DTU Library

Methods for Evaluating Learning Analytics and Learning Analytics Dashboards in Adaptive Learning Platforms

A Systematic Review

Tretow-Fish, Tobias Alexander Bang; Khalid, Md Saifuddin

Published in:
Electronic Journal of e-Learning

Link to article, DOI:
10.34190/ejel.21.5.3088

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Methods for Evaluating Learning Analytics and Learning Analytics Dashboards in Adaptive Learning Platforms: A Systematic Review

Tobias Alexander Bang Tretow-Fish and Md. Saifuddin Khalid
Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark

tabtr@dtu.dk
skhalid@dtu.dk

Abstract: This research paper highlights and addresses the lack of a systematic review of the methods used to evaluate Learning Analytics (LA) and Learning Analytics Dashboards (LAD) of Adaptive Learning Platforms (ALPs) in the current literature. Addressing this gap, the authors built upon the work of Tretow-Fish and Khalid (2022) and analyzed 32 papers, which were grouped into six categories (C1-6) based on their themes. The categories include C1) the evaluation of LA and LAD design and framework, C2) the evaluation of user performance with LA and LAD, C3) the evaluation of adaptivity, C4) the evaluation of ALPs through perceived value, C5) the evaluation of Multimodal methods, and C6) the evaluation of the pedagogical implementation of ALP's LA and LAD. The results include a tabular summary of the papers including the categories, evaluation unit(s), methods, variables and purpose. While there are numerous studies in categories C1-4 that focus on the design, development, and impact assessment of ALP's LA and LAD, there are only a few studies in categories C5 and C6. For the category of C5), very few studies applied any evaluation methods assessing the multimodal features of LA and LADs on ALPs. Especially for C6), evaluating the pedagogical implementation of ALP's LA and LAD, the three dimensions of signature pedagogy are used to assess the level of pedagogy evaluation. Findings showed that no studies focus on evaluating the deep or implicit structure of ALP's LA. All studies examine the structural surface dimension of learning activities and interactions between students, teachers, and ALP's LA and LAD, as examined in categories C2-C5. No studies were exclusively categorized as a C6 category, indicating that all studies evaluate ALP's LA and LAD on the surface structure dimension of signature pedagogy. This review highlights the lack of pedagogical methodology and theory in ALP's LA and LAD, which are recommended to be emphasized in future research and ALP development and implementation.

Keywords: Adaptive Learning Platform, Learning Analytics, Systematic literature review, Methods of evaluation

1. Introduction

The field of Adaptive Learning (AL) is a relatively new area of research that spans multiple disciplines and involves numerous synonyms and definitions. While terms such as personalized learning, individualized learning, intelligent tutoring, and customized learning are sometimes used interchangeably, adaptive learning is the most used term (Shemshack and Spector, 2020). Adaptive learning platforms (ALPs) have gained significant attention in recent years, leading to the development of various methods to design and evaluate personalized activities and content. However, assessing ALPs is a complex and multi-dimensional process involving several factors such as Learning Analytics (LA), Learning Analytics Dashboards (LAD), system usability, user perception, use, and pedagogy.

While numerous reviews on adaptive learning and learning analytics exist, few studies have specifically focused on evaluating ALPs. Martin, Denne and Bonk (2020) synthesizes several systematic reviews on adaptive learning and learning analytics. The synthesis on adaptive learning includes: (1) A meta-analysis to address the question: Can students improve their knowledge when the system adapts to their profile and performance? (2) An analysis of 42 studies on source of adaptation focusing on learner and learner environment interaction; analysis of 29 studies on adaptation of content, presentation and instruction; 25 studies on rule-based, probability-based, or other adaptive pathways, (3) Content analysis involving 70 studies addressing learning styles in adaptive educational hypermedia systems. (4) Analysis of 98 studies determining characteristics for learner models in adaptive systems. (5) Document analysis of 78 studies and reviews on learning styles in adaptive systems, (6) Review of 78 studies that explored learner, dimensions of these traits, and identification techniques for these traits in adaptive learning systems, (7) 61 empirical studies were reviewed on adaptive sources based on learner models and adaptive targets based on content and instructional models, and (8) A qualitative thematic analysis of 62 studies and a subset of 12 studies on experimental designs for meta-analysis to study the effects of adaptivity in educational games. The review of reviews on learning analytics (Martin, Dennen and Bonk, 2020) include (1) Analysis of 40 studies identifying the research objectives and methods applied in learning analytics and educational datamining studies, (2) reviewed 44 studies to examine learning analytics methods, benefits, and challenges in higher education, (3) examined 43 studies on applied research design, topic of study,

Reference this paper: Tretow-Fish, T.A.B. and Khalid, M.S., 2023. Methods for Evaluating Learning Analytics and Learning Analytics Dashboards in Adaptive Learning Platforms: A Systematic Review, *The Electronic Journal of e-Learning, 21(5)*, pp 430-449, available online at www.ejel.org

educational context, learning scenario, pedagogical practices, learning platform, technology tools, and methodological techniques, (4) conducted a review on 52 studies focusing on visual learning analytics of educational data identifying approaches, audience, purposes, contexts, and data sources. (5) did a systematic review of 107 studies identifying what data was collected, modeling methods, research themes, system evaluation, and similarities and differences between open learner models and learning analytic dashboards, (6) reviewed 29 papers examining learning analytic dashboards from a self-regulated learning perspective, (7) 11 studies were reviewed to examine the efficacy of learning analytics interventions in higher education which (8) build upon where 46 studies were inspected to address whether and to what extent learning analytics were successful in providing study success in higher education.

Mousavinasab et al. (2021) conducted a review of 53 studies that investigated the variant characteristics of Intelligent Tutoring Systems across various educational fields. In their review, they raised several questions concerning the methods employed to assess these systems. For example, the importance of the learners' role in evaluating intelligent tutoring systems is evident, especially when assessing system usability. Only 5.66% of the reviewed studies evaluated intelligent tutoring systems based solely on learner experiences. However, in studies where learner experiences were combined with other variables, such as learner and system performance, learner experiences were more frequent. The review fails to provide information on the specific methods utilized to obtain the learner experience or the types of usability tests employed. Understanding the various evaluation perspectives and assessment methods utilized in evaluating Learning Analytics (LA) and Learning Analytics Dashboards (LADs) is crucial. The absence of comprehensive insights on evaluation methods for LA and LADs is the driving force behind this systematic review. So, this review expands the work of Tretow-Fish and Khalid (2022) and aims to synthesize evaluation methods used in the design, development, and implementation of ALP's LA and LAD to support pedagogical and learning-related decisions for educators and students. The review will analyze the research design, frameworks, methods, and instruments utilized to evaluate ALP's LA and LAD, the resulting impact, and the integration of students' and educators' perceptions of LAD and LA into the evaluation methods. The research will significantly contribute to the field of usability engineering, user experience, and digital learning technology. Investigating the evaluation methods applied to ALPs is crucial to enhance the quality of the learning experience and outcomes, improving educators' teaching experiences and their technology adoption, aiding the development process in companies, and ensuring the proper implementation of evaluation methods. The scope mentioned above, and motivation led us to devise the research question:

How to evaluate the Learning Analytics and Learning Analytics Dashboards of Adaptive Learning Platforms?

The objective is to identify a set of methods for evaluating the technological features' functionalities and perceived experiences and another set of methods for demonstrating the evidence of improving learning outcomes, learning experience, and teaching quality. The researchers and practitioners will be able to apply the synthesized methods and instruments for the evaluation of learning analytics and dashboards in the contexts of digital platforms and for the assessment of impact. The review will provide insights for identifying the scope of future research by providing an overview of how LA and LADs of ALPs are evaluated, with what purpose, and on which variables. Furthermore, the outcome can advance the field of interaction design, while the latter can contribute to the broader domain of service design and innovation in education and training.

2. Methods

The selection of papers and the process of analysis and synthesis is conducted using two distinct established methods.

2.1 Selection of Papers: PRISMA

The paper selection process follows the four phases of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol (Page *et al.*, 2021), which are identification, screening, eligibility, and inclusion (see Figure 1). To review the evaluation methods utilized for Learning Analytics (LA) and Learning Analytics Dashboards (LAD) on Adaptive Learning Platforms (ALPs), a range of keywords including evaluation, adaptive learning, learning analytics, learning analytics dashboards, assessment, and others are employed in different combinations. The search was limited to articles that had been peer-reviewed and published in English, Danish, and Norwegian (considering the authors' language abilities) from 2011 up to the search deadline of September 1, 2021.

