

# Architecting Analytics Across Multiple E-learning Systems to Enhance Learning Design

Katerina Mangaroska, Boban Vesin, Vassilis Kostakos, Peter Brusilovsky, *Senior member, IEEE*, and Michail Giannakos, *Senior member, IEEE*,

**Abstract**—With the wide expansion of distributed learning environments the way we learn became more diverse than ever. This poses an opportunity to incorporate different data sources of learning traces that can offer broader insights into learner behavior and the intricacies of the learning process. We argue that combining analytics across different e-learning systems can potentially measure the effectiveness of learning designs and maximize learning opportunities in distributed settings. As a step toward this goal, in this study, we considered how to broaden the context of a single learning environment into a learning ecosystem that integrates three separate e-learning systems. We present a cross-platform architecture that captures, integrates, and stores learning-related data from the learning ecosystem. To demonstrate the feasibility and the benefits of cross-platform architecture, we used regression and classification techniques to generate interpretable models with analytics that can be relevant for instructors in understanding learning behavior and sensemaking of the instructional method on learner performance. The results show that combining data across three e-learning systems improve the classification accuracy compared to data from a single learning system by a factor of 5. Our work highlights the value of cross-platform learning analytics and presents a springboard for the creation of new cross-system data-driven research practices.

**Index Terms**—Cross-platform analytics, architecture for educational systems, distributed learning settings, distance education.

## I INTRODUCTION

DIGITAL learning has grown significantly with the rapid expansion of Information and Communication Technology (ICT) and the concept of ubiquitous computing. This trend is a catalyst for learning to happen everywhere and at any time, across many different platforms and learning systems, situated and shaped by the tasks, the content resources, and the dynamics of distributed learning environments [1]. Although learning happens anytime and across many diverse learning settings, we still lack insights how to effectively optimize the learning context and the learner experience in these settings [2]. In addition, as learners are rapidly embracing the use of novel data-intensive learning technologies, they are becoming more demanding and critical, creating a challenge how to engage and support them when learning takes place in distributed settings [3].

One promising approach lies in a wider application of learning-related data collected from various e-learning systems, that once merged, can support a learning ecosystem of “dynamic, interconnected, and ever-evolving community of learners, instructors, tools, and content” [4]. The idea draws on the work presented in [5], who advocates that understanding learning and knowledge creation in dis-

tributed settings, requires multi-level analyses on learners’ traces fragmented across time, numerous e-learning systems, and media (i.e., digital substrates where communication modes are encoded). However, mainstream methods and tools often rely on *metrics derived from single and many times limited data sources* such as grades, submission of assignments, self-reported data, or test performances [6]. On the one hand, findings based on metrics extracted from limited data sources, represent only a small proportion of the learning process and the activities students engage with. This, in turn, only partially help educators to understand *when and how students learn*, and *how effectively they use the opportunities for learning as given in the learning design*. Thus, current approaches often display the ongoing limitations in the learning analytics field, in which many researchers and educators miss the opportunity to make effective and meaningful refinements in the learning designs that can encourage, enable, and advance learning.

On the other hand, learners often make decisions (e.g., whether and what technologies to use) based on the perception of what might maximize their chances to succeed [7]; hence, their focus is often on assessment [8], [9]. However, changes in the instructional methods (i.e., adding personalized feedback) can change learners’ single focus on assessment [10] and usage of technologies for reasons other than solely succeeding in the course assessments [6]. Past research has shown that learning design and the instructional conditions strongly affect what technologies and tools students use [6], [11], as well as their level of engagement and performance [12]. Yet, we do not really know *how to create and measure the effectiveness of learning designs that can maximize learning opportunities in distributed learning settings*. In that respect, learning

Manuscript received May 27, 2020; revised March 9, 2021; accepted ???, 2021. (Corresponding author: Boban Vesin)

K. Mangaroska and M. Giannakos are with with Department of Computer Science, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway (e-mail: mangaroska@ntnu.no; michailg@ntnu.no).

B. Vesin is with the School of Business, University of South-Eastern Norway, Vestfold, Norway (e-mail: boban.vesin@usn.no).

V. Kostakos is with School of Computing and Information Systems, University of Melbourne, Australia (e-mail: vassilis.kostakos@unimelb.edu.au).

