 2015 to 2025. Rather than treating "personalization" as a rhetorical label, we treat it as an operational construct: an intervention that adapts some aspect of the learning experience to an individual learner (or to explicitly defined learner subgroups) using learner-specific information. The contribution of the paper is therefore integrative rather than algorithmic: it offers an auditable comparative map of how personalization is instantiated across two influential communities, the types of signals used to drive it, and the kinds of evidence used to justify impact. We do not aim to produce a pooled effect size of "whether personalization works"; instead, we clarify what counts as personalization in this literature and characterize where each community concentrates its efforts and how strongly those efforts are tied to outcome-focused evaluation.

Building from the operational definition, we distinguish (i) what is personalized (e.g., task sequencing, recommendations, feedback, dashboards, assessment criteria), (ii) which indicators are used to drive adaptation (behavioral traces, knowledge-state estimates, text, demographics/context, affect/motivation), and (iii) which techniques implement it (reinforcement learning, BKT / IRT and related cognitive models, recommender algorithms, supervised ML, deep learning, visual analytics, rules, and qualitative / UX / co-design). Because personalization claims often risk conflating usage with learning, we also attend to evaluation designs and explicitly differentiate studies that measure effects on learning outcomes from those that report only process, engagement, or usability, treating process measures as outcomes only when a study provides an explicit validity argument that they are learning-proximal in context.

Two additional considerations motivate a careful synthesis at this time. First, the proliferation of powerful modeling techniques, contextual bandits and deep reinforcement learning, sequence-aware recommenders, and transformer-based approaches for text and multimodal traces, has increased the feasibility of fine-grained adaptation while raising the stakes around fairness, privacy, and explainability. Second, institutions and platforms are deploying analytics at scale, creating opportunities for rigorous evaluation (e.g., pragmatic trials, stepped-wedge designs) while also exposing persistent gaps in construct validity and generalizability across courses and populations. A review that connects what is being personalized to how it is decided and what evidence supports it can help the field move beyond isolated successes toward dependable practice.

Our contribution is threefold. First, we build a transparent corpus by identifying LAK and EDM papers that mention personalization (2015–2025), then screening for studies that are substantively about personalization rather than merely naming it. This keeps the review focused on instantiated mechanisms rather than aspirations. Second, we extract a common set of variables from the included studies, definitions, targets of personalization, indicators, techniques, contexts and scales, evaluation designs, and reported limitations/future work, and organize papers into interpretable clusters defined by the object of personalization, yielding community-specific profiles that can be compared directly. Third, we use those profiles to answer a compact set of research questions that address where each community focuses its personalization efforts, which signals it relies on, and what gaps remain most salient for future work.

We structure the inquiry around three research questions that bridge description and critique:

RQ1: In studying personalization, where do EDM and LAK place their main focus, and how do those foci differ?

RQ2: Which learner- and context-level indicators are used to drive personalization in EDM and LAK?

RQ3: What research gaps and future directions emerge across the two communities?

Answering these questions yields both a map and a mirror: a map of the conceptual and methodological terrain of personalization across two influential communities, and a mirror in which each community can see its strengths and blind spots. By anchoring personalization in concrete targets and signals, and by contrasting EDM and LAK along those axes, the review aims to move the conversation from whether personalization "works" in the abstract to how we can design effective, equitable, and adoptable personalization in practice.

4. METHODOLOGY

4.1 Review design and reporting framework

We conducted a structured literature review of personalization research in the LAK and EDM conference proceedings for the years 2015–2025. Our goal is to map how personalization is instantiated (RQ1), what signals are used to drive it (RQ2), and what limitations/future directions are reported (RQ3). In terms of review type, this work aligns most closely with a *scoping review* (i.e., aiming to characterize and map a body of work rather than estimate an effect size). To improve transparency and reproducibility, we report our search, screening, and selection steps in a PRISMA-compatible structure (flow counts and explicit eligibility criteria). Where PRISMA terminology is used, it is adopted as reporting guidance rather than implying a registered SLR protocol.

4.2 Information sources and search strategy (high-recall)

We searched the official proceedings and hosting digital libraries for LAK and EDM for each year in scope. To maximize recall while keeping the query auditable, we used a single high-recall lexical stem centered on `personali*`. This stem captures spelling variants and common morphological forms, including (non-exhaustively) *personalization*, *personalisation*, *personalized / personalised*, *personalizing / personalising*, and *personalizability / personalisability*. We executed the query separately for each venue-year and community. Overall, we record the authors, identifier / DOI, year, venue, title, and abstract, and exported all data to a CSV and concatenated into a raw "intake" file. Each row corresponds to one retrieved record prior to de-duplication and screening.

4.3 Eligibility criteria (inclusion/exclusion)

We predefined inclusion and exclusion criteria before screening and applied them consistently throughout.

Inclusion criteria - A record was eligible if it:

- is a paper in LAK or EDM proceedings within 2015 - 2025 (excluding workshop-only volumes if not part of the main proceedings, unless explicitly in scope);
- contains an instantiated personalization/adaptation mechanism (see operational definition below).

Exclusion criteria - A record was excluded if it:

- is not a LAK/EDM proceedings paper in scope (e.g., abstract-only items, workshops);
- mentions personalization only rhetorically (e.g., as motivation or aspiration) without an implemented adaptive mechanism.

We define personalization as an *intentional adaptation* of content, sequence/path, feedback, support, interface, or assessment to individual learners (or explicitly defined learner subgroups) using learner-specific information (data, models, or explicit rules). Mere parameter tuning at the population level, generic "one-size-fits-all" recommendations, or aspirational statements without a realized adaptive mechanism do not qualify.

4.4 Screening and selection process

Screening proceeded in two stages and was documented in a dedicated workbook sheet with validation rules.

Stage 1: about-personalization decision - We screened each unique record to determine whether it is substantively about personalization under the operational definition above.
Records were labeled `Include`, `Exclude`, or `Unclear`. Unclear cases were resolved via discussion using decision notes and adjudicated conservatively (i.e., inclusion required evidence of an instantiated adaptive mechanism).

Stage 2: learning-outcomes evaluation decision - In Stage 2, we focus on outcome evidence because this review aims to synthesize support for learning impact claims, i.e., whether personalization changes what learners know or can do, rather than only whether an intervention is used, accepted, or engaging. Process characteristics (e.g., clicks, time-on-task, persistence, navigation patterns, or strategy indicators) are often meaningful as mediators or early warning signals, but they are not consistently identifiable with learning across contexts and can be driven by novelty, interface constraints, or compliance effects. Given the field's well-known risk of conflating "more activity" with "more learning", we treat process measures as learning outcomes only when a study provides an explicit validity argument that the process metric is learning-proximal in that context (e.g., validated linkage to performance/competency or a preregistered/justified surrogate endpoint). Otherwise, process measures are coded as indicators or intermediate outcomes rather than learning outcomes. For records included after Stage 1, we assessed whether the study measured effects on learning outcomes. We treated learning outcomes as changes in knowledge, skill, or performance evidenced by assessments (pre/post or delayed), mastery probabilities, grades explicitly tied to competencies, validated performance tasks, or completion when justified as competency-proximal. Engagement, usability, satisfaction, or time-on-task were not counted as learning outcomes unless explicitly validated within the study design as learning-proximal endpoints.