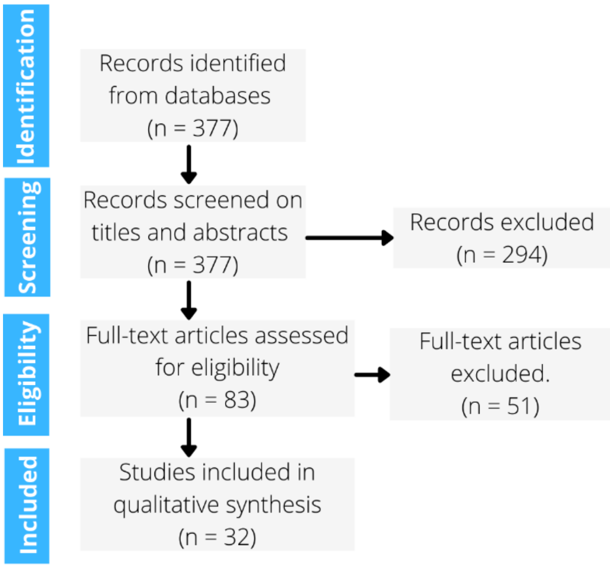

Figure 1: PRISMA Flow-Chart

After testing different keyword combinations and consulting with a librarian, four databases were selected. Various combinations of the keywords returned the following results: Scopus (n=75), ACM (n=144), ScienceDirect (n=106), ERIC (n=14), and Taylor Francis (n=38). The screening and eligibility stages involved applying specific exclusion criteria, which are as follows: (1) Papers that do not primarily investigate the LA or LAD of an ALP were excluded. (2) Papers focusing on LA and LAD in other e-learning environments that do not meet the adaptivity requirements of the learning platform were excluded. (3) Papers without empirical data examining LA or LAD on ALP or those that do not primarily focus on LA and LAD evaluation were excluded. (4) Only included papers published in the main conference proceedings, while workshop papers and posters were excluded. The two authors independently screened different databases, and only the papers selected by one author (n=83) were included in this document. For this review, 32 articles and two reviews were included for analysis and synthesis.

2.2 Constant Comparative Analysis Method

We applied the constant comparative analysis method (Hewitt-Taylor, 2001) for the analysis and synthesis. The articles were encoded according to themes and then divided into categories and subcategories. During this process, the coded sections were regularly compared to similar parts of texts containing the same codes. The intention was to connect the texts and ensure the continuity of the codes' definitions (ibid.). Each included paper was read to identify methods, variables, and purpose of evaluating LA and LAD. The data extracted from the papers are tabulated to synthesize: (1) The methods used when evaluating LA and LAD. (2) Variables measured by the methods to evaluate LA and LAD. (3) The purpose of the evaluation method applied. A thematic analysis was initiated from the identified purposes, and categories were developed.

2.3 Signature Pedagogy for Analytical Thematization in C6

To thematize the category of C6) an evaluation of the pedagogical implementation of ALP's LA Shulman's theory of signature pedagogics was applied (Shulman, 2005). The application was to identify and categorize the levels of pedagogy evaluations which coincide with signature pedagogy's dimensions of the surface, deep, and implicit structure. Surface structure describes learning activities in an educational context, such as interactions between teacher, student, and technology, or concrete learning activities, such as reading, discussing, and completing assignments on an ALP. Deep structure describes the context and pedagogic structures that learning activities exist in, such as Flipped Classroom, Inquiry-Based Learning, and Dialogic Teaching. Lastly, the implicit structure represents the inherent values associated with the surface, and deep structures of the signature pedagogy, such as Flipped Classroom's purpose of distributing the workload to student preparation implies an intrinsic value contribution to in-class learning activities.

3. Analysis and Synthesis

We report the results of a qualitative synthesis of a systematic review. Evaluation methods identified in the papers are grouped into six categories based on their main focus. These categories are: C1) evaluating design and framework, C2) performance, C3) adaptivity, C4) perceived value, C5) multimodal methods, and C6) pedagogical implementation. These categories are further described in the following sections. Each category has multiple themes to specify the studies' focus. Papers may be presented in multiple themes within a category. Some studies used multiple evaluation methods. We review and map these studies at the end of each category's section (see Table 2).

The distribution of the identified evaluation categories and how they are connected to each other is visualized in figures 2 and 3 showing a low representation of studies evaluating multimodal methods (C5) and pedagogical implementation (C6). 27 studies contribute to multiple evaluation categories and only five papers fall under single category. In addition, in literature, there is a preference for variables used to assess and describe students and platforms compared to variables assessing and describing teachers (see Table 1).

Furthermore, the most frequently applied methods of evaluation are visualized in figure 4 where Logs on user activity (n=11), Interviews (n=5), and Pre- and Post-test (n=5) are the most preferred methods used.

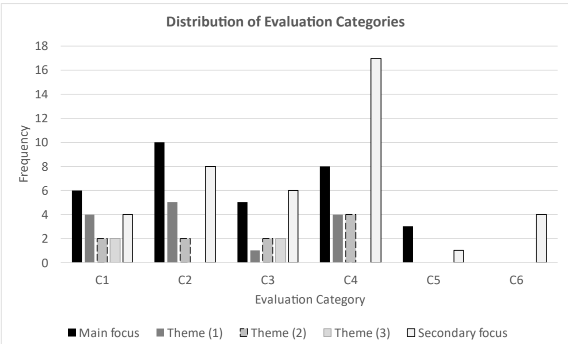

Figure 2: The distribution of evaluation categories on the evaluated papers

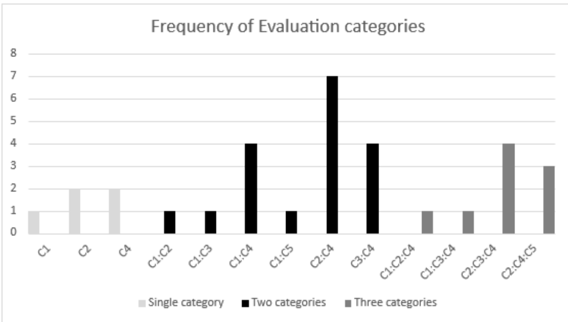

Figure 3: Frequency of primary focus evaluation categories and their combinations in studies

Table 1: Variables and their category association used by Learning Analytics and Learning Analytics Dashboards either describing or being described by students, teachers, or the platform when evaluating Adaptive Learning Platforms

	Variables describing			Variables being described by
Students	(C2) Time	(C2) Competence	(C2) Score	(C2) Self-assessment
	(C2) Reading competence	(C2) Completion time	(C2) Performance score	(C2:C4:C5) Perception on performance
	(C1:C4) Learning styles	(C1:C4) Learners	(C1:C4) Confidence and trust	(C2) Self-completion
	(C1:C4) User behaviour	(C1:C4) User interactions	(C1:C5) Effort	
	(C1:C5) Response time	(C1:C5) Performance	(C1:C5) Arousal	
	(C2:C4) Students' learning performance	(C2:C4) Performance	(C2:C4) Students' performance	
	(C2:C4) LAD use	(C2:C4) Satisfaction with LAD use	(C2:C4) Confidence	
	(C2:C4) Satisfaction	(C2:C4) Learning attitude	(C2:C4) Technology acceptance	

	Variables describing			**Variables being described by**
	(C2:C4) Problem solving activities	(C2:C4) Learning performance	(C2:C4) Learning experience	
	(C2:C4) Learning flow	(C2:C4) Learning performance	(C2:C4) Perceived learning experience	
	(C2:C4) Students' attitude on system	(C2:C4) Learning outcomes	(C3:C4) Linguistic category features	
	(C3:C4) Users attitudes towards the system	(C3:C4) User performance	(C3:C4) Learner's personality	
	(C3:C4) User experience of cognitive load	(C3:C4) Language complexity measures	(C1:C2:C4) Performance	
	(C1:C2:C4) Learner satisfaction	(C1:C2:C4) Motivation	(C1:C2:C4) Study time	
	(C1:C2:C4) Cognitive styles	(C1:C2:C4) Previous relevant knowledge	(C2:C3:C4) Improvement of learning scores	
	(C1:C3:C4) Learning styles dimensions	(C1:C3:C4) Learning effectiveness	(C2:C3:C4) Learner's satisfaction scores	
	(C2:C3:C4) Learning effectiveness	(C2:C3:C4) Performance scores	(C2:C4:C5) Response time	
	(C2:C4:C5) Learner physiological signals	(C2:C4:C5) Learners affective state	(C2:C4:C5) Emotions	
	(C2:C4:C5) Arousal	(C2:C4:C5) Cognitive load	(C2:C4:C5) Attention	
	(C2:C4:C5) Playing accuracy	(C2:C4:C5) Playing speed	(C2:C4:C5) Goal orientation	
Teachers	(C1:C4) Teacher perceptions	(C2:C4) Technology acceptance	(C2:C4) Teacher's performance	(C1:C4) Teacher expectations
	(C2:C4) LAD use	(C2:C4) Satisfaction with LAD use	(C2:C4) Technology acceptance	
	(C2:C4:C5) Educators' tacit experiences,	(C2:C4) Teachers' perspective on system in praxis		
Platform	(C1) User centred design principles	(C1) E-learning life cycles	(C1) Layered evaluation of adaption features	(C1:C2) Prediction of performance
	(C2) Reading Performance assessment	(C4) Feasibility for virtual mentoring	(C4) Adjustability	
	(C4) Feasibility for virtual mentoring	(C4) Satisfaction	(C4) Reliability	(C3:C4) System performance
	(C4) Delightfulness	(C1:C4) Context	(C1:C4) Learning objects	(C1:C3) Recommendations meeting requirements of learners
	(C1:C3) Accuracy of learning material estimation	(C1:C4) Threshold levels	(C1:C4) Effectiveness	(C2:C4) Homogeneity in grouping through peer collaboration features in system.