P. Brusilovsky is with School of Information Sciences, University of Pittsburgh, Pittsburgh PA 15260 USA (e-mail: peterb@pitt.edu).

Digital Object Identifier ???/TLT.???

57 analytics has the potential to provide insights into what is  
 58 happening in each and across the different learning systems,  
 59 and thus, examine the effectiveness of the learning design.  
 60 For example, if educators have learning-related data from  
 61 other systems learners use to master skills (e.g., GitHub for  
 62 learning programming), rather than solely from the assigned  
 63 tasks and the learning management system (LMS) in use,  
 64 educators can have improved overview of learners' progress  
 65 and potential misconceptions, and make pertinent decisions  
 66 to (re)design the learning context when learning behavior  
 67 deviates from the pedagogical intention.

68 Consequently, we designed and implemented a study to  
 69 explore if *architecting analytics across multiple e-learning*  
 70 *systems can enhance the analytics capacities of the individual*  
 71 *systems*, and discussed how the findings can support  
 72 and enhance learning design practices in distributed settings  
 73 [6], [13], [14]. However, developing cross-platform systems  
 74 is a complex and data intensive process [15]. On the one  
 75 hand, the procedure of standards development is inherently  
 76 challenging [16] and there is a lack of data interoperability  
 77 standards for handling and processing data generated from  
 78 different systems [17]. On the other hand, cross-platform  
 79 systems additionally increase the complexity of orchestrating  
 80 learning activities in distributed settings, as educators  
 81 need to deal with the requirements that stem from different  
 82 learning designs [18], [19]. Therefore, it was also necessary  
 83 to *explore and define the minimum technical architecture*  
 84 *requirements* essential for setting the foundations to develop  
 85 cross-platform systems.

86 To demonstrate proof of concept, we implemented  
 87 a cross-platform architecture that integrates three inde-  
 88 pendent personalized e-learning systems (i.e., ProTuS,  
 89 MasteryGrids, and Visual Learning Analytics System for  
 90 Programming–VLASP) into one learning ecosystem for data  
 91 collection, integration, and harmonization. Since these sys-  
 92 tems automatically capture all user interactions, learning  
 93 analytics was utilized to shed light and give rise to a larger  
 94 phenomenon in digital learning and predictive modeling—  
 95 *how to develop predictive models that not only predict per-*  
 96 *formance and success, but also reveal significant elements*  
 97 *for teaching practice* (e.g., generic and specific), that can  
 98 be applied to improve the quality of learning designs and  
 99 instructional methods [20]. As a result, we present the po-  
 100 tential of cross-platform learning analytics generated from  
 101 behavioral log data, and utilized to show how predictive  
 102 models can be constituted to inform teaching practices as  
 103 a “diagnostic” tool that can support data-driven changes  
 104 in the learning design, pertinent to the optimization of the  
 105 various technologies used during the course. In that regard,  
 106 we addressed the following research questions:

- 107 • **RQ1:** What are the benefits of implementing cross-  
 108 platform architecture and harnessing cross-platform  
 109 learning analytics for digital education?
- 110 • **RQ2:** What implications cross-platform learning an-  
 111 alytics can offer to learning design?

112 In sum, the contribution of this paper is threefold: 1)  
 113 *Conceptual*—to present an ideational model of a digital  
 114 learning ecosystem which supports and harnesses cross-  
 115 platform analytics, 2) *Operational*—to display the implemen-  
 116 tation of a cross-platform architecture, and 3) *Empirical*—

117 to validate the value of cross-platform data integration for  
 118 building predictive models that carry the opportunity to  
 119 reveal significant elements for teaching practice, rather than  
 120 the long-standing focus on identifying learners at risk of  
 121 failing a course or solely for predicting learner performance,  
 122 as it is commonly done in the learning analytics and edu-  
 123 cational data mining communities.

## II BACKGROUND

124 Optimizing the learning context and making valid and  
 125 informed changes in the learning design utilizing learning-  
 126 related data, was probably one of the first motivations  
 127 for the emergence of learning analytics [21]. Nowadays,  
 128 the wide proliferation of distributed learning environments  
 129 gives rise to opportunities in learning analytics and pre-  
 130 dictive modeling, to explore how analytics from various  
 131 learning systems (i.e., cross-platform learning analytics) can  
 132 be harnessed to enhance the quality of learning designs and  
 133 instructional methods.