PRISMA-style accounting of records - To support reproducibility, we report record counts at each step:

- Records retrieved by searches in EDM and LAK: **503** (LAK: $n = 370$, EDM: $n = 133$)
- Records excluded in Stage 1 (not personalization): **269** (LAK: $n = 226$, EDM: $n = 43$)
- Records included after Stage 1: **234** (LAK: $n = 144$, EDM: $n = 90$)
- Records excluded in Stage 2 (no learning-outcome effects measured): **174** (LAK: $n = 117$, EDM: $n = 57$)
- Records included after Stage 2 (learning outcomes measured): **60** (LAK: $n = 27$, EDM: $n = 33$)

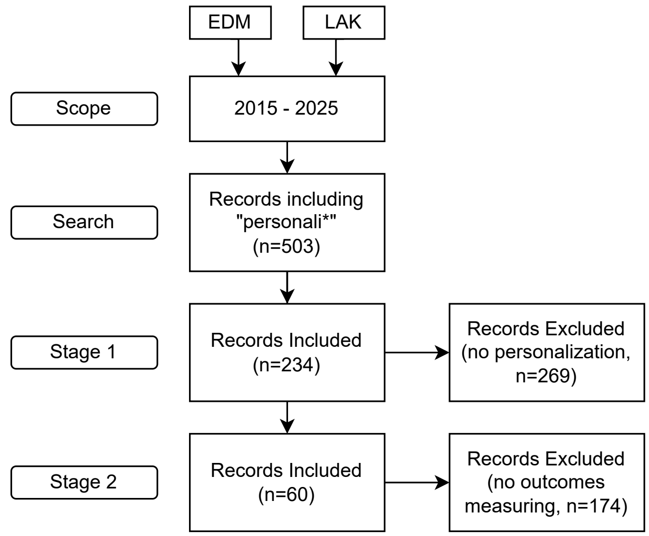

Figure 1: Inclusion and Exclusion of Records.

These counts correspond to the workbook filters and enable reconstruction of the selection process. Fig. 1 visualizes the accounting of records.

4.5 Data extraction, codebook, and coder reliability

Workbook and codebook - We maintained a multi-sheet Excel workbook as the study database. A `Dictionary` sheet defined each variable and its admissible values; `Metadata` stored bibliographic fields and provenance; `Screening` captured inclusion decisions with validation rules and inconsistency flags; and `Coding` contained analytical variables with controlled vocabularies (e.g., object of personalization, indicator families, techniques).

Coding variables - For all papers included after Stage 1, we extracted variables aligned with RQ1–RQ3. For RQ1, we coded the *object(s) of personalization* as non-exclusive categories (e.g., content, sequence/path, feedback, interface, support, assessment) and recorded domain and scale descriptors (participants and/or interaction records). For RQ2, we coded the indicators used to drive personalization (multi-select), such as knowledge-state estimates, performance history, demographics / learner profile, outcomes, behavioral traces, or social / context signals. For RQ3, we transcribed limitations and future-work statements as close paraphrases and coded them into thematic categories.

Coders and disagreement resolution - Coding was performed by **6** coders. To ensure consistency, we conducted a pilot on

30 papers to refine the codebook and decision rules. Each paper was coded by **two independent coders**. Disagreements were resolved through discussion. When ambiguous, both codings have been included. We assessed reliability on a double-coded subset of **40** papers using **Cohen's** $\kappa = .78$ for key categorical variables (Stage 1 decision, object of personalization, indicator families). Agreement was **Cohen's** $\kappa = .84$ in stage 2 and informed final refinements to the codebook prior to full coding.

4.6 Analysis

Analysis proceeded in three coordinated steps mapped to the research questions. For RQ1, we treated the object-of-personalization codes as the primary organizing axis and computed category distributions separately for EDM and LAK, yielding community-specific profiles of where personalization efforts concentrate. Because these object categories directly reflect how adaptation is instantiated, this approach supports straightforward validity checks and avoids the opacity of unsupervised topic modeling. To enrich the profiles, we examined within-category co-occurrence of indicators.

For RQ2, we compared the prevalence of indicator families across communities both overall and conditional on object category (e.g., whether knowledge-state estimates co-occur more frequently in EDM than in LAK, or whether LAK draws more heavily on learner profiles and social/context signals). For RQ3, we synthesized coded limitations and future-work themes to identify common gaps and directions, reported overall and stratified by venue and object category where informative.

In summary, this review identifies where EDM and LAK concentrate their personalization efforts (RQ1), which signals they rely on (RQ2), and which gaps are most salient for future research (RQ3).

5. RESULTS

Based on how each paper operationalizes personalization and the target goal to which personalization is optimized, we identified seven non-exclusive personalization types:

- **Learning Path (Learning Material)**: Adaptive selection and sequencing of course materials, activities, and difficulty to guide each learner's path.
- **Recommender System (Course, External Resources)**: Recommendations for courses/programs and external resources (videos, articles, apps) tailored to learners' goals and performance.
- **Dashboard**: Personalized analytics/visualizations that surface progress, risks, and suggested actions to support self-regulated learning.
- **Feedback/Support/Notifications**: Tailored hints, explanations, formative feedback, reminders, and intervention messages triggered by learner state.
- **Instruction**: Adaptation of instructional strategies and supports (e.g., tutoring policies, scaffolding, differentiated activities, teacher support allocation).
- **Scheduling**: Personalized timing and pacing - practice/test scheduling, spacing and review intervals, and ordering of assignments.
- **Collaboration/Forum**: Social personalization - peer / mentor recommendations, forum thread suggestions, group-based differentiation, and support for collaborative activities.

Papers were allocated to the personalization types as shown in the extract of Table 1. The entire table with all codings can be found at `https://doi.org/10.5281/zenodo.20023758`[79]. Six papers were excluded from type-based analyses because they could not be linked to any of the seven types ([92, 73, 51, 45, 48, 76]). Fig. 2 illustrates the absolute numbers of personalization-related papers clustered by personalization type, including all included papers (top) and the subset measuring learning outcomes (bottom). Absolute counts are shown to reflect overall research output, given the larger volume of LAK papers in stage 1, rather than emphasizing relative proportions.

5.1 RQ1: Where do EDM and LAK place their main focus, and how do those foci differ?

To test whether EDM and LAK differ in their emphasis across personalization types, we first applied Levene's tests to determine whether equal variances could be assumed. We then used independent samples t-tests when Levene's tests were non-significant and Welch tests otherwise.

Because RQ1 (and RQ2) involve comparing multiple categories, we interpret the p-values jointly rather than as independent tests. We therefore report exact p-values and effect sizes for transparency and interpret exploratory / descriptive. Given the review's descriptive / scoping aim and the non-independence of categories, we did not apply a strict family-wise correction (e.g., Bonferroni), which would be overly conservative and increase Type II error in an already power-limited setting, especially for the learning-outcomes subset ($n = 60$).

As a sensitivity check, we note that the key community differences in the full corpus (recommender systems, dashboards, and feedback / support with $p < .001$) would remain statistically significant under common false-discovery-rate control (e.g., Benjamini–Hochberg at $q = .05$), whereas marginal findings should be interpreted as exploratory and are not relied upon for central claims.

Considering all personalization-related papers ($n = 234$), Levene's tests were significant for learning path, recommender system, dashboard, and feedback/support ($p \leq .001$). Levene's tests for the remaining types were also significant ($p \leq .05$). The two-sided p-values of the Welch tests were below .001 for recommender system, dashboard, and feedback / support, indicating statistically significant differences between communities. For the remaining types, p-values were above .05, indicating no statistically significant differences at $p \leq .05$. Effect sizes were medium to large for feedback / support (Cohen's $d = -.616$), medium for recommender systems (Cohen's $d = .526$) and dashboards (Cohen's $d = -.477$), and small for learning path (Cohen's $d = .274$).

Table 1: Personalization types and indicators of LAK and EDM papers, which include personalization effects on the learning outcome (extract).