	Variables describing			Variables being described by
	(C1:C4) Devices	(C1:C4) Usefulness	(C1:C4) Information adequacy	
	(C1:C4) Usefulness	(C1:C4) Novelty	(C1:C4) Serendipity	
	(C1:C4) Accuracy	(C1:C4) Level of detail	(C1:C4) Richness	
	(C1:C4) Usability	(C1:C4) Context awareness	(C1:C4) Domain coverage	
	(C1:C4) Coverage	(C3:C4) Adaptability and variability	(C1:C3:C4) User centric effects	
	(C1:C4) Information diversity	(C2:C3:C4) Adaptive to learning styles	(C2:C3:C4) Accuracy of identifying students' cognitive styles	
	(C1:C3:C4) System performance	(C2:C3:C4) Adaption to learning styles	(C2:C4:C5) Learners' engagement patterns	
	(C2:C3:C4) System impact on students learning engagement	(C2:C4:C5) Level change timing		

As an example of how to read the table, the variable of Time is used to describe student performance (C2) whereas, self-assessment is a variable being described by students.

Figure 4: Word Cloud on methods from table 1 showing methods (n=235) with a frequency of <1

3.1 C1) Evaluation of LA and LAD Design and Framework

In reviewing the included studies, the category of Evaluation of LA and LAD design and Framework emerged. This category is divided into the themes of (1) Users' perception of framework and design, (2) System evaluation as a feature of the ALP, and (3) Requirement and comparative analysis of framework and design. Six papers focus on evaluating the framework and design of LA and LAD on ALPs, whereas four papers apply the evaluation as a secondary contribution of their paper. These four will be assessed at the end of the C1) category.

3.1.1 Users' perception on framework and design

The first theme of Users' Perception on Framework and Design consists of four papers where users' perception was primarily used for evaluating the LA and LAD design and framework in ALPs. Abech et al. (2016) proposed EduAdapt, an architectural model for the adaptation of learning objects which considered device characteristics, learning style, and students' contextual information in its content recommendation to users. In evaluating the architectural model, scenarios, use cases, prototypes, a learning style survey, and a self-developed user experience survey were used. The study developed a mobile application prototype and applied it in an undergraduate course called Ubiquitous and Mobile Computing with learners (n=20) from the Computer Science area for one month of teaching. The main scientific contribution was proposing a learning object adaptation model employing inferences, rules, and learning styles in a varied context ontology. Lau, Lee, and Singh (2015) developed a recommender system for personalizing system recommendations using students' annotated metadata through a schema. The system was evaluated in two stages by students (n=92) enrolled in the course Introduction to Personal Finance, with 42% completing the evaluation. The first stage evaluated the ontology's quality using a questionnaire on 12 competency statements. The second stage used a novel seven-pillar evaluation framework, which included user-perceived accuracy, novelty, domain coverage, confidence, information adequacy, usefulness, and usability. The study's contribution was a framework that enabled learners to use their peers' opinions to locate relevant and quality resources. Tahmasebi, Ghazvini, and Esmaeili (2018) developed a feature-based educational recommender ranker system that interacts with users based on their learning styles. The system was, among other methods, assessed through a user questionnaire to evaluate the general perception of the proposals from the recommender system. The empirical study was done on science and engineering students (n=77) through a questionnaire during two semesters. The findings showed that the proposed method outperforms the general search algorithm and that the tool could be used for other systems. Fasihuddin, Skinner, and Athauda (2017) also used user perception to identify patterns and threshold levels that lead to optimal precision in detecting learning styles by tracking learners' behaviors in open environments. User perception was applied in their framework through the Indexed Learning Styles questionnaire (ILS). The empirical study was done on a prototype Cloud Adaptive Learning Courses (CALC) of an open learning environment developed and piloted on undergraduate students (n=83) taking an undergraduate IT course. It was shown that threshold values derived from literature and customized to suit open learning environments provided high accuracy in identifying learning styles.

3.1.2 System evaluation as a feature of the ALP

In the theme System Evaluation as a Feature of the ALP, two studies where the evaluation of the framework and design of the ALP is a feature of the ALP were found. Dounas, Salinesi, and Beqqali (2019) presented a framework to evaluate the INSPIREus adaptive educational hypermedia environment. The study analyzed the system's suitability for different learner types and examined if it fulfilled four requirements. The system was monitored using a novel tool, RMAS, and the collected data was checked against automatically derived constraints from the requirements. The study involved informatics and telecommunications students (n=21) over a three-month course. The results showed INSPIREus was most accurate for theorist learners and least accurate for reflector learners. It fulfilled two requirements but struggled with communication and control sharing. The system also needed more flexibility in addressing individual student needs. Fasihuddin, Skinner, and Athauda (2017) evaluated the precision of the threshold levels for identifying students' learning styles in a literature-based method. The precision was computed as the similarity of the system-identified learning style to the learning style determined by the ILS survey responded to by the students.

3.1.3 Requirement and comparative analysis of framework and design evaluations

In this section, the theme of Requirement and Comparative Analysis of Framework and Design Evaluations is presented constructed from two studies, where Abech et al. (2016) set up a comparative analysis of an ontology specified for adaptive learning systems called OntoAdapt against other ontologies. The comparative analysis was done through scenarios, analyzing the quality and fidelity of OntoAdapt. The comparative analysis was done with evaluation metrics from the Full Ontology Evaluation (FOEval) method, some additional variables provided by the software Protége and complemented with the Manchester— Web Ontology Language (OWL) tool Ontology Metric. In the study's second phase, the ontology was assessed on the platform EduAdapt to investigate if the use of ontology matches the learning objects adaptation scope. Santos and Boticario (2015) proposed the TORMES methodology for eliciting recommendation opportunities in a recommender system using six steps: (1) Context of use, (2) User requirements, (3) Modeling of the design solution, (4) Publication of the design solution, (5) Usage to gather evaluation data, and (6) Feedback from evaluating design requirement. The study demonstrated the methodology's use in two contexts, namely the Discovering the Platform course in dotLRN LMS and the EBIFE course in the Willow free-text adaptive computer-assisted assessment system, each with two iterations. Results revealed that TORMES can detect problematic affective situations and react efficiently, but the system was intrusive in gathering affective data, causing discomfort among participants.

In addition to the three presented themes of (1) Users' perception of framework and design, (2) System evaluation as a feature of the ALP, and (3) Requirement and comparative analysis of framework and design evaluations, four papers applied the evaluation of LA and LAD design and framework as a secondary focus. Mavroudi et al. (2016) presented a framework to frame the user requirements of an adaptive system. Beckmann, Bertel, and Zander (2015) presented a framework to adapt learning material. Rincón-Flores et al. (2019) developed an algorithm from three different forecast models, and this algorithm was based on unstructured data implemented in an adaptive learning system. Lastly, Sharma, Papamitsiou, and Giannakos (2019) developed a prediction algorithm for an adaptive learning system applied in an experimental design.

3.2 C2) Evaluation of Users Performance with LA and LAD

This category emerges from the frequent use of user performance in LA is applied as a criterion for evaluating ALPs. We further thematize the category nuancing how user performance is applied. The themes are (1) sorting LA content based on user performance and (2) the effect of LA features on user performance. Ten papers focus on evaluating user performance with LA and LAD on ALPs. Eight papers also evaluate the category but as a secondary contribution of their paper.