### II.1 Cross-Platform Learning Analytics

135 Current research often relies on metrics derived from data  
 136 sources such as grades, submission of assignments, the  
 137 time learners spent in e-learning, self-reported data, or test  
 138 performances [6]. However, it often falls behind to consider  
 139 data from more than one learning system, particularly when  
 140 learning happens in distributed settings [22]. Nonetheless,  
 141 the more complex data researchers capture across settings  
 142 (e.g., interactions with learning materials via LMSs, learn-  
 143 ing trajectories via problem-based learning), the harder it  
 144 becomes to synchronize and analyze that data [23], [24].  
 145 Although frameworks that describe how to capture and  
 146 classify data from different sources exist [25]–[27], there is a  
 147 lack of available tools that could assist researchers to easily  
 148 establish cross-platform and sometimes multimodal systems  
 149 [28]–[30].

150 Learning is distributed across multiple media, locations,  
 151 and online environments; yet, researchers' scope is often  
 152 limited to a single virtual learning environment (VLE) or  
 153 LMS. This is a common drawback in the field of learning an-  
 154 alytics that depicts the present-day reality where researchers  
 155 depend on one-sided learning analytics measures due to  
 156 the difficulties of extracting, harmonizing, and sensemaking  
 157 of data from various sources [2]. In a previous work, a  
 158 conceptual model named Group Learning Unified Environ-  
 159 ment with Pedagogical Scripting, Monitoring, Analysis, and  
 160 Across-Spaces Support (GLUEPS-MAASS) was presented,  
 161 describing how data from multiple sources should be col-  
 162 lected and integrated, encompassing learning activities in  
 163 the web, the physical, and the 3D virtual space [31]. How-  
 164 ever, to the best of our knowledge, this is still a conceptual  
 165 model, that has not been placed into practice yet. Moreover,  
 166 there are a few cross-platform patent models [32]–[34] that  
 167 currently have a pending status.

168 Consequently, we try to overcome some of the on-  
 169 going issues (e.g., one-sided learning analytics measures,  
 170 data integration, and interoperability) by *proposing a cross-*  
 171 *platform architecture that automatically collects, integrates,*  
 172 *and harmonizes data from several e-learning systems* (i.e., Pro-  
 173 TuS, MasteryGrids, and VLASP). These data is then used

175 to explore if *combining metrics extracted from multi-system log*  
 176 *data can increase the predictive power* of the individual systems  
 177 with respect to estimating student performance at the end  
 178 of the course, as well as reveal significant metrics that can  
 179 further *refine the personalization and the design of learning*  
 180 *activities and instructional methods*.

## 181 **II.2 Predictive Modeling in Learning Analytics**

182 Data-driven approaches to further our understanding of  
 183 learning are particularly relevant for e-learning and learning  
 184 analytics research [35], with predictive modeling being an  
 185 important topic [36]. Many researchers have already utilized  
 186 predictive modeling techniques to identify students  
 187 at risk and increase their retention [37], [38], to provide  
 188 early insights about students' performance and to generate  
 189 interpretable performance models [39]–[41], to improve the  
 190 quality and to scale up feedback [42], and to create intervention  
 191 methods that can improve students' mental health and  
 192 their university experience [43].

193 Predictive modeling involves statistical models or data  
 194 mining algorithms to find patterns in the data, and predict  
 195 new or future events [44]. Most of the research in predictive  
 196 modeling forecasts what may happen, and as such has  
 197 shown opportunities for advancing the field of learning  
 198 analytics. However, it has yet to mature to offer a wide-scale  
 199 impact [45]. In other words, predictive models intent to offer  
 200 actionable insights for learners and instructors, so that these  
 201 groups of users can take further actions, rather than increase  
 202 the frequency of feedback (i.e., informing students how  
 203 they stand with respect to meeting the course list criteria)  
 204 [46]. Hints (e.g., the traffic light metaphor in [37]) whether  
 205 learners are at risk of not meeting certain course criteria,  
 206 can help learners to be aware of their current progress, they  
 207 do not offer much beyond that [45]. Therefore, learners and  
 208 instructors do not always find much value and use of such  
 209 predictive models, as these models are limited in provoking  
 210 reflection and action [46], [47]. In the same vein, predictive  
 211 models do not always generate actionable insights,  
 212 resulting in limited information for instructors to improve  
 213 their practices and the overall learning design [45], [46]. A  
 214 step toward a more insightful and actionable information  
 215 can be generated by combining predictive with explanatory  
 216 modeling [44], to develop interpretable models with under-  
 217 lying variables that are relevant for instructors and learners  
 218 in understanding learning behavior and making sense of  
 219 instructional methods on learning performance [20].