EDM/LAK		Year	Personalization Type							Indicators											
			Learning Path	Recommender System	Dashboard	Feedback/Support/Notifications	Instruction	Scheduling	Collaboration/Forum	Outcomes & Achievement	Knowledge/Mastery State	Assessment & Item Properties	Temporal Pacing & Workload	Behavioral Interaction	Strategy, Process & Help-Seeking	Learner Profile & Context	Social & Community Signals	Content/Curriculum	Model/Policy Signals & Diagnostics	Work Products & Artifact Features	Records & Administrative Data
EDM	[58]	2017	-	-	-	-	-	-	✓	-	-	-	✓	✓	-	✓	✓	-	-	-	-
EDM	[78]	2017	✓	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-
EDM	[69]	2017	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EDM	[92]	2018	-	-	-	-	-	-	-	-	-	-	-	✓	-	✓	-	-	-	-	-
EDM	[73]	2019	-	-	-	-	-	-	-	-	-	-	✓	✓	-	✓	-	-	-	-	-
EDM	[50]	2019	✓	-	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	-	-	-
EDM	[87]	2019	✓	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-
EDM	[53]	2020	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	✓	-	-	-
EDM	[31]	2020	-	-	-	-	-	✓	-	-	✓	-	✓	-	-	-	-	-	-	✓	-
EDM	[56]	2020	-	✓	-	-	-	-	-	✓	✓	-	✓	-	-	✓	-	-	✓	-	-
EDM	[83]	2020	✓	-	-	-	-	-	-	✓	✓	-	-	-	-	✓	-	✓	-	-	-
EDM	[81]	2021	-	-	-	-	-	✓	-	✓	-	-	✓	✓	-	✓	✓	-	✓	-	-
EDM	[21]	2021	✓	-	-	-	-	-	-	✓	-	-	-	-	-	✓	-	-	-	-	-
EDM	[66]	2021	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	✓	✓	-
EDM	[44]	2022	-	✓	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-
EDM	[74]	2022	✓	-	-	-	✓	-	-	✓	-	-	✓	-	-	-	-	-	-	-	-
EDM	[80]	2023	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-
EDM	[60]	2023	-	-	-	✓	-	-	-	-	-	-	✓	✓	-	-	-	-	-	✓	-
EDM	[7]	2024	✓	-	-	-	-	-	-	✓	-	✓	✓	-	-	-	-	-	-	-	-
EDM	[67]	2025	-	-	✓	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-
EDM	[71]	2025	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-
EDM	[89]	2025	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EDM	[72]	2025	-	-	-	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
LAK	[26]	2017	-	-	✓	✓	-	-	-	-	-	-	✓	✓	-	✓	✓	-	-	-	-
LAK	[48]	2017	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-
LAK	[76]	2018	-	-	-	-	-	-	-	-	-	-	-	✓	-	✓	✓	-	-	-	-
LAK	[64]	2019	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-
LAK	[59]	2020	-	-	✓	-	-	-	-	✓	✓	-	-	-	-	-	-	-	-	-	-
LAK	[68]	2020	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LAK	[54]	2021	-	-	-	✓	-	-	-	✓	-	-	-	✓	-	✓	-	-	-	-	✓
LAK	[57]	2021	-	✓	-	-	-	-	-	✓	-	-	-	✓	✓	✓	✓	-	-	-	✓
LAK	[70]	2021	-	-	-	✓	-	-	-	✓	✓	-	-	-	-	-	-	-	-	-	-
LAK	[91]	2022	-	-	-	✓	-	-	-	✓	-	-	-	-	-	✓	-	-	-	-	-
LAK	[49]	2022	✓	-	-	-	-	-	-	✓	-	-	-	✓	-	✓	-	-	-	-	-
LAK	[65]	2022	-	-	-	✓	-	-	-	✓	-	-	-	-	-	✓	-	-	-	✓	-
LAK	[39]	2023	-	-	-	✓	-	-	-	✓	-	-	-	-	-	-	-	-	✓	-	-
LAK	[6]	2024	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-
LAK	[36]	2024	-	✓	-	-	-	-	-	-	-	-	-	-	-	✓	✓	✓	-	-	✓
LAK	[84]	2024	✓	-	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	✓	-	-
LAK	[45]	2025	-	-	-	-	-	-	-	-	✓	-	✓	✓	-	-	-	-	-	✓	-
LAK	[52]	2025	-	-	-	✓	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	-
LAK	[51]	2025	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LAK	[63]	2025	-	-	-	✓	-	-	-	✓	-	-	-	-	-	✓	-	-	-	-	-

We repeated the same analysis for papers that measure personalization effects on learning outcomes ($n = 60$). Levene's tests were significant under $p \leq .05$ for learning path ($p = .002$), dashboard ($p = .012$), feedback/support ($p = .012$), instruction ($p \leq .001$), and collaboration ($p = .007$), indicating unequal variances. Neither the p-values of the two-sided Welch tests nor those of independent samples t-tests were significant under $p \leq .05$. Thus, while absolute counts suggest different emphases in the learning-outcomes subset, the subset is too small to detect statistically reliable differences.

In summary, this finding answers RQ1 and reveals that EDM and LAK papers published between 2015 and 2025 differ in terms of personalization type: LAK papers focus more prominently on feedback / support and dashboards, while EDM papers focus more often on recommender systems.

A minor difference is observed for learning-path personalization, where LAK papers occur slightly more frequently overall. For papers that also measure effects on learning outcomes, the number of included papers is too small to identify statistically significant differences. However, descriptively, 42% of EDM papers in the learning-outcomes subset focus on learning-path adaptations, while in LAK, only 22% focus on learning-path adaptations when learning outcomes are measured, although there is no remarkable difference when papers are not filtered to explicitly focus on learning outcomes.

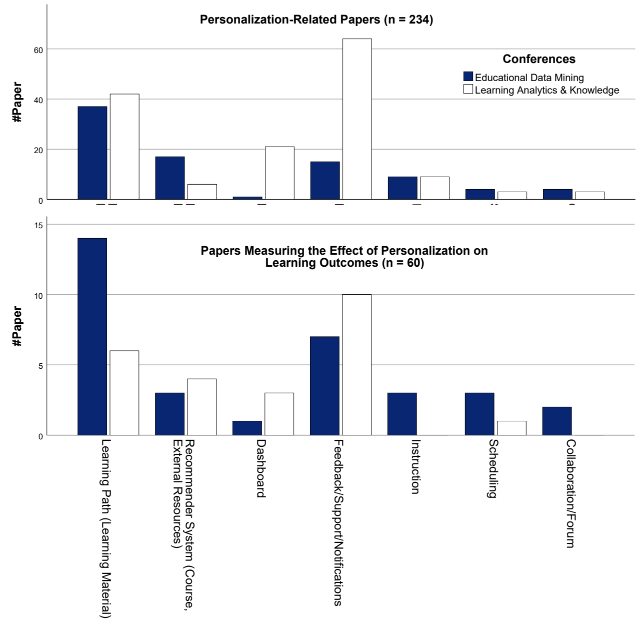

Figure 2: Personalization types in EDM and LAK papers.