Table 2: Review of Results

Author	Category	Evaluated unit	Methods	Variables	Purpose
Di Mascio et al. (2013)	C3, C4	Adaptive learning system TERENCE	Heuristic evaluation, expert reviewing, cognitive walk-through, observations, think-aloud and verbal protocols, controlled experiments, simulation and system performance indicators.	Users' attitudes towards the system, users' performance and system performance.	The qualitative methods such as Heuristic, expert reviewing, and cognitive walk-through evaluations are used to evaluate design choices, while simulations and system performance indicators are used to evaluate usability.
Bresó, et al. (2016)	C3, C4	A mechanism that adapts to stamina/mood	Surveys and simulations	Adaptability and variability.	The simulation evaluated system outputs, while surveys and a pilot case assessed perceived variability and adaptability levels.
Tlili et al. (2023)	C3, C4	Method for modelling to learners' personalities	Survey and LA student personality scores.	Learner personality	LA estimated learners' personalities and surveys validated personality models

Author	Category	Evaluated unit	Methods	Variables	Purpose
Hsu and Li (2015)	C2, C3, C4	Adaptive learning algorithm	Surveys, pre- and post-tests, performance scores, and user satisfaction scores.	Learner satisfaction scores and learning effectiveness	How well did the algorithm perform in terms of student performance and learning satisfaction with LAD?
Nye, et al. (2021)	C4	MentorPal, adaptive framework for virtual mentors	Formative user testing interviews, log data, pre- and post-surveys for career attitudes, and post-survey for usability.	Feasibility for virtual mentoring	The LAD's statistical evaluation was verified against users' subjective quality assurance testing.
Abech et al. (2016)	C1, C4	Ontology model for LA and LAD	FOEval, user feedback, surveys, measurements of survey reliability, user scenarios, competence questions and usage patterns.	Learners, learning objects, devices, context, context awareness, coverage, richness, and level of detail.	The ontology undergoes two phases of evaluation: the first for development and the second for comparison with other ontologies and performance in a learning context.
Mavroudi, et al. (2016)	C1, C4	Teacher-led design on the envisioned adaptive system	Evaluation questionnaire and Qualitative Comparative Analysis	Teacher perceptions and expectations	A methodology to frame requirements for critical success factors to meet user expectations of the system.
Khawaja, Chen, and Marcus (2014)	C3, C4	Adaptive multitouch tabletop interaction application.	Subjective ratings, Linguistic Inquiry and Word Count, and Advanced Text Analyzer.	User's experienced cognitive load, Language Complexity Measures, and Linguistic Category Features.	A none-intrusive, non-manipulative adaptive learning method that adjusts to users' cognitive load.
Santos, et al. (2016)	C2, C4, C5	Ambient Intelligence Context-aware Affective Recommender Platform	Tutor Oriented Recommendations Modelling for Educational Systems methodology, user-centred design methods, data mining techniques, interviews, and SUS questionnaire.	Learners' affective state, educators' tacit experiences, learner physiological signals	Exploring ambient intelligence's sensory feedback and its impact on personalized support through a recommender system.
Zhang et al. (2023)	C2, C4	Student-centred online one-to-one tutoring system	Pre- and post test of students' academic performance and system log files.	Students' learning performance (academic), teachers' performance (attracting students)	Evaluation of the system's practical value.
Kim, Jo, and Park (2016)	C2, C4	LAD for students in a virtual learning	Logs on students' frequency of use, survey on satisfaction,	Students' performance, LAD use, and satisfaction with LAD use.	Evaluating LAD use to improve

Author	Category	Evaluated unit	Methods	Variables	Purpose
		environment called Cyber Campus	and final scores on tests/exams. Analysis of the relationship among three variables.		student performance.
Lau, Lee, and Singh (2015)	C1, C4	Recommender system on user annotated metadata.	Questionnaires to experts, users, and logs on user activity	Accuracy, novelty and serendipity, domain coverage and information diversity, confidence and trust, information adequacy, usefulness and effectiveness, and usability.	Evaluating and enhancing ontology quality based on expert suggestions.
Hooshyar et al. (2018)	C2, C4	Solution-based Intelligent Tutoring System for flowchart development to improve students' problem-solving skills.	A questionnaire assessing students' learning attitude, learning interest and technology acceptance and a pre- and post-test.	Learning interest, learning attitude, technology acceptance, and problem-solving activities.	Assessing the improvement of students' problem-solving skills using the system.
Troussas et al. (2020)	C2, C4	An intelligent tutoring application over Facebook called i-LearnC# is used for learning C# programming.	Population characteristics survey, CIAO framework survey, teacher interviews, statistical hypothesis test, and system's log files.	Students' attitude on system, teachers' perspective on system in praxis, Learning outcomes, and homogeneity in grouping through peer collaboration features in system.	Evaluating social network's potential to support learners in universities and technological institutes.
Lo et al. (2012)	C2, C3, C4	An adaptive web-based learning system focusing on students' cognitive styles.	Cognitive style questionnaire based on Myers-Briggs and calculations with log files.	Accuracy of identifying students' cognitive styles and impact of the proposed adaptive web-based system on students' engagement in learning.	Unobtrusively identify students' cognitive styles through a multi-layer feedforward neural network compared to self-reported cognitive styles to provide adaptive content to students.
Dounas, Salinesi, and Beqqali (2019)	C1, C3	INSPIREus an adaptive educational hypermedia environment that provides personalized content and adaptive navigation support for each learner	Runtime Monitoring for Adaptive Systems (RMAS) tool	Accuracy of learning material estimation/recommendations meeting requirements of features used for the learning process.	To determine whether INSPIREus meets its own presented requirements.
Beckmann, Bertel, and Zander (2015)	C1, C2, C4	An adaptive framework on Moodle	Revised Verbaliser-Visualiser Questionnaire, Index of Learning Styles questionnaire, and performance scores from the test.	Learner satisfaction, motivation, Study time, Individual cognitive styles, Previous relevant knowledge, Performance	To assess whether adapting eLearning material to inter-individual differences in learning styles can improve learning

Author	Category	Evaluated unit	Methods	Variables	Purpose
					effectiveness and efficiency, learner motivation, and satisfaction.
Latham, et al. (2012)	C2, C3, C4	Adaptive online conversational intelligent tutoring system	The Index of Learning Styles Survey, Pre-post-tests, Self-developed user evaluation questionnaire, Log files.	Adaption to learning styles, Performance scores.	Assessing whether Oscar adapts to learning styles for higher student performance.
Effenberger and Pelánek (2019)	C2	RoboMission is an adaptive learning game.	Statistical analysis, Log files, and Diagnostic visualizations	Performance scores	Present discrete performance levels with universal interpretations rather than a binary failure success outcome.
Latham, Crockett, and McLean (2014)	C2, C3, C4	A conversation intelligent tutoring system called Oscar.	Index of Learning Styles, Performance score, Log files, pre- and post-test.	Adaptive to learning styles and improvement of learning scores.	Improvement of learning through an automated online conversational tutorial by presenting tutor material adapted to a student's learning style.
Dirin, Laine, and Nieminen (2017)	C4	Adaptive Mobile learning application	Emotional engagement analysis method, Web questionnaire, Diary on daily activities, Semi-structured interviews, Scenarios, Paper-based prototyping, Usability evaluation, and Observations	Delightfulness, Reliability, Adjustability, Satisfaction.	Defining a conceptual model of digital service sustainability and its measurement indicators and criteria.
Tahmasebi, Ghazvini and Esmaeili (2018)	C1, C3, C4	Feature-based educational recommender ranker system	Cronbach's Alpha Structural Equation Modelling (SEM) Fornelland Larcker's measure of average variance extracted (AVE) User perception questionnaire on the general perception of the proposals from the recommender system Index learning styles questionnaire Web crawler on meta-data of sample Open Courseware Consortium's web pages Pre- and post-test Log files.	Learning styles dimensions, System performance, learning effectiveness, and User-Centric effects.	Developing a model for a feature-based educational recommender ranker system.
Fasihuddin, Skinner and Athauda (2017)	C1, C4	CALC - open learning environments-threshold for	ILS survey Log files (Automated calculation/pattern tracking)	Learning Styles, User interactions, User behaviour, and threshold levels.	To determine whether the literature-based method