220 To move beyond predictive analytics [20], [47]–[49] and  
 221 to investigate and apply sophisticated and innovative ap-  
 222 proaches, we focused on *harnessing cross-platform learning*  
*223 analytics in predictive modeling*. This way we aimed to explore  
 224 if combining analytics across various systems can increase  
 225 the predictive power of individual learning systems with  
 226 respect to estimating student performance (i.e., grades), as  
 227 well as develop models that reveal significant elements  
 228 for teaching practice, that in future, can *help learners, to*  
*229 understand the value of different learning resources* apart from  
 230 solely maximizing their chances to succeed by getting good  
 231 grades.

## 232 **II.3 Standards, Integration, and Interoperability of** **233 Learning Systems**

234 Systems operate by "understanding" the data structures  
 235 they share [50]. Therefore, to perform meaningful analysis  
 236 and produce applicable outcomes, interoperability of  
 237 data format is paramount. The interoperability challenge  
 238 is present in the learning analytics community, creating  
 239 obstacles in implementing a standardized specification at  
 240 scale that each "data supplier" or "tools developer" has  
 241 to conform to [16]. In fact, "interoperability and scalability  
 242 are evolution features embodied in the architecture of the  
 243 software system" [16, p.32]. Considering this issue, much  
 244 research in technology-enhanced learning has been focused  
 245 on enhancing interoperability [51]. Thus, several conceptual  
 246 frameworks [52] and software architectures [16], [53], [54]  
 247 have been proposed to effectively store and retrieve large  
 248 amounts of data generated in e-learning settings.

249 The interoperability issue is not a new one. Since 2001,  
 250 several learning resource specifications have been developed,  
 251 including ADL (Advanced Distributed Learning),  
 252 SCORM (Shareable Content Object Reference Model) [55],  
 253 IMS Learning Resource Metadata Specification [56], and  
 254 IEEE LOM (Learning Object Metadata) [57]. These examples  
 255 have been considered as drivers toward re-usability and  
 256 interoperability of learning resources [58]. Furthermore, sev-  
 257 eral industrial solutions, such as the Learning Tools Inter-  
 258 operability [59] and the Experience API (xAPI) [60], [61] are  
 259 widely applied, to enhance the interoperability of e-learning  
 260 systems and tools [16]. xAPI is a standardized approach that  
 261 clarifies how the collection, storage, analysis, and exploita-  
 262 tion of data are taking place. The prominence of xAPI con-  
 263 sists of system independence, easy implementation, and the  
 264 focus on learner activities [62]. Slowly, but effectively, xAPI  
 265 specification [61] emerged as a standard vocabulary for  
 266 communication with distributed data in learning systems,  
 267 due to its inherent extensibility to accommodate unforeseen  
 268 data collection needs.

269 Findings from past research present an architecture that  
 270 tackles the challenge of collecting and managing data from a  
 271 variety of services and feeds, and with a focus on simplicity  
 272 and flexibility [30]. The work published in [30] emphasized  
 273 the implementation and the importance of trackers as main  
 274 connectors between the activity provider, the LMS, and the  
 275 data storage component. To that end, our work aims to *set*  
*276 up a learning ecosystem consisting of several integrated e-learning*  
*277 systems that rely on distributed and diverse data, which will*  
*278 satisfy the requirements for data format interoperability*  
*279 and harness the potential of combining cross-platform learning*  
*280 analytics*. Moreover, with the *proposed cross-platform archi-  
 281 tecture*, we aim to present a *proof-of-concept emphasizing the*  
*282 importance of holistic understanding of learners' behavior and*  
*283 progress, relevant for supporting data-driven changes in the learn-  
 284 ing design, and toward improving and sustaining student*  
*285 engagement utilizing personalized feedback methods* [10].