5.2 RQ2: Which learner- and context-level indicators are used to drive personalization in EDM and LAK?

Across included papers, we identified 12 indicator categories to which each paper was linked:

- **Outcomes & Achievement**: Observable end results such as grades, test scores, pass/fail, completion, and proficiency.
- **Knowledge/Mastery State**: What the learner knows now and how it changes - KC mastery/ability estimates, probability of correctness, learning/retention.
- **Assessment & Item Properties**: Features of tasks/items - difficulty, discrimination, format, Q-matrix/skills mapping, steps/subparts.
- **Temporal Pacing & Workload**: When and how long - start/submit vs. deadlines, time-on-task, lags between sessions, workload over time.
- **Behavioral Interaction & Engagement**: Interaction traces - clicks, views, plays/pauses, navigation, resource use, attendance, dwell time.
- **Strategy, Process & Help-Seeking**: How learners work - attempt patterns, hints/scaffolds/JIT use, SRL signals (planning / monitoring), path / strategy metrics (e.g., TTS / CFA).
- **Learner Profile & Context**: Who the learner is - demographics, prior achievement / knowledge, goals / motivation, affect / anxiety, interests, availability.
- **Social & Community Signals**: Peer/community activity - posts, comments, messages, reactions, collaboration / co-authoring, centrality, participation roles.
- **Content / Curriculum**: Structure/semantics of learning materials and alignment - concepts / topics, tags, prerequisites, knowledge graphs, similarity / relevance.
- **Model / Policy Signals & Diagnostics**: System-side predictions and controls - predicted risk / performance, utilities / importance, bandit / adaptive policies, latent weights / biases, model diagnostics.
- **Work Products & Artifact Features**: Properties of learner-produced artifacts - writing (linguistic / structural features), code (tokens / AST / correctness), revisions / edits, talk / multimedia quality.
- **Records & Administrative Data**: Institutional / operational records - attendance / absences, mobility / enrollment, suspensions, assignment status, calendars / notifications, merged official logs.

To test indicator differences between EDM and LAK, we employed the same procedure as for RQ1. Levene's test was only significant for "Knowledge/Mastery State" ($p = .017$). However, across all indicators, there was no statistically significant difference in the independent samples t-tests or Welch tests when examining all personalization-related papers (Fig. 3).

Considering only papers that measured the effect of personalization on learning outcomes ($n = 60$), Levene's tests were significant for "Temporal Pacing & Engagement" only ($p \leq .001$). Again, no statistically significant differences were identified. This answers RQ2: there are no statistically reliable differences in the indicators used for personalization in EDM and LAK within the corpus, either overall or within the learning-outcomes subset.

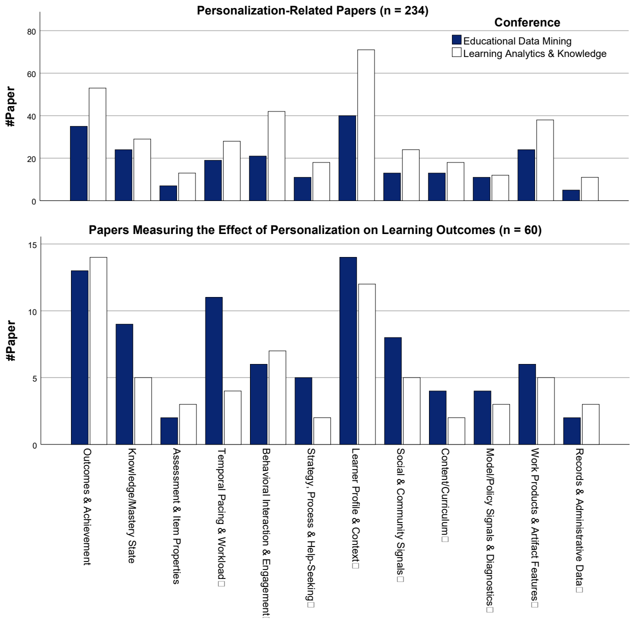

Figure 3: Personalization indicators in EDM and LAK papers.

5.3 RQ3: What research gaps and future directions emerge across the two communities?

To address RQ3, we analyzed the limitations and future-work statements reported in the included papers and synthesized them into recurrent themes. We report themes that appear across both communities, followed by community-leaning emphases.

Cross-cutting limitations (both communities) - Across EDM and LAK, reported limitations converged on threats to external validity and data quality. Many studies relied on small, context-bound samples or single-course deployments and noted limited generalizability to other settings and populations [1, 6, 27]. Cold-start and sparsity issues were frequently highlighted, especially for new learners, low-activity learners, or new content where personalization signals are weak [20, 12]. Studies also reported measurement noise, incomplete outcomes, and missing ground truth for learning constructs, which constrained robust evaluation and cross-study comparability [88, 1]. These recurring issues collectively limit confidence in transferability and reduce the feasibility of cumulative evidence building.

LAK-leaning limitations: adoption and ecological validity - Within LAK, limitations more often problematized ecological validity and adoption: platform- or course-specific interventions that do not generalize; institutional constraints on integration and data flows; instructor workload and interpretation burdens; and low voluntary engagement/compliance when interventions are deployed in authentic settings [1, 6, 15, 27, 49]. These limitations often framed personalization as a socio-technical intervention that depends on organizational fit, teacher workflow, and sustained user uptake rather than on predictive accuracy alone.

EDM-leaning limitations: model complexity and weak causal validation - Within EDM, limitations were predominantly model-centric. Papers emphasized parameter tuning difficulty, computational complexity, and overfitting risks (especially for deep or KT-style models), as well as performance degradation under distribution shift [88, 9, 25]. Another recurring theme was the gap between offline evaluation and causal impact: algorithmic improvements measured on logs or counterfactual simulations were often not validated through randomized/causal designs in authentic learning environments [38, 9].

Cross-cutting future directions (both communities) - Future-work statements commonly called for stronger evaluation and more realistic deployment contexts: online experimentation and adaptive policies embedded in courses; richer data beyond clickstream (e.g., multimodal traces); and clearer reporting standards to improve reproducibility and benchmarking [5, 41, 47, 42]. Ethical and governance themes also appeared, including fairness, privacy, and responsible use of personalization, alongside interoperability needs for cross-platform integration [33, 47, 43].

LAK-leaning future directions: human-centered orchestration and infrastructure - LAK papers more frequently emphasized moving from analytics to action through situated, human-centered interventions. Recurrent directions included teacher-in-the-loop orchestration (dashboards and workflow support), real-time adaptation in authentic contexts, interoperability and standards, and causal evaluation designs that move beyond correlational evidence [5, 11, 29, 35, 4, 15, 24, 28, 47]. Multimodal data and richer recommendation/sequencing support were also recurring motifs [41, 14, 3, 30, 46, 65].

EDM-leaning future directions: scalable policy optimization and benchmarking - EDM proposals concentrated on recommendation and curriculum sequencing, representation learning, and sequence / knowledge modeling (often extending KT) to improve personalization. Many papers also pointed to benchmarking, replicability, and robustness testing to strengthen external validity, and a growing emphasis on fairness / privacy and human-centeredness in adaptive systems [2, 23, 55, 75, 77, 9, 16, 53, 71, 82, 19, 40, 62, 69]. Overall, these results indicate that both communities articulate similar high-level needs (generalization, data quality, evaluation), but diverge in emphasis: LAK foregrounds adoption and socio-technical integration, while EDM foregrounds algorithmic generalization and policy/model validation.

6. DISCUSSION AND FUTURE WORK

The results paint a picture of two communities that are aligned in aspiration but differentiated in how personalization is *realized*, *justified*, and *validated*. This discussion interprets the three results in relation to the paper's central goal: clarifying how personalization is enacted and evaluated in EDM and LAK, and what that implies for credible progress.

6.1 Interpreting differences in personalization targets (RQ1)

The observed differences in focus can be read as a division of labor between *actionability* and *optimization*. Dashboards and feedback mechanisms are often designed around human interpretability and immediate use in learning/teaching workflows, which aligns with LAK's longstanding emphasis on making analytics actionable. Recommender systems and sequencing policies foreground algorithmic decision-making under uncertainty, fitting EDM's tradition of optimizing interventions given data and constraints. Importantly, these are not competing endpoints but complementary components of personalization at scale: optimized policies that are not interpretable or adoptable may fail in authentic settings, while actionable interfaces without validated adaptive mechanisms risk becoming advisory overlays.