Author	Category	Evaluated unit	Methods	Variables	Purpose
		estimate Learning Styles			can achieve a satisfactory level of precision in identifying learning styles in open learning environments.
Yuksel et al., (2016)	C2, C4, C5	BACh is an adaptive brain-computer system that teaches piano.	LIBSVM, Bitwig, fNIRS data, Perception of performance, questionnaire, Interview on level change	Playing accuracy, Playing speed, Level change timings, Perception on performance.	To investigate whether BACh can dynamically adapt to increasing difficulty levels in a musical learning task based on pianists' cognitive workload.
Katuk, Kim, and Ryu, (2013)	C2, C4	A dynamic content sequencing system (DCSS)	Activity-followed-by-survey method, pre- and post-tests, learning experience questionnaire	Learning performance, Learning experience	Examining the learning experience in conjunction with learning performance to assess the adaptive learning system.
Lynch and Ghergulescu (2017)	C2, C4	Adaptemy system	Log files, Predication algorithm developed on the item Response Theory, and Questionnaire on Learning Experience.	Learning flow, Learning performance, Perceived learning experience.	Examining the learning experience and learning performance to assess the adaptive learning system.
Fadljević et al. (2020)	C2	Adaptive Moodle learning system on medical content.	K-means ANOVA, analysis, clustering, TukeyHSD, One-Way Correlation	Reading Performance assessment, Time, Competence, Score, Self-Completion	Assessing temporal behaviour as a predictor of performance on the system and whether students were fast or slow because text difficulty was unsuitable.
Papamitsiou et al. (2020)	C2, C4, C5	Multimodal self-assessment adaptive learning system	Fuzzy set qualitative comparative analysis, Pre-test goal expectations survey, Multimodal (eye-tracking, wristband, cameras, and EEG cap)	Learners' engagement patterns, Response time, Arousal, Cognitive load, Emotions, Attention, Goal-orientation	Measuring learner engagement in a multimodal learning setting.
Rincón-Flores et al. (2019)	C1, C2	Forecasting algorithm	Test and control group setup, K-nearest Neighbour, Random Forest, Logs, Grades, Photographs, Semi-structured interviews, Student t-distribution.	Prediction of performance	Using three forecast models, AI can predict student performance on unstructured data

Author	Category	Evaluated unit	Methods	Variables	Purpose
Sharma, Papamitsiou, and Giannakos (2019)	C1, C5	Developing a multimodal forecasting algorithm	Support Vector Machine algorithm, Decision trees, Gaussian process regression, Machine Learning, Principal Component Analysis, and Random Forrest.	Effort (Response time effort), Performance, Arousal	Evaluating which combination of physiological data from students explains effortful engagement and learning performance in an adaptive learning system.
Santos and Boticario (2015)	C1	A design for developing adaptive learning management systems	Brainstorming, Focus groups, Wizard of Oz, Observational study, Questionnaires, Data logs, Interviews, and Problem scenarios	User-centred design principles, E-learning life cycles, Layered evaluation of adaption features	Developing adaptive learning platforms according to ISO and other design protocols.
Al-Shanfari et al. (2020)	C2, C4	OLMlet	Bias score calculation, ANNOVA, Non-parametric tests, independent t-test, Mann-Whitney U tests, Bonferroni correction, Semi-structured interviews.	Confidence, Performance	Displaying performance and confidence levels to students to improve their performance.

3.2.1 Sorting LA content on user performance

In this section, the Sorting LA content on User Performance theme is presented which is constructed from four studies. Hsu and Li (2015) developed a new algorithm called the competency-based guided-learning algorithm (CBGLA). CBGLA-based learning system guided learners in achieving the learning objectives through personalized learning paths on the student's performance on the platform. A pilot study of the system was tested on third-year college students of electrical engineering (n=6) before an experiment on the same type of students (n=59) was conducted. The findings showed that the CBGL system supported students' learning. Yuksel et al. (2016) presented a study on an adaptive brain-computer system (BACh). The system increased the difficulty of a musical learning course aimed at the piano when cognitive workload levels became low. In a within-subject test design, study participants(n=6) undertook a training task playing 15 easy and 15 complex pieces on the piano. Pieces were learned through a typical approach (control), and two were learned afterward through BACh. Participants then played the four pieces, and performance data was used to assess the system. Results showed that learning with BACh increased accuracy and speed compared to the control setup. Fadljević et al. (2020) presented an e-learning system to support students' acquisition of health literacy with content developed in collaboration between clinical psychologists, pedagogues, and medical students. The Moodle-based system adapted text difficulty depending on students' reading competence, performance score, and self-assessment of students. From the LA, students were grouped into four competence levels. The study's participants were students [n=196] from 6th to 8th grade and worked using the system for four weeks. The results showed that at each difficulty level, students could be separated into a class of slow and a class of fast students. The text difficulty was for no students deemed unsuitable. Rincón-Flores et al. (2019) presented a predictive algorithm applied in an adaptive learning environment where instructors (n=3) taught a Physics II course. The algorithm was trained on photographs, grades, and log data from instructors' previous course students (n=182). The results showed that the algorithm provided a good forecast of the performance of each group. Al-Shanfari et al. (2020) presented an open learner modeling system (OLMlet) that used student question answers to provide adaptive feedback to students based on the correctness of their answers. The mixed method study was done with undergraduate engineering students (n=32) in an introduction to Java programming course. Students were split into a control and a test group. The control group was presented with only the skill meter (performance score). The test group was presented with the alignment between the system's evaluation of students' performance and confidence levels. The findings showed that low-achieving students benefited highly from being presented with a visual alignment model. The visual model of alignment was associated with positive changes in their performance.

3.2.2 Effect of LA features on user performance

Five papers constitute the theme of Effect of LA Features on User Performance. Zhang et al. (2023) presented the Student-Centered Online One-to-one Tutoring system (SCOOT), where students could ask questions outside school to expand the flexibility of posing questions. The study sought to evaluate the efficiency of SCOOT and examine how students' prior knowledge and simple patterns of tutoring sessions affected student learning. The evaluation included integrating students' learning performance and behavior log files instead of conducting between-subject experiments. The empirical study comprises 40 tutoring sessions randomly selected over 50 days with a pre-test and a post-test. The study participants (n=810) were selected from Grade 7 mathematics. The results showed that the flexibility element of SCOOT was necessary and that SCOOT further increased performance differences between high- and low-achieving students. In Kim, Jo, and Park (2016), a LAD for students in the virtual learning environment Cyber Campus was implemented, which distributed video lectures and quizzes and enabled students to submit assignments. Through an experimental design on college students (n=151), the LAD was assessed on whether it would lead to higher performance. The findings showed a significant difference between the treatment and control groups. Students who had access to the LAD performed better, but an exciting finding showed that few uses of the LAD led to higher satisfaction compared to more frequent uses. Furthermore, learners who used the LAD frequently and performed well were less satisfied with the LAD. Latham et al. (2012) applied Oscar, an adaptive online conversational intelligent tutoring system that delivered a personalized natural language tutorial by predicting and adapting to students' learning styles. Participants in the study were undergraduates of science and engineering (n=70) who previously worked with the content and were to revise the topics with Oscar. Results showed a significant difference in the group with mismatched learning styles; students performed better when presented with materials matching their learning styles. In a subsequent study, Latham, Crockett, and McLean (2014) Oscar was assessed on whether it supported students' discussions in constructing knowledge. Students' learning styles were defined through Oscar's adaptation algorithms, and the validity of the system's categorization of students' learning styles was tested on student perception. The empirical study involved undergraduates (n=62) in science and engineering in assessing whether they improved their performance using Oscar. Findings showed a significant difference between match/mismatch groups. Effenberger and Pelánek (2019) presented RoboMission, which was an adaptive learning system for introductory programming. RoboMission took the form of a programming game. The case study examined the task sessions from students (n=3.800), illustrating that designing performance measures is nontrivial but possible. Three papers (Hsu and Li, 2015; Yuksel *et al.*, 2016; Al-Shanfari *et al.*, 2020) presented in C2) (1) also evaluate the effect of LA as described in C2) (1).

In addition, Santos et al. (2016) evaluated the effectiveness of supporting the learning process by, e.g., giving affective and sensory input to help calm users in a stressful learning context and whether the input was helpful in the students' performance. Hooshyar *et al.* (2018) evaluated students' improvement of learning achievement on their Solution-based intelligent tutoring system through pre- and post-tests, showing a positive impact on the learning achievements of the experimental group. Troussas *et al.* (2020) evaluated students' learning outcomes through a t-test on log data from the adaptive system. Beckmann, Bertel, and Zander (2015) used performance statistics to assess a correlation between learning styles, learning content format, and other statistics. Lo, Chan, and Yeh (2012) evaluated the students' engagement with learning on the platform, and Katuk, Kim, and Ryu (2013) assessed the students' learning performance by applying a pre and post-test setup. Lynch and Ghergulescu (2017) assessed students' performance and performance improvements with log files and developed prediction algorithms based on the Item Response Theory. Finally, Papamitsiou et al. (2020) used students' performance to compare and evaluate their self-assessment of their preparation and their engagement.

3.3 C3) Evaluation of Adaptivity

This category surfaces from the different studies that Evaluate the Adaptivity of ALPs. Evaluation of adaptivity in the ALPs is divided into three themes: (1) adaptivity based on psychological inclinations, (2) adaptivity based on users' affective capabilities, and (3) adaptivity based on users' cognitive capabilities. Five papers evaluate the adaptivity of LA and LAD on ALPs. Six additional papers mention adaptivity in their studies but do not present it as their main focus.