## 286 **III ARCHITECTURE OF THE PROPOSED LEARNING** **287 ECOSYSTEM**

288 Our motivation for designing and developing a cross-  
 289 platform architecture lies in:

- 290 1) Offering a modular system that can be easily modified by adding new data sources;
- 291 2) Exploring the trade-off between interoperability, 348
- 292 flexibility, and scalability of the system; 349
- 293 3) Initiating communication among various stakeholders 350 (designers, educators, students) to investigate 351 how learning analytics might contribute to personalization 352 and flexibility vs. scalability and standardization 353 of learning; 354
- 294 4) Demonstrating proof-of-concept for the feasibility 355 and the potential of combining analytics across various e-learning systems.

302 The proposed architecture is developed addressing five 303 core functionalities (see Table 1) that *the next generation of 304 learning ecosystems* should have [4]. In addition, considering 305 the nature of the learning setting, the following are the 306 requirements that have been taken into account during the 307 design of the architecture, as suggested by [63]:

- 308 • **Data accessibility.** Accessibility and data latency 309 are two crucial factors that affect data usage for 310 instructional improvement [64]. Therefore, the 311 architecture model integrates heterogeneous data using 312 APIs for mining and retrieving common data formats 313 such as JavaScript Object Notation (JSON), comma- 314 separated values (CSV), or database storage. This 315 way, the infrastructure can support and promote 316 standardization, while facilitating data integration 317 and harmonization [65], [66].
- 318 • **Extensibility.** The design needs to follow modular 319 architecture with clearly defined and separated 320 components. This approach increases the extensibility of 321 the system and decreases the level of effort required 322 to implement future functionalities [67].
- 323 • **Scalability.** Although scalability is a growing concern 324 for e-learning systems [68], majority of these 325 systems are implemented to support their current 326 users, with less consideration for future user-base 327 growth. The model of our proposed architecture aims 328 to provide a better code structure, ability to run as 329 a distributed application with faster resource usage, 330 and thus, support future scaling of the user base.

331 To reach the goals of the proposed architecture, the 332 design and development stages followed approaches defined 333 in system development research [69], best practices in 334 software design [66], and principles of software engineering 335 for learning systems [70]. Thus, the architectural design 336 decisions have been emphasized through two views [71]:

- 337 • **The conceptual view** shows the composition of the 338 concepts necessary for system execution. This view 339 represents the conceptual model of the system and 340 explains the communication and data aggregation 341 processes between the different components.
- 342 • **The implementation view** shows the topology of the 343 implemented solution, the architectural layers, and 344 the physical connection between the three e-learning 345 systems.

346 The conceptual view represents the generic overview of 347 the system and contains the elements required for collecting

348 cross-platform data and analytics, while the implementation 349 view presents the actual execution of the proposed 350 architecture with the use of several existing systems as data 351 providers. Since the purpose of the study is to demonstrate 352 proof-of-concept of the feasibility and applicability of 353 analytics across learning systems, this paper presents only the 354 overall idea that lies behind the proposed cross-platform 355 architecture.

### III.1 Conceptual Model of Cross-Platform Architecture

356 To minimize challenges (e.g., data formats, undocumented 357 data, or noise in the data) when working with multiple data 358 streams, we propose a conceptual model that promotes and 359 supports integration and interoperability among various 360 data sources. The aim is to develop an integrated ecosystem, 361 that would eliminate the need to manually log in, gather, 362 and synchronize data from different systems. The proposed 363 integration encompasses several functional layers as shown 364 in Fig. 1:

- 366 • **The data processing layer** imports, aggregates, 367 transforms, normalizes, and processes data. This 368 layer is responsible for collecting and preparing data 369 for further use and analysis.
- 370 • **The data analysis layer** interacts with the stored data 371 to extract business intelligence.
- 372 • **The report generator** visualizes data and generates 373 reports based on educators and designers' preferences.
- 374 • **The data source layer (i.e., learning record store)** 375 stores data in standardized and consistent format.
- 376 • **The application front end (i.e., learning analytics 377 dashboard)** accommodates different reports, visualizations, 378 and solutions, for report customization and 379 personalized feedback.