A key nuance is that the learning-outcomes subset is small enough that statistical differences disappear even when descriptive differences remain. This suggests a practical constraint on the field's evidence base: personalization is frequently implemented and studied, but comparatively rarely evaluated with outcome measures in a way that supports stable comparisons across venues. Strengthening the outcome-evaluation base is therefore not merely a methodological preference; it is a prerequisite for cumulative knowledge.

6.2 Why indicators look similar across communities (RQ2)

Despite the differences in personalization *targets*, indicator usage does not differ statistically between EDM and LAK. One interpretation is that both communities draw from a shared measurement toolkit (performance, behavioral traces, context/profile, and model-derived signals), especially given overlapping platforms and datasets. Another interpretation is that the indicator taxonomy is broad enough to absorb community differences at a finer granularity: for example, both communities may use "behavioral interaction" indicators, but for different purposes (policy learning vs. teacher-facing explanation), or with different operationalizations (raw logs vs. engineered SRL constructs). In other words, the lack of indicator differences at category level should not be taken to mean that measurement practices are identical; rather, it highlights that divergence may lie in *how* signals are modeled, validated, and translated into adaptation.

This result also clarifies an opportunity: if both communities already rely on similar signal families, then shared reporting standards (feature definitions, preprocessing, missingness handling, construct validity arguments) could significantly improve reproducibility and cross-study comparability without requiring convergence on a single methodological paradigm.

6.3 From "gaps" to research programs (RQ3)

Research gaps are empirically grounded in what authors themselves repeatedly report. The discussion here focuses on what those recurring themes imply for a forward research program.

First, the cross-cutting limitations (context-bound samples, sparsity/cold-start, noisy and incomplete outcomes) point to a deeper structural issue: personalization research often advances through *bespoke* deployments and datasets, which hinders external validity. This suggests that progress depends on infrastructure-level changes: shared benchmarks where appropriate, multi-site studies, and more consistent definitions of what constitutes "personalization" and "learning outcomes" across contexts. Without these, even strong individual studies remain difficult to synthesize.

Second, the community-leaning emphases are complementary and could be integrated into more powerful evaluation and design patterns. LAK's adoption/ecological-validity concerns indicate that personalization should be evaluated as a socio-technical intervention embedded in workflows, constraints, and incentives. EDM's model-centric concerns indicate that personalization should be stress-tested for robustness, generalization, and policy validity beyond offline gains. A combined program would therefore evaluate personalization *end-to-end*: from signal quality and model robustness, through interface and orchestration, to causal impact on learning.

Third, the repeated call for stronger causal validation suggests that "better models" and "better dashboards" are insufficient if research designs cannot distinguish personalization effects from confounds. Pragmatic experimentation (e.g., online A/B tests, stepped-wedge rollouts, teacher-mediated randomization) and careful measurement design (validated outcomes, delayed post-tests, or competency-proximal assessments) are needed to turn personalization into reliable practice. The rarity of learning-outcomes evaluations in the corpus implies a field-level bottleneck: many studies stop at plausibility or engagement metrics, leaving learning impact under-evidenced.

Finally, ethics and governance themes (fairness, privacy, explainability) appear as an emergent but not yet uniformly operationalized agenda. A practical implication is that these considerations should be connected to the same objects / indicators / techniques framework used in this review. For example, fairness risks differ between a course recommender and a mastery-based sequencer; privacy risks differ between administrative records and multimodal sensing; explainability requirements differ between teacher-facing dashboards and learner-facing feedback. Treating ethics as design- and context-dependent, rather than as a generic checklist, is likely to make it more actionable across both communities.

6.4 Implications for cross-fertilization

Taken together, the results suggest a productive synthesis: EDM contributes scalable modeling and policy optimization methods, while LAK contributes human-centered orchestration, interpretability, and deployment realism. Cross-fertilization could be accelerated by (i) adopting shared reporting standards for personalization objects, signals, and evaluation designs; (ii) building evaluation pipelines that explicitly connect offline metrics to online impact; and (iii) designing systems where policy learning and human oversight are co-designed rather than bolted together. Such work would directly address the dominant limitations reported in both communities while preserving their strengths.

7. LIMITATIONS

This review offers a structured, comparative synthesis of personalization research in EDM and LAK between 2015 and 2025, but several limitations constrain the interpretation and generalization of the findings. The corpus was intentionally limited to papers published in the LAK and EDM conference proceedings to support direct comparisons between two influential communities. While this restriction improves internal coherence, it also narrows coverage by excluding relevant contributions published in journals and adjacent venues where personalization is frequently studied and where longer-term deployments, multi-institution evaluations, and theory-building work are often reported. Consequently, the patterns identified here should be understood as reflecting the proceedings-centered discourse of these two conference ecosystems rather than the full landscape of educational personalization research.

Our retrieval strategy employed a single high-recall lexical stem (`personali*`) to ensure auditability and capture common spelling and morphological variants. Although this choice strengthens reproducibility, it can lead to both false negatives and false positives. Relevant work may be missed when personalization is instantiated under alternative terminology (e.g., adaptive tutoring, individualized sequencing, targeted intervention, mastery-based progression, or recommendation framed without explicit "personalization" language), and records may be retrieved where personalization is invoked primarily as motivation rather than as an implemented mechanism. Accordingly, the resulting corpus depends not only on the search term but also on subsequent screening decisions.

The review further relies on an operational definition of personalization as system-driven, individual-level adaptation (or explicitly defined subgroups) based on learner-specific information. This definition provides conceptual clarity, but it may exclude studies in which personalization is enacted primarily through human mediation (e.g., instructor decision-making supported by analytics) or where personalization is realized as infrastructure, workflow integration, or institutional practice rather than as an explicit adaptive mechanism in software. Borderline cases are inevitable in a heterogeneous literature, and although adjudication procedures were used, some degree of classification error remains possible. Similarly, the second-stage distinction between studies that measure learning outcomes and those that report engagement, usability, or satisfaction requires interpretive judgment because papers do not always report outcomes with sufficient construct validity arguments or consistent operationalization. Proxies such as completion or time-on-task may be treated as learning-proximal in some contexts but not in others, and ambiguity in reporting can affect categorization.

The coding scheme necessarily abstracts across substantial within-category heterogeneity. Targets of personalization and indicator families were treated as non-exclusive categories, reflecting the fact that many systems combine mechanisms and signals. However, this design complicates inference because category frequencies do not correspond to mutually exclusive partitions, and coarse taxonomies can mask differences in how signals are operationalized and validated. For example, "behavioral interaction" may denote raw clickstream counts, engineered self-regulated learning constructs, or multimodal engagement indicators; "knowledge/mastery state" may arise from classical cognitive models, deep knowledge tracing variants, or hybrid approaches. As a result, small differences at the category level should not be interpreted as equivalence in measurement practice or modeling choices.

Quantitative comparisons are based on counts of papers coded into categories and employ standard significance tests. This approach treats papers as independent observations and does not explicitly account for dependencies such as recurring authorship, shared datasets, or coordinated research programs that may cluster publication patterns. Moreover, the subset of included studies that evaluate learning outcomes is comparatively small, limiting statistical power to detect stable differences between communities in that subset. Non-significant findings in the learning-outcomes analyses should therefore be interpreted cautiously as insufficient evidence for differences rather than evidence of no differences.