3.3.1 Adaptivity based on users' psychological inclinations

Only one study evaluated adaptivity on users' personalities. Tlili *et al.* (2023) assessed students (n=50) in exploratory research on building an evidence-based model to map users' personalities on Big Five Inventory dimensions to adapt learning content through the iMoodle LMS. The findings showed that the Bayesian network makes it possible to model learners' personalities compared to BFI for the three personality dimensions of extraversion, openness, and neuroticism.

3.3.2 Adaptivity on users' affective capabilities

Bresó *et al.* (2016) evaluated adaptability and variability of content from simulations and user feedback in the Personal Health System, a part of Help4Mood. The personal health system was developed to support users, so they do not relapse into depression. The system adapted its content to users' stamina or mood. Simulations (n=20.000) were done on 19 tasks and 31 subtasks; one or more subtasks could form a task. The paper concluded that the framework provided adaptive and varied sessions, improving users' use experience. Santos *et al.* (2016) presented an Ambient Intelligence Context-aware Affective Recommender Platform (AICARP) that applied Tutor Oriented recommendations Modeling for Educational Systems (TORMES) elicitation methodology to sense changes in learners' affective state from sensory communication channels. In the exploratory empirical case study, participants (n=6) completed tasks in a Wizard of Oz setting with a psycho-educational expert as the wizard. Findings showed that recommendations from an intelligent ambient system could tackle affective issues during the second language learning process.

3.3.3 Adaptivity on users' cognitive capabilities

Khawaja, Chen, and Marcus (2014) presented a model for improving performance in complex and time–critical situations by dynamically deploying more appropriate output strategies to reduce cognitive load on linguistic behavioral features. The study examined a session where participants (n=44) managed firefighting tasks as a team. The findings showed that an interaction system could apply speech and linguistic patterns to determine cognitive load and adapt system responses minimizing users' cognitive load to maintain performance. Lo, Chan, and Yeh (2012) presented an adaptive web-based learning system that adapted learning material to students' cognitive styles. In an initial study, cognitive styles were identifiable from students' (n=162) browsing behavior in the adaptive learning system. Evaluating the impact of the adaptive web-based learning system on students' engagement, another study was set up on college students (n=170) from Computer Science and Informatics. The results showed that the adaptive learning system significantly impacted temporal elements' effect on students' learning engagement. The study demonstrated that the adaptive web-based learning system based on students' cognitive styles could effectively enhance students' engagement in learning for Interpersonal and Mastery styles. In addition to the five papers synthesized above, six papers evaluated adaptivity. Di Mascio et al. (2013) presented usability associated with adaptivity and Hsu and Li (2015) assessed the effectiveness of adaptivity. Latham *et al.* (2012) and Latham, Crockett, and McLean (2014) used learning styles to adapt material delivered to students and used their performance scores to evaluate the adaptivity of their method. Tahmasebi, Ghazvini and Esmaeili (2018) used users' perception of the proposed learning material fit users learning styles as an evaluation of the adaptivity. Papamitsiou et al. (2020) used students' self-assessments to evaluate performance from multimodal data.

3.4 C4) Evaluation of ALPs Through Perceived Value

The evaluation of ALPs through perceived value is divided into the following themes: (1) users' perspectives on usability and (2) self-efficacy elements and learning styles. Eight papers focus on evaluating LA or ALPs from users' perceived value or evaluation of users' perceived value. All the papers analyzed in this review used some element of users' perspective, but most with it as a secondary focus.

3.4.1 Users' perspectives on usability

Di Mascio et al.(2013) developed the TERENCE system's Graphical User Interface (GUI) prototypes for supporting poor comprehenders and their educators. Three groups participated in the development: primary-school students (n=170), educators (n=10), and experts (n=10) such as psychologists and linguists. Usability evaluations, including expert-based evaluation, observations, think-aloud, and verbal protocol of experts and users, were conducted to identify users' requirements and context of use. Findings highlighted the importance of considering the timing and focus of users' participation and system performance during the execution of users' tasks in usability testing. Nye et al. (2021) evaluated the usability of MentorPal, a virtual mentor system that gave career advice to high school students (n=31) attending STEM internships, with the Unified Theory of

Acceptance and Use of Technology constructs (UTAUT) survey and a survey generated from variants of the CAPA system. MentorPal's career advice focused on STEM careers in the Navy, and the researchers observed the students' usage unobtrusively. Although the study had limited sample size, diversity, and impact, the students found the MentorPal experience compelling and valuable. The study's findings suggested the need to improve the diversity representation and coverage of students' main career interests by mentors. Troussas et al. (2020) studied i-LearnC, an intelligent tutoring application for learning C programming. The application overlaid Facebook, and as students made mistakes, a virtual coach (ViC) provided suitable learning material to correct misconceptions. ViC advised the pace of instruction based on students' profiles built on current and previous knowledge levels, collaboration preferences, and types of misconceptions. In an experimental design, second-year students (n=400) in an undergraduate course on Object-Oriented Development of Applications used i-LearnC, and assessment was done through a survey on the CIAO! framework, which evaluated students' use of technology-based teaching and learning. Results showed students had a positive attitude toward using Facebook for educational purposes, appreciated the communication and collaboration features, and found them helpful. Peer recommendation for collaboration showed a significant difference in acceptance of recommendations provided by i-LearnC compared to the conventional system. Dirin, Laine, and Nieminen (2017) presented an adaptive mobile learning application that provided theory and assessment for driving school students based on their learning competence and progress. The empirical study was done on participants with driver’s licenses (n=7) and instructors (n=5) to identify user needs and requirements for the target application to develop a paper-based prototype. The prototype was tested in a usability laboratory on the user experience factors of delightfulness, adjustability, satisfaction, and reliability. The findings demonstrated that users’ emotional attachment is essential for the target users.

3.4.2 Self-efficacy elements and learning styles

Hooshyar et al. (2018) presented a solution-based intelligent tutoring system (SITS) with an automatic text-to-flowchart conversion approach for engaging students in flowchart development aimed at improving students’ problem-solving skills in an experimental design. Participants were university students (n=32) in an introductory programming course completing a self-developed questionnaire on students’ learning attitude, -interest, and technology acceptance, assessing the ease of use and usefulness of SITS. From the questionnaire, it was seen that using SITS, students experienced an enjoyable learning context, were motivated to use it and experienced the content proposals from the system as helpful. Beckmann, Bertel, and Zander (2015) presented an adaptive learning system that targets performance, motivation, satisfaction, and previous knowledge to assess the effectiveness of a Moodle platform using an adaptive learning ’layer’ to distribute content to students’ learning styles. In a mixed-methods study design, students (n=53) of Computer Science and Media studies participated under laboratory conditions completing a questionnaire on learning motivation and satisfaction. Analysis with non-parametric statistical methods resulted in no significance of a good or bad fit between visual/verbal format and individual learning style on study time and learning outcomes. Also, there was a significant influence of matching learners’ learning styles with learner satisfaction and motivation. Katuk, Kim, and Ryu (2013) presented IT Tutor, which is an adaptive learning system. In a one-way between-subject design study, participants (n=80) assessed the e-learning application, the tutorial session, and the learning experience using a learning experience questionnaire and other instruments. The results indicated that the lower or medium achievers gained certain benefits from the platform, while the high achievers in learning performance might suffer from boredom. Lynch and Ghergulescu (2017) assessed the Adaptemy system, which used curriculum-mapped content to provide personalized learning journeys to students. Secondary students participated in an objective study (n=7.614) assessing the adaptivity content and a subjective study (n=80) assessing the perceived learning experience. The subjective study revealed that the students felt increased confidence in solving math questions and an increase in enjoyment, confidence, and improved learning. In addition, 17 papers reported the evaluation of perceived value. Five studies (Abech *et al.*, 2016; Bresó *et al.*, 2016; Santos *et al.*, 2016; Tlili *et al.*, 2023; Zhang *et al.*, 2023) presented the evaluation of perceived value as a method for further informing the performance of LA. Two studies (Bresó *et al.*, 2016; Santos *et al.*, 2016) used the evaluation of perceived value to evaluate adaptability and variability and to assess the usability of the LA. Hsu and Li (2015) presented the perceived value of students to determine satisfaction levels of LA, Khawaja, Chen, and Marcus (2014) estimated the perceived level of cognitive load, Zhang et al. (2023) evaluated the practical value of the LA, and Abech et al. (2016) and Mavroudi et al. (2016) developed the application with input on the perceived value from students. Lo et al. (2012) used Myers-Briggs definitions to create cognitive style questionnaires to get students’ insights on their cognitive styles. Latham et al. (2012) and Latham, Crockett, and McLean (2014) used a self-developed user evaluation questionnaire to understand students’ perceived value of the intelligent conversational agent system Oscar. Tahmasebi, Ghazvini and Esmaeili (2018) used users’ perceptions for the assessment of the adaptivity of their system. In the assessment process, they also evaluated users' experienced value of the system. Fasihuddin, Skinner, and Athauda (2017) applied the Felder and Silverman Learning Styles (ILS) survey to compare the automated calculation of Learning Styles on behavior and interactions with students' perceived learning styles. Yuksel et al. (2016) used questionnaires to evaluate participants' perceived performance with BACh and used interviews to assess whether the level changes were done adequately. Al-Shanfari et al. (2020) used semi-structured interviews to understand student experiences and explain behavior identified in the data logs acquired from the system.