### III.2 The Implemented Architecture of the Integrated System

381 This section presents the implemented learning ecosystem 382 that encompasses three e-learning systems, i.e., ProTuS, 383 MasteryGrids, and VLASP. The proposed architecture of 384 the learning ecosystem aggregates data from four different 385 data providers, and thereby, supports cross-platform learning 386 analytics. The following are the e-learning systems we 387 integrated:

- 388 • **ProTuS** is an intelligent e-learning system for learning 389 programming basics. ProTuS allows educators to 390 design and implement their own learning content, 391 in addition to the option for easy integration of 392 learning content from third-party providers, such 393 as wiki pages or YouTube videos. For this study, 394 lectures from Confluence wiki pages were used to 395 cover the basic Java concepts. ProTuS also provides 396 personalization techniques and several methods for 397 recommending learning content [72].
- 398 • **MasteryGrids** is an open social learner modeling 399 interface, written in JavaScript [73]. The interface 400 shows learners' progress in different topics 401 compared to other learners or the class. It also provides 402

TABLE 1  
The Core Functionalities of a Learning Ecosystem

Goals	Core functionalities
Identify learners' characteristics, goals, skills, strategies, and needs.	Personalization
Monitor, assess, and predict students' behavior, progress, and performance.	Analytics and learning assessment
Process, interpret, and utilize data across learning systems.	Interoperability and integration
Provide real-time actionable feedback.	Advising and support
Visualize metrics based on cross-platform analytics and educational theories.	Explanation and interpretation

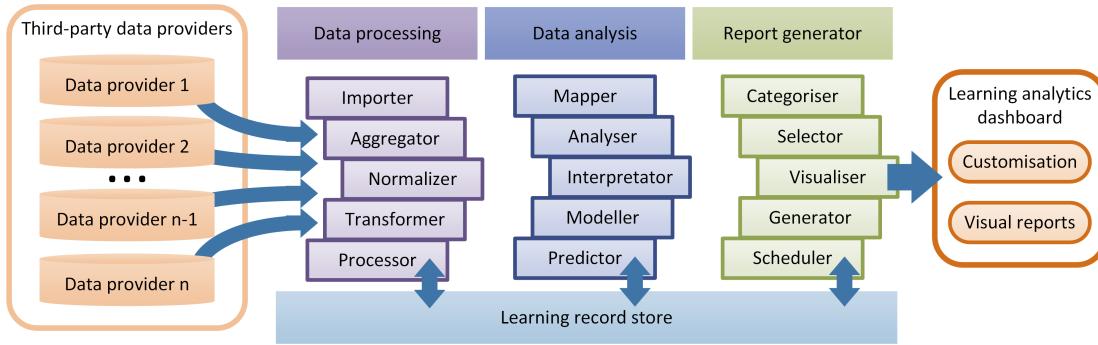


Fig. 1. The general architecture.

adaptive navigation support for learning content with stars indicating recommendations. The system tracks learners' activities and updates learner knowledge levels in a centralized user modeling server. This allows MasteryGrids to report the progress level (i.e., based on activities) and the knowledge level (i.e., based on estimated learner knowledge). MasteryGrids collects activity data from two data providers:

- 404 – **PCLab** includes interactive examples and challenges developed at University of Pittsburgh 405 [74]. The system tracks learner activity, including 406 students' trial and error approaches.
- 407 – **Programming Course Resource System** 408 (**PCRS**) includes coding exercises developed 409 at University of Toronto [75]. This system tests 410 learners' solutions against a set of unit tests 411 for a particular problem, while the results are 412 stored in the data source layer.
- 433 • **Visual learning analytics system for programming** 434 (**VLASP**) is an Eclipse plug-in that monitors learners' 435 progress in programming, tracks learner behavior 436 while learners develop/debug code in Java, and 437 reflects progress to learners as a mirroring tool [76]. 438 The environment monitors progress and visualizes 439 metrics (e.g., how many times a student has run 440 an individual test, how many times the code has 441 been compiled) associated with learner behavior and 442 performance during programming/debugging activities. 443 The main goal of the system is twofold: 1) to 444 collect data about learner activities, so that educators 445 can better understand how learners program/debug; 446 and 2) to mirror learners' own actions back to them, 447 as a way to increase awareness and motivation, foster 448

449 self-reflection, and facilitate improvements in their 450 programming habits [77].