Finally, this work is not a meta-analysis and does not estimate pooled effect sizes. The included studies vary widely in context, intervention types, outcome measures, duration, and evaluation designs, ranging from offline analyses and observational studies to quasi-experiments and randomized evaluations. This heterogeneity limits the appropriateness of aggregating effects and complicates causal interpretation, particularly when reports omit information needed to assess internal validity (e.g., randomization fidelity, attrition, contamination, or multiple testing control). In addition, the synthesis of limitations and future-work statements depends on what authors choose to report and may be affected by reporting bias, page constraints, and selective emphasis, potentially underrepresenting negative results, failed deployments, or governance constraints such as privacy and fairness trade-offs. Taken together, these considerations indicate that the present review is best read as a transparent, reproducible map of how personalization is operationalized and evaluated in LAK and EDM proceedings during 2015–2025, rather than as an exhaustive census of all personalization research or a definitive causal estimate of personalization effectiveness.

8. CONCLUSION

EDM and LAK both pursue personalization, but they do so through different lenses, as this literature review has examined. This review compared how EDM and LAK operationalize personalization (2015 – 2025) by what is adapted, the indicators that guide personalization, and how effects are evaluated. We find distinct foci - LAK leans toward feedback and dashboards, EDM toward recommender/sequencing, but no systematic differences in indicator usage. Shared limitations include context-bound samples, data sparsity, and measurement noise; LAK stresses adoption/ecological validity, while EDM emphasizes model complexity and weak causal validation. Progress hinges on combining EDM's policy/model advances with LAK's human-centered orchestration, supported by practice-grounded trials, cross-platform benchmarks, and clear reporting of objects, indicators, and evaluation designs.

REFERENCES

[1] S. Abdi, H. Khosravi, S. Sadiq, and D. Gasevic. Complementing educational recommender systems with open learner models. In *Proceedings of the tenth international conference on learning analytics & knowledge*, pages 360–365, 2020.

[2] K. Abhinav, V. Subramanian, A. Dubey, P. Bhat, and A. D. Venkat. Lecore: A framework for modeling learner's preference. In *Proceedings of the 11th International Conference on Educational Data Mining, Buffalo NY*, 2018.

[3] S. A. Adjei, A. F. Botelho, and N. T. Heffernan. Sequencing content in an adaptive testing system: the role of choice. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pages 178–182, 2017.

[4] K. Aghaei, M. Hatala, and A. Mogharrab. How students' emotion and motivation changes after viewing dashboards with varied social comparison group: A qualitative study. In *LAK23: 13th International Learning Analytics and Knowledge Conference*, pages 663–669, 2023.

[5] A. P. Aguinalde and J. Shin. Talking in sync: How linguistic synchrony shapes teacher-student conversation in english as a second language tutoring environment. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, pages 395–406, 2025.

[6] Q. U. Ain, M. A. Chatti, P. A. Meteng Kamdem, R. Alatrash, S. Joarder, and C. Siepmann. Learner modeling and recommendation of learning resources using personal knowledge graphs. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pages 273–283, 2024.

[7] N. Alam, B. Mostafavi, S. Tithi, M. Chi, and T. Barnes. How much training is needed? reducing training time using deep reinforcement learning in an intelligent tutor. In *Proceedings of the 17th International Conference on Educational Data Mining*, 2024.

[8] H. Alamri, V. Lowell, W. Watson, and S. L. Watson. Using personalized learning as an instructional approach to motivate learners in online higher education: Learner self-determination and intrinsic motivation. *Journal of Research on Technology in Education*, 52(3):322–352, 2020.

[9] H. Almoubayyed, S. Fancsali, and S. Ritter. Generalizing predictive models of reading ability in adaptive mathematics software. In *Proceedings of the 16th International Conference on Educational Data Mining*, 2023.

[10] K. T. Arnesen, C. R. Graham, C. R. Short, and D. Archibald. Experiences with personalized learning in a blended teaching course for preservice teachers. *Journal of online learning research*, 5(3):275–310, 2019.

[11] S. Aslan, E. Okur, N. Alyuz, S. E. Mete, E. Oktay, U. Genc, and A. A. Esme. Students' emotional self-labels for personalized models. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pages 550–551, 2017.

[12] D. Azcona, P. Arora, I.-H. Hsiao, and A. Smeaton. user2code2vec: Embeddings for profiling students based on distributional representations of source code. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 86–95, 2019.

[13] C. Baek and T. Doleck. Educational data mining versus learning analytics: A review of publications from 2015 to 2019. *Interactive Learning Environments*, 31(6):3828–3850, 2023.

[14] M. Bannert, I. Molenaar, R. Azevedo, S. Järvelä, and D. Gašević. Relevance of learning analytics to measure and support students' learning in adaptive educational technologies. In *proceedings of the seventh international learning analytics & knowledge conference*, pages 568–569, 2017.

[15] C. Borchers, J. Ooge, C. Peng, and V. Aleven. How learner control and explainable learning analytics about skill mastery shape student desires to finish and avoid loss in tutored practice. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, pages 810–816, 2025.

[16] C. Botta, A. Segal, and K. Gal. Sequencing educational content using diversity aware bandits. In *Proceedings of the 16th International Conference on Educational Data Mining*, 2023.

[17] P. Brusilovsky, J. Eklund, and E. Schwarz. Web-based education for all: a tool for development adaptive courseware. *Computer networks and ISDN systems*, 30(1-7):291–300, 1998.

[18] P. Brusilovsky and C. Peylo. Adaptive and intelligent web-based educational systems. *International journal of artificial intelligence in education*, 13(2-4):159–172, 2003.

[19] S. Bulathwela, M. Verma, M. Pérez-Ortiz, E. Yilmaz, and J. Shawe-Taylor. Can population-based engagement improve personalisation? a novel dataset and experiments. *arXiv preprint arXiv:2207.01504*, 2022.

[20] H. Bydzovská. Course enrollment recommender system. In *Educational Data Mining*, 2016.

[21] N. Chakraborty, S. Roy, W. L. Leite, M. K. S. Faradonbeh, and G. Michailidis. The effects of a personalized recommendation system on students' high-stakes achievement scores: A field experiment. *International Educational Data Mining Society*, 2021.

[22] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs. A reference model for learning analytics. *International journal of Technology Enhanced learning*, 4(5-6):318–331, 2012.

[23] W. Chen, C. Joe-Wong, C. G. Brinton, L. Zheng, and D. Cao. Principles for assessing adaptive online courses. *International Educational Data Mining Society*, 2018.

[24] A. Claassen, N. Mirriahi, V. Kovanović, and S. Dawson. From data to design: Integrating learning analytics into educational design for effective decision-making: From data to design. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, pages 558–567, 2025.

[25] B. Clement, P.-Y. Oudeyer, and M. Lopes. A comparison of automatic teaching strategies for heterogeneous student populations. In *EDM 16-9th international conference on educational data mining*,

2016.

[26] D. Davis, I. Jivet, R. F. Kizilcec, G. Chen, C. Hauff, and G.-J. Houben. Follow the successful crowd: raising mooc completion rates through social comparison at scale. In *Proceedings of the seventh international learning analytics & knowledge conference*, pages 454–463, 2017.

[27] S. Dawson, A. Pardo, F. Salehian Kia, and E. Panadero. An integrated model of feedback and assessment: From fine grained to holistic programmatic review. In *LAK23: 13th International Learning Analytics and Knowledge Conference*, pages 579–584, 2023.

[28] E. De Quincey, C. Briggs, T. Kyriacou, and R. Waller. Student centred design of a learning analytics system. In *Proceedings of the 9th international conference on learning analytics & knowledge*, pages 353–362, 2019.

[29] H. Deininger, C. Parrisius, R. Lavelle-Hill, D. Meurers, U. Trautwein, B. Nagengast, and G. Kasneci. Who did what to succeed? individual differences in which learning behaviors are linked to achievement. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, pages 771–782, 2025.