3.5 C5) Evaluation of Multimodal Methods

Only three papers pertained to category C5, eliminating the need for themes to further refine the studies. Papamitsiou et al. (2020) introduced a self-assessment adaptive learning system using multimodal data analysis in various configurations. The study involved 32 undergraduates who participated in an online adaptive self-assessment procedure. Multimodal data were collected using cameras, wristbands, eye-tracking, clickstreams, and EEG caps, measuring variables such as cognitive load, heart rate, blood volume pressure, temperature, EDA, attention, and emotions. Six configurations explained learners' high performance, while three configurations explained learners' medium/low performance based on engagement measures from the collected data. Sharma, Papamitsiou, and Giannakos (2019) studied an online adaptive self-assessment system for a Web Technologies course with 32 undergraduates. They collected EEG, eye-tracking, facial expressions, and wristband data, used feature selection algorithms, and employed Machine Learning techniques for prediction purposes. In contrast, Yuksel et al. (2016) used brain and MIDI data to evaluate piano performance in the BACh platform.

3.6 C6) Evaluation of the Pedagogical Implementation of ALP's LA and LAD

For the category of C6) the three dimensions (surface structure, implicit structure and deep structure) of signature pedagogy (Shulman, 2005) are used to assess the papers' level of pedagogy evaluation. We do not include the theme for evaluating ALP's LA surface structure or implicit structure in this category. All included studies examine learning activities and/or interactions between students, teachers, and ALP's LA and LAD as examined in the categories of C2-C5. No studies were exclusively categorized as a C1 category which follows that all studies evaluate ALP's LA and LAD on signature pedagogy's surface structure dimension. Moreover, the signature pedagogy dimension of ALP's LA and LAD's implicit structure was evaluated by no papers. This leaves the evaluation category of the pedagogical implementation of ALP's LA and LAD with only one theme: Evaluation of ALP's LA deep structure. Four papers have a secondary focus on evaluating ALP's LA deep structure.

3.6.1 Evaluation of ALP's LA deep structure

Four papers mentioned pedagogical theory as a contextual factor for their studies. None of the studies evaluated how pedagogical theory was evaluated in either LA, LAD, or frameworks. Di Mascio et al. (2013) applied expert evaluation consisting of 10 learning experts, who evaluated the TERRENCE system prototype, they included a pedagogical direction described as the pedagogical stimulation plan. The results from the user evaluation consisted of users (n=170) who assessed whether the expectations of the pedagogical stimulation plan were met. The evaluations were done through observational, think-aloud, verbal protocols, and controlled experiments. Abech et al. (2016)reviewed other works on an ontology that had a pedagogical approach. The reviewed ontologies were compared to their own ontology's adaption to learning styles, but their ontology was not assessed on any pedagogical parameters. Hsu et al. (2015) used competency-based learning to develop their Competency-Based Guided-Learning Algorithm (CBGLA). However, their study does not mention how CBGLA could or should be implemented in a pedagogical context, nor how CBGLA resulted in the development of users' competencies. Troussas et al. (2020) mentioned the Revised Bloom Taxonomy and collaborative learning theory as the fundamental theory for their research on the adaptive learning system i-LearnC. Still, no further description or evaluation of deep structure was included.

4. Conclusion

This study extends the research conducted by Tretow-Fish and Khalid (2022) through a systematic literature review, encompassing the analysis of 32 empirical papers and two reviews. A comprehensive assessment of Learning Analytics (LA) and Learning Analytics Dashboards (LAD) on Active Learning Platforms (ALPs) involves the examination of 27 studies contributing to multiple evaluation categories, with only five papers falling within a single category (Figure 3).

A notable trend in the literature review is the prevalence of variables employed for assessing and characterizing students and platforms, in contrast to the limited attention given to variables focused on evaluating and describing teachers (see Table 1). The preferred methods for assessment predominantly include questionnaires (surveys), log files, and interviews (see Figure 4). Notably, the evaluations encompass ontologies, frameworks, methodologies, experimental designs, mathematical models, and LA statistics-components integral to Learning Analytics. However, assessments of pedagogical elements are conspicuously absent from the literature.

Few studies have addressed the pedagogical implementation of ALPs, with the majority treating it as a secondary focus rather than a primary concern. Despite various investigations into ALPs' LA and LADs, the existing literature predominantly evaluates the surface structure (i.e. operational acts – demonstrating, questioning, etc.), neglecting the deep structure (i.e. know-how of discipline – math by derivation practice, design by doing iterative design) and implicit structure (i.e. moral values – nursing for physical and mental health) in impact and method evaluation. Specifically, none of the reviewed studies, whether as a primary or secondary focus, systematically assess the deep and implicit structures in impact and method evaluation related to ALP's LA and LAD.

Our review reveals a noteworthy gap in the educational exploration of ALP's LA and LAD as tools for informing pedagogical or didactic decision-making among students and educators. The deficiency extends to the scarcity of studies employing established pedagogical methodologies, theories, and concepts—such as Flipped Learning, Inquiry-Based Learning, Simulation Laboratories, and Gamification. Consequently, the overall understanding of ALP's potential as a learning tool within an educational context remains underdeveloped, hindering the application of informed pedagogical or didactic choices informed by relevant theories.

For future research endeavors, a comprehensive examination of multimodal features (C5) is warranted, delving into the nuanced ways various modes of interaction influence the efficacy of LA and LADs in ALPs. Subsequent investigations should scrutinize the deep or implicit structures underpinning the pedagogical implementation of ALP's LA and LAD (C6). Methodologies ought to be developed for aligning robust pedagogical theories and concepts, thereby informing the development and assessment processes. Furthermore, future studies should incorporate evaluations encompassing teachers' perspectives to attain a holistic understanding of the impact of LA and LAD on ALPs implementation.

An essential question emerges: How can we enhance the quality of learning and teaching through LA when data collection and presentation lack a foundational methodology, framework, or theory? This query underscores the need to associate actions with data rather than merely presenting metrics such as learning objectives' difficulty, time spent on the platform, or active users. The proposed future investigations must intricately link pedagogy to LA and LAD of ALPs, providing essential support for teachers and students as they navigate cognitive and meta-cognitive impacts, behavioral changes, and social learning activities.

In essence, the envisioned assessments should integrate ontologies, frameworks, methodologies, experimental designs, mathematical models, and LA statistics—comprehensive components constituting the foundational elements of LA. Notably absent in the current landscape are evaluations of pedagogical elements, a critical gap that future research must address to comprehensively advance the field.

References

Abech, M. *et al.* (2016) 'A model for learning objects adaptation in light of mobile and context-aware computing', *Personal and Ubiquitous Computing*, 20(2), pp. 167–184. Available at: https://doi.org/10.1007/s00779-016-0902-3

Al-Shanfari, L. *et al.* (2020) 'Visualising alignment to support students' judgment of confidence in open learner models', *User Modeling and User-Adapted Interaction*, 30(1), pp. 159–194. Available at: https://doi.org/10.1007/s11257-019-09253-4

Beckmann, J., Bertel, S. and Zander, S. (2015) 'Performance & Emotion-A Study on Adaptive E-Learning Based on Visual/Verbal Learning Styles', in *International Association for Development of the Information Society*. *International Conference on e-Learning*, Las Palmas de Gran Canaria, Spain: International Association for the Development of the Information Society, pp. 27–34. Available at: https://eric.ed.gov/?id=ED562495 (Accessed: 9 November 2023).

Bresó, A. *et al.* (2016) 'A novel approach to improve the planning of adaptive and interactive sessions for the treatment of Major Depression', *International Journal of Human-Computer Studies*, 87, pp. 80–91. Available at: https://doi.org/10.1016/j.ijhcs.2015.11.003.

Di Mascio, T. *et al.* (2013) 'Design choices: affected by user feedback? affected by system performances?: lessons learned from the TERENCE project', in *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*. *CHItaly '13: Biannual conference of the Italian chapter of SIGCHI*, Trento Italy: ACM, pp. 1–9. Available at: https://doi.org/10.1145/2499149.2499171.

Dirin, A., Laine, T.H. and Nieminen, M. (2017) 'Sustainable usage through emotional engagement: a user experience analysis of an adaptive driving school application', *Cognition, Technology & Work*, 19(2), pp. 303–313. Available at: https://doi.org/10.1007/s10111-017-0406-6.