451 ProTuS, MasteryGrids, and VLASP are separately 452 designed and implemented systems; thus, their data 453 models are different. The integrated learning 454 environment has to provide access to different 455 data structures, combine those 456 data structures, and harmonize the data 457 formats. Therefore, Visualized Education 458 NTNUI (VENT) [78] has been created 459 and presented as a layer on top of the modules of each 460 data source, consisting of a VENT system object 461 notation (VSON) 462 model and a VENT controller. This 463 layer contains the data 464 source controllers that act as conversion 465 layers from the 466 source model (e.g., JSON format) to VSON 467 format which is 468 then exposed by the VENT controller. Finally, because 469 three 470 e-learning 471 systems were utilized in this 472 study, we selected 473 ProTuS to be a portal for seamless 474 integration of different 475 content 476 providers. The overview of the data 477 sources and 478 integrated 479 learning 480 environments employed in the 481 study is 482 shown in Fig. 2.

## IV METHODOLOGY

### IV.1 Research Approach

459 The approach adopted in this study is based on 460 design-based research (DBR) [79]. DBR utilizes an 461 iterative process of design, implementation, 462 analysis, and revision of models, 463 with two primary goals: to construct 464 knowledge and to 465 develop solutions [80]. Hence, a series of DBR 466 cycles were 467 performed to develop the learning 468 analytics component (i.e., 469 first DBR cycle) [72], the 470 adaptability 471 feature, i.e., 472 adaptive 473 assessment (i.e., 474 the second DBR cycle) [81], and the 475 cross- 476 platform 477 architecture (i.e., 478 the third DBR cycle) [82].

479 In the first DBR cycle, a focus group was organized with 480 12 teaching 481 assistants (TAs), to understand and generate 482

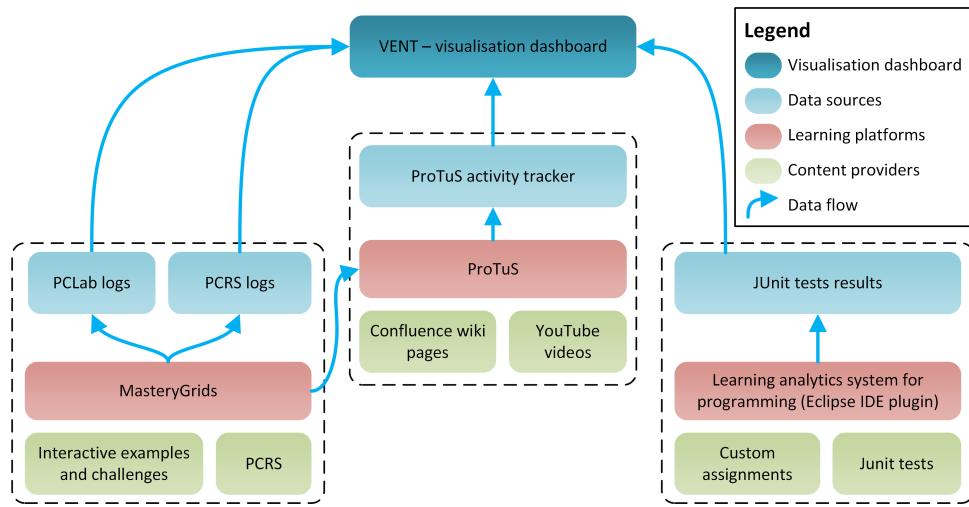


Fig. 2. Aggregated data sources.

470 the best practices they had accumulated over the last few  
 471 years, by closely working with students from introductory  
 472 programming courses. The TAs were computer science (CS)  
 473 majors, that were in their third or fourth semester of bachelor  
 474 CS studies at the Norwegian University of Science and  
 475 Technology (NTNU). The focus in the first DBR cycle  
 476 was on participatory [83] and human-centered [84] design  
 477 approaches in the development of the learning analytics  
 478 component. These approaches were employed to support  
 479 the design of seamless user experience in personalized e-  
 480 learning systems [3]. Applying affinity diagram technique  
 481 and usability survey, we transformed the generated best  
 482 practices into design guidelines and applied them in the  
 483 second DBR cycle.