[30] D. Di Mitri, M. Scheffel, H. Drachsler, D. Börner, S. Ternier, and M. Specht. Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. In *Proceedings of the seventh international learning analytics & knowledge conference*, pages 188–197, 2017.

[31] K. C. Dieter, J. Studwell, and K. P. Vanacore. Differential responses to personalized learning recommendations revealed by event-related analysis. *International Educational Data Mining Society*, 2020.

[32] S. Dormezil, T. M. Khoshgoftaar, and F. Robinson-Bryant. Differentiating between educational data mining and learning analytics: A bibliometric approach. In *EDM (Workshops)*, pages 17–22, 2019.

[33] S. Doroudi and E. Brunskill. Fairer but not fair enough on the equitability of knowledge tracing. In *Proceedings of the 9th international conference on learning analytics & knowledge*, pages 335–339, 2019.

[34] A. Downer, A. Shapoval, O. Vysotska, I. Yuryeva, and T. Bairachna. Us e-learning course adaptation to the ukrainian context: lessons learned and way forward. *BMC medical education*, 18(1):247, 2018.

[35] L. Faucon, J. K. Olsen, and P. Dillenbourg. A bayesian model of individual differences and flexibility in inductive reasoning for categorization of examples. In *Proceedings of the tenth international conference on learning analytics & knowledge*, pages 285–294, 2020.

[36] J. Frej, N. Shah, M. Knezevic, T. Nazaretsky, and T. Käser. Finding paths for explainable mooc recommendation: A learner perspective. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pages 426–437, 2024.

[37] B. Garrick, D. Pendergast, and D. Geelan. Introduction to the philosophical arguments underpinning personalised education. In *Theorising Personalised Education: Electronically Mediated Higher Education*, pages 1–16. Springer, 2016.

[38] J. P. González-Brenes and Y. Huang. " your model is predictive–but is it useful?" theoretical and empirical considerations of a new paradigm for adaptive tutoring evaluation. *International Educational Data Mining Society*, 2015.

[39] A. Gurung, S. Baral, K. P. Vanacore, A. A. Mcreynolds, H. Kreisberg, A. F. Botelho, S. T. Shaw, and N. T. Hefferna. Identification, exploration, and remediation: Can teachers predict common wrong answers? In *LAK23: 13th International Learning Analytics and Knowledge Conference*, pages 399–410, 2023.

[40] J. He-Yueya and A. Singla. Quizzing policy using reinforcement learning for inferring the student knowledge state. In *14th International Conference on Educational Data Mining*, pages 533–539. educationaldatamining. org, 2021.

[41] K. Holstein, G. Hong, M. Tegene, B. M. McLaren, and V. Aleven. The classroom as a dashboard: Co-designing wearable cognitive augmentation for k-12 teachers. In *Proceedings of the 8th international conference on learning Analytics and knowledge*, pages 79–88, 2018.

[42] Q. Hu and H. Rangwala. Reliable deep grade prediction with uncertainty estimation. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 76–85, 2019.

[43] M. Huptych, M. Bohuslavek, M. Hlosta, and Z. Zdrahal. Measures for recommendations based on past students' activity. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pages 404–408, 2017.

[44] P. Hur, H. Lee, S. Bhat, and N. Bosch. Using machine learning explainability methods to personalize interventions for students. *International Educational Data Mining Society*, 2022.

[45] G. D. Jaiyeola, A. Y. Wong, R. L. Bryck, C. Mills, and S. Hutt. One size does not fit all: Considerations when using webcam-based eye tracking to models of neurodivergent learners' attention and comprehension. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, pages 24–35, 2025.

[46] W. Jiang, Z. A. Pardos, and Q. Wei. Goal-based course recommendation. In *Proceedings of the 9th international conference on learning analytics & knowledge*, pages 36–45, 2019.

[47] K. Kitto, N. Sarathy, A. Gromov, M. Liu, K. Musial, and S. Buckingham Shum. Towards skills-based curriculum analytics: Can we automate the recognition of prior learning? In *Proceedings of the tenth international conference on learning analytics & knowledge*, pages 171–180, 2020.

[48] S. Kopeinik, E. Lex, P. Seitlinger, D. Albert, and T. Ley. Supporting collaborative learning with tag recommendations: a real-world study in an inquiry-based classroom project. In *Proceedings of the seventh international learning analytics & knowledge conference*, pages 409–418, 2017.

[49] W. L. Leite, S. Roy, N. Chakraborty, G. Michailidis, A. C. Huggins-Manley, S. D'Mello, M. K.

Shirani Faradonbeh, E. Jensen, H. Kuang, and Z. Jing. A novel video recommendation system for algebra: An effectiveness evaluation study. In *LAK22: 12th International learning analytics and knowledge conference*, pages 294–303, 2022.

[50] B. Levy, A. Hershkovitz, O. Tzayada, O. Ezra, A. Segal, K. Gal, A. Cohen, and M. Tabach. Teacher vs. algorithm: Double-blind experiment of content sequencing in mathematics. In *12th International Conference on Educational Data Mining, EDM 2019*, pages 603–606. International Educational Data Mining Society, 2019.

[51] H. Li, W. Xing, C. Li, W. Zhu, B. Lyu, F. Zhang, and Z. Liu. Who should be my tutor? analyzing the interactive effects of automated text personality styles between middle school students and a mathematics chatbot. In *Proceedings of the 15th international learning analytics and knowledge conference*, pages 910–917, 2025.

[52] T. Li, D. Nath, Y. Cheng, Y. Fan, X. Li, M. Raković, H. Khosravi, Z. Swiecki, Y.-S. Tsai, and D. Gašević. Turning real-time analytics into adaptive scaffolds for self-regulated learning using generative artificial intelligence. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, pages 667–679, 2025.

[53] Z. Li, L. Yee, N. Sauerberg, I. Sakson, J. J. Williams, and A. N. Rafferty. Getting too personal (ized): The importance of feature choice in online adaptive algorithms. *International Educational Data Mining Society*, 2020.

[54] L.-A. Lim, D. Gasevic, W. Matcha, N. Ahmad Uzir, and S. Dawson. Impact of learning analytics feedback on self-regulated learning: Triangulating behavioural logs with students' recall. In *LAK21: 11th international learning analytics and knowledge conference*, pages 364–374, 2021.

[55] B. Ma, G. P. Hettiarachchi, and Y. Ando. Format-aware item response theory for predicting vocabulary proficiency. In *Proceedings of the 15th International Conference on Educational Data Mining*, page 695, 2022.

[56] B. Ma, Y. Taniguchi, and S. Konomi. Course recommendation for university environments. *International educational data mining society*, 2020.

[57] R. Matz, K. Schulz, E. Hanley, H. Derry, B. Hayward, B. Koester, C. Hayward, and T. McKay. Analyzing the efficacy of ecoach in supporting gateway course success through tailored support. In *LAK21: 11th international learning analytics and knowledge conference*, pages 216–225, 2021.

[58] F. Mi and B. Faltings. Adaptive sequential recommendation for discussion forums on moocs using context trees. *International Educational Data Mining Society*, 2017.

[59] I. Molenaar, A. Horvers, R. Dijkstra, and R. S. Baker. Personalized visualizations to promote young learners' srl: The learning path app. In *Proceedings of the tenth international conference on learning analytics & knowledge*, pages 330–339, 2020.

[60] L. Mouchel, T. Wambsganss, P. Mejia-Domenzain, and T. Käser. Understanding revision behavior in adaptive writing support systems for education. 2023.

[61] A. Muhammad, Q. Zhou, G. Beydoun, D. Xu, and J. Shen. Learning path adaptation in online learning systems. In *2016 IEEE 20th international conference on computer supported cooperative work in design (CSCWD)*, pages 421–426. IEEE, 2016.