Dounas, L., Salinesi, C. and Beqqali, O.E. (2019) 'Requirements Monitoring and Diagnosis for Improving Adaptive E-Learning Systems Design', *Journal of Information Technology Education: Research*, pp. 1–20.

Effenberger, T. and Pelánek, R. (2019) 'Measuring Students' Performance on Programming Tasks', in *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*. *L@S '19: Sixth (2019) ACM Conference on Learning @ Scale*, Chicago IL USA: ACM, pp. 1–4. Available at: https://doi.org/10.1145/3330430.3333639.

Fadljević, L. *et al.* (2020) 'Slow is good: the effect of diligence on student performance in the case of an adaptive learning system for health literacy', in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. *LAK '20: 10th International Conference on Learning Analytics and Knowledge*, Frankfurt Germany: ACM, pp. 112–117. Available at: https://doi.org/10.1145/3375462.3375502.

Fasihuddin, H., Skinner, G. and Athauda, R. (2017) 'Towards adaptive open learning environments: Evaluating the precision of identifying learning styles by tracking learners' behaviours', *Education and Information Technologies*, 22(3), pp. 807–825. Available at: https://doi.org/10.1007/s10639-015-9458-5.

Hewitt-Taylor, J. (2001) 'Use of constant comparative analysis in qualitative research', *Nursing Standard*, 15(42), pp. 39–42. Available at: https://doi.org/10.7748/ns2001.07.15.42.39.c3052.

Hooshyar, D. *et al.* (2018) 'SITS: A solution-based intelligent tutoring system for students' acquisition of problem-solving skills in computer programming', *Innovations in Education and Teaching International*, 55(3), pp. 325–335. Available at: https://doi.org/10.1080/14703297.2016.1189346.

Hsu, W.-C. and Li, C.-H. (2015) 'A competency-based guided-learning algorithm applied on adaptively guiding e-learning', *Interactive Learning Environments*, 23(1), pp. 106–125. Available at: https://doi.org/10.1080/10494820.2012.745432.

Katuk, N., Kim, J. and Ryu, H. (2013) 'Experience beyond knowledge: Pragmatic e-learning systems design with learning experience', *Computers in Human Behavior*, 29(3), pp. 747–758. Available at: https://doi.org/10.1016/j.chb.2012.12.014.

Khawaja, M.A., Chen, F. and Marcus, N. (2014) 'Measuring Cognitive Load Using Linguistic Features: Implications for Usability Evaluation and Adaptive Interaction Design', *International Journal of Human–Computer Interaction*, 30(5), pp. 343–368. Available at: https://doi.org/10.1080/10447318.2013.860579.

Kim, J., Jo, I.-H. and Park, Y. (2016) 'Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement', *Asia Pacific Education Review*, 17(1), pp. 13–24. Available at: https://doi.org/10.1007/s12564-015-9403-8.

Latham, A. *et al.* (2012) 'Adaptive Tutoring in an Intelligent Conversational Agent System', in N.-T. Nguyen (ed.) *Transactions on Computational Collective Intelligence VIII*. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), pp. 148–167. Available at: https://doi.org/10.1007/978-3-642-34645-3_7.

Latham, A., Crockett, K. and McLean, D. (2014) 'An adaptation algorithm for an intelligent natural language tutoring system', *Computers & Education*, 71, pp. 97–110. Available at: https://doi.org/10.1016/j.compedu.2013.09.014.

Lau, S.B.-Y., Lee, C.-S. and Singh, Y.P. (2015) 'A folksonomy-based lightweight resource annotation metadata schema for personalized hypermedia learning resource delivery', *Interactive Learning Environments*, 23(1), pp. 79–105. Available at: https://doi.org/10.1080/10494820.2012.745429.

Lo, J.-J., Chan, Y.-C. and Yeh, S.-W. (2012) 'Designing an adaptive web-based learning system based on students' cognitive styles identified online', *Computers & Education*, 58(1), pp. 209–222. Available at: https://doi.org/10.1016/j.compedu.2011.08.018.

Lynch, T. and Ghergulescu, I. (2017) 'Large Scale Evaluation of Learning Flow', in *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*. *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, pp. 62–64. Available at: https://doi.org/10.1109/ICALT.2017.98.

Martin, F., Dennen, V.P. and Bonk, C.J. (2020) 'A synthesis of systematic review research on emerging learning environments and technologies', *Educational Technology Research and Development*, 68(4), pp. 1613–1633. Available at: https://doi.org/10.1007/s11423-020-09812-2.

Mavroudi, A. *et al.* (2016) 'Teacher-led design of an adaptive learning environment', *Interactive Learning Environments*, 24(8), pp. 1996–2010. Available at: https://doi.org/10.1080/10494820.2015.1073747.

Mousavinasab, E. *et al.* (2021) 'Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods', *Interactive Learning Environments*, 29(1), pp. 142–163. Available at: https://doi.org/10.1080/10494820.2018.1558257.

Nye, B.D. *et al.* (2021) 'Feasibility and usability of MentorPal, a framework for rapid development of virtual mentors', *Journal of Research on Technology in Education*, 53(1), pp. 21–43. Available at: https://doi.org/10.1080/15391523.2020.1771640.

Page, M.J. *et al.* (2021) 'The PRISMA 2020 statement: An updated guideline for reporting systematic reviews', *International Journal of Surgery*, 88, p. 105906. Available at: https://doi.org/10.1016/j.ijsu.2021.105906.

Papamitsiou, Z. *et al.* (2020) 'Utilizing Multimodal Data Through fsQCA to Explain Engagement in Adaptive Learning', *IEEE Transactions on Learning Technologies*, 13(4), pp. 689–703. Available at: https://doi.org/10.1109/TLT.2020.3020499.

Rincón-Flores, E.G. *et al.* (2019) 'Adaptive learning based on AI with predictive algorithms', in *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*. New York, NY, USA: Association for Computing Machinery (TEEM'19), pp. 607–612. Available at: https://doi.org/10.1145/3362789.3362869.

Santos, O.C. *et al.* (2016) 'Toward interactive context-aware affective educational recommendations in computer-assisted language learning', *New Review of Hypermedia and Multimedia*, 22(1–2), pp. 27–57. Available at: https://doi.org/10.1080/13614568.2015.1058428.

Santos, O.C. and Boticario, J.G. (2015) 'Practical guidelines for designing and evaluating educationally oriented recommendations', *Computers and Education*, 81, pp. 354–374. Available at: https://doi.org/10.1016/j.compedu.2014.10.008.

Sharma, K., Papamitsiou, Z. and Giannakos, M. (2019) 'Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach', *British Journal of Educational Technology*, 50(6), pp. 3004–3031. Available at: https://doi.org/10.1111/bjet.12854.

Shemshack, A. and Spector, J.M. (2020) 'A systematic literature review of personalized learning terms', *Smart Learning Environments*, 7(1), p. 33. Available at: https://doi.org/10.1186/s40561-020-00140-9.

Shulman, L.S. (2005) 'Signature Pedagogies in the Professions', *Daedalus*, 134(3), pp. 52–59.

Tahmasebi, M., Ghazvini, F.F. and Esmaeili, M. (2018) 'Implementation and Evaluation of a Resource-Based Learning Recommender Based on Learning Style and Web Page Features', *Journal of Web Engineering*, pp. 284–304.

Tlili, A. *et al.* (2023) 'Automatic modeling learner's personality using learning analytics approach in an intelligent Moodle learning platform', *Interactive Learning Environments*, 31(5), pp. 2529–2543. Available at: https://doi.org/10.1080/10494820.2019.1636084.

Tretow-Fish, T.A.B. and Khalid, Md.S. (2022) 'Evaluating Learning Analytics of Adaptive Learning Systems: A Work in Progress Systematic Review', in E. Brooks, J. Sjöberg, and A.K. Møller (eds) *Design, Learning, and Innovation*. Cham: Springer International Publishing (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), pp. 37–52. Available at: https://doi.org/10.1007/978-3-031-06675-7_3.

Troussas, C. *et al.* (2020) 'Intelligent and adaptive tutoring through a social network for higher education', *New Review of Hypermedia and Multimedia*, 26(3–4), pp. 138–167. Available at: https://doi.org/10.1080/13614568.2021.1908436.

Yuksel, B.F. *et al.* (2016) 'Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty Based on Brain State', in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI '16), pp. 5372–5384. Available at: https://doi.org/10.1145/2858036.2858388.

Zhang, L. *et al.* (2023) 'Evaluation of a student-centered online one-to-one tutoring system', *Interactive Learning Environments*, 31(7), pp. 4251–4269. Available at: https://doi.org/10.1080/10494820.2021.1958234.