[62] R. Murali, C. Conati, and D. Poole. A comparison of real-time user classification methods using interaction data for open-ended learning. 2025.

[63] I. Musabirov, M. Reza, H. Song, S. Moore, P. Chen, H. Kumar, T. Li, J. Stamper, N. Bier, A. Rafferty, et al. Platform-based adaptive experimental research in education: Lessons learned from the digital learning challenge. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, pages 13–23, 2025.

[64] T. NeCamp, J. Gardner, and C. Brooks. Beyond a/b testing: Sequential randomization for developing interventions in scaled digital learning environments. In *Proceedings of the 9th International Conference on learning analytics & knowledge*, pages 539–548, 2019.

[65] S. Nicoll, K. Douglas, and C. Brinton. Giving feedback on feedback: An assessment of grader feedback construction on student performance. In *LAK22: 12th International Learning Analytics and Knowledge Conference*, pages 239–249, 2022.

[66] J. W. Orr and N. Russell. Automatic assessment of the design quality of python programs with personalized feedback. *International Educational Data Mining Society*, 2021.

[67] A. Ozturk, R. Schmucker, T. Mitchell, and A. T. Kumtepe. Heterogeneous treatment effects of learning analytics dashboards: Do all learners benefit equally? In *Proceedings of the 18th International Conference on Educational Data Mining*, pages 418–426, 2025.

[68] Z. A. Pardos and W. Jiang. Designing for serendipity in a university course recommendation system. In *Proceedings of the tenth international conference on learning analytics & knowledge*, pages 350–359, 2020.

[69] T. Patikorn, N. T. Heffernan, and J. Zou. An offline evaluation method for individual treatment rules and how to find heterogeneous treatment effect. In *Workshop and Tutorials Chairs*, page 390. ERIC, 2017.

[70] J. Pesonen, O.-M. Palo-oja, and K. Ng. Personalized learning analytics-based feedback on a self-paced online course. In *Companion Proceedings of the 11th International Conference on Learning Analytics and Knowledge (LAK'21): The impact we make: The contributions of learning analytics to learning: April 12-16, 2021, Online, Everywhere*, pages 13–14. Society for Learning Analytics Research, 2021.

[71] T. Phung, N. Kotalwar, M. Liut, J. Leinonen, P. Denny, and A. Singla. Humanizing automated programming feedback: Fine-tuning generative models with student-written feedback. 2025.

[72] V. Prakash, R. Bansal, and S. Sharma. Quest-genius: An ai-driven personalised learning platform bridging educational gaps for secondary students - a quasi-experimental study on academic performance, engagement, and equity. In C. Mills, G. Alexandron, D. Taibi, G. L. Bosco, and L. Paquette, editors,

Proceedings of the 18th International Conference on Educational Data Mining, pages 630–634, Palermo, Italy, July 2025. International Educational Data Mining Society.

[73] Praseeda, S. Srinivasa, and P. Ram. Validating the myth of average through evidences. In *Proceedings of the 11th International Conference on Educational Data Mining, Buffalo NY*, 2019.

[74] E. Prihar, M. Syed, K. Ostrow, S. Shaw, A. Sales, and N. Heffernan. Exploring common trends in online educational experiments. In *Proceedings of the 15th International Conference on Educational Data Mining*, 2022.

[75] A. N. Rafferty, R. Jansen, and T. L. Griffiths. Using inverse planning for personalized feedback. *EDM*, 16:472–477, 2016.

[76] M. J. Rodríguez-Triana, L. P. Prieto, A. Martínez-Monés, J. I. Asensio-Pérez, and Y. Dimitriadis. The teacher in the loop: Customizing multimodal learning analytics for blended learning. In *Proceedings of the 8th international conference on learning analytics and knowledge*, pages 417–426, 2018.

[77] J. Rollinson and E. Brunskill. From predictive models to instructional policies. *International Educational Data Mining Society*, 2015.

[78] I. Rushkin, Y. Rosen, A. M. Ang, C. Fredericks, D. Tingley, M. J. Blink, and G. Lopez. Adaptive assessment experiment in a harvardx mooc. In *EDM*, 2017.

[79] S. Rüdian. Personalization in EDM and LAK - Dataset, 2026.

[80] A. Schütt, T. Huber, I. Aslan, and E. André. Fast dynamic difficulty adjustment for intelligent tutoring systems with small datasets. 2023.

[81] F. Sense, M. Krusmark, J. Fiechter, M. G. Collins, L. Sanderson, J. Onia, and T. Jastrzembski. Combining cognitive and machine learning models to mine cpr training histories for personalized predictions. *International Educational Data Mining Society*, 2021.

[82] H. Shaikh, A. Modiri, J. J. Williams, and A. N. Rafferty. Balancing student success and inferring personalized effects in dynamic experiments. In *Proceedings of the 11th International Conference on Educational Data Mining, Buffalo NY*, 2019.

[83] K. Thaker, L. Zhang, D. He, and P. Brusilovsky. Recommending remedial readings using student knowledge state. *International Educational Data Mining Society*, 2020.

[84] D. R. Thomas, J. Lin, E. Gatz, A. Gurung, S. Gupta, K. Norberg, S. E. Fancsali, V. Aleven, L. Branstetter, E. Brunskill, et al. Improving student learning with hybrid human-ai tutoring: A three-study quasi-experimental investigation. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pages 404–415, 2024.

[85] S. Thomas. Future ready learning: Reimagining the role of technology in education. 2016 national education technology plan. *Office of Educational Technology, US Department of Education*, 2016.

[86] J. Turnbull, D. Lea, D. Parkinson, P. Phillips, B. Francis, S. Webb, V. Bull, and M. Ashby. Oxford advanced learner's dictionary. *International Student's Edition*, 2010.

[87] O. Vainas, Y. Ben-David, R. Gilad-Bachrach, M. Ronen, O. Bar-Ilan, R. Shillo, G. Lukin, and D. Sitton. Staying in the zone: Sequencing content in classrooms based on the zone of proximal development. In *12th International Conference on Educational Data Mining, EDM 2019*, pages 659–662. International Educational Data Mining Society, 2019.

[88] J. Vassoyan, J.-J. Vie, and P. Lemberger. Towards scalable adaptive learning with graph neural networks and reinforcement learning. In M. Feng, T. Käser, and P. Talukdar, editors, *Proceedings of the 16th International Conference on Educational Data Mining*, pages 351–361, Bengaluru, India, July 2023. International Educational Data Mining Society.

[89] J. Woodrow, C. Piech, and S. Koyejo. Improving generative ai student feedback: Direct preference optimization with teachers in the loop. In *Proceedings of the 18th International Conference on Educational Data Mining*, pages 442–449, 2025.

[90] H. Yu, C. Miao, C. Leung, and T. J. White. Towards ai-powered personalization in mooc learning. *npj Science of Learning*, 2(1):15, 2017.

[91] A. Zavaleta Bernuy, Z. Han, H. Shaikh, Q. Y. Zheng, L.-A. Lim, A. Rafferty, A. Petersen, and J. J. Williams. How can email interventions increase students' completion of online homework? a case study using a/b comparisons. In *LAK22: 12th International Learning Analytics and Knowledge Conference*, pages 107–118, 2022.

[92] M. Zhang, J. Hao, P. Deane, and C. Li. Defining personalized writing burst measures of translation using keystroke logs. In *Proceedings of the 11th International Conference on Educational Data Mining, Buffalo NY*, pages 549–552, 2018.

[93] H. Zo. Personalization vs. customization: Which is more effective in e-services? 2003.