484 In the second DBR cycle we focused more on learners'  
 485 behavior and requirements, because personalized e-learning  
 486 systems need to acknowledge and model users' natural  
 487 behavior, so that the interaction is intuitive and minimizes  
 488 users' cognitive workload. Therefore, we designed an ex-  
 489 periment to explored how students interact with the new  
 490 learning analytics module in ProTuS. The purpose of the  
 491 second DBR cycle was to explore learners' trajectories dur-  
 492 ing five quiz activities. A total of 66 students participated  
 493 in the study and each student was asked to fill out one  
 494 quiz at a time. After every quiz, the students were asked  
 495 to reflect and monitor its own progress with the help of  
 496 the generated reports utilizing learning analytics, and then  
 497 continue to the next quiz assignment. All 66 students were  
 498 CS majors in their second semester of bachelor studies. The  
 499 insights generated from this study were used to develop the  
 500 adaptive assessment feature in ProTuS.

501 The last DBR cycle is the focus of this study, which is the  
 502 development, the implementation, and the evaluation of the  
 503 proposed cross-platform architecture. All three DBR cycles  
 504 have used the framework for modeling personalization di-  
 505 mensions proposed by [85]. This framework was selected  
 506 to develop personalization features in ProTuS following six  
 507 personalization dimensions in intelligent tutoring systems  
 508 (ITSs) and adaptive educational hypermedia [85].

## IV.2 Implementation

### IV.2.1 Context and Participants

509 The research context for this study was an introductory  
 510 object-oriented programming (OOP) course offered to un-  
 511 dergraduate students at NTNU. The course content was  
 512 delivered online (e.g., reading materials, assignments, ex-  
 513 amples) and once a week in a classroom setting (e.g.,  
 514 lectures and labs). During the course (which lasted for 3  
 515 months) the students were required to submit ten individual  
 516 assignments and undertake a final mandatory exam. The  
 517 grade students get at the end of the course is based only  
 518 on the final exam. The instructor used the university LMS  
 519 to distribute the relevant course materials and Eclipse in-  
 520 tegrated development environment (IDE) for the submission  
 521 of the individual assignments. In addition, the instructor  
 522 introduced ProTuS and MasteryGrids, as non-mandatory  
 523 learning systems, that students could use to practice and  
 524 learn Java.

525 The sample was comprised of 153 participants, freshman  
 526 CS majors, who were in their second semester. All par-  
 527 ticipants had already taken an introductory programming  
 528 course in Python in their first semester; thus, it was assumed  
 529 that they have already mastered a basic knowledge in pro-  
 530 cedural programming. The study focused on a set of online  
 531 activities and participants' interaction with the educational  
 532 content. ProTuS has been used as a portal for seamless in-  
 533 tegration of content from different content providers, while  
 534 both ProTuS and VENT have been used to access, record,  
 535 and collect activity data. The data were collected over the  
 536 academic year 2018-2019 from logs of the three e-learning  
 537 systems: ProTuS, MasteryGrids (PCLab, PCRS), and VLASP  
 538 (see Table 2).

### IV.2.2 Study design and data collection.

539 Before the start of the study, the participants were intro-  
 540 duced to the NTNU policy for ethical and data privacy  
 541 issues, as well as with the purpose of the study and the e-  
 542 learning systems that they could interact with. The learning  
 543 content encompassed four types of activities that support  
 544 individual work aligned with self-regulated learning prac-  
 545 tices [86]. Participants that used the system signed up with  
 546

TABLE 2  
Overall Data Collected

Academic year 2018-2019	Value
Overall number of users	389
Number of active users	153
Number of user sessions	2.727
Number of student activities (content visits, submissions, etc)	906
Data sources	ProTuS, PCLab, PCRS and VLASP

549 their university email address; however, in the system they  
 550 got an ID number (e.g., StudentID001) that has been linked  
 551 across the three systems. The three systems provided five  
 552 types of learning content, which are briefly described in the  
 553 following:

- 554 1) **Explanations (ProTuS).** ProTuS contains reading  
 555 content (i.e., tutorials) on 15 topics that are aligned  
 556 with the curriculum presented in the course. These  
 557 learning materials help students to master concepts  
 558 in OOP (Java language) based on their existing  
 559 knowledge in procedural programming (Python).
- 560 2) **Examples (MasteryGrids-PCLab).** For each topic  
 561 learners can start with a worked-out example from  
 562 Program Construction EXamples (PCEX) set [87],  
 563 which explains why certain programming con-  
 564 structs are used in the code. Explanations are avail-  
 565 able for almost all lines of code in the example,  
 566 and are hidden until a learner clicks on the lines  
 567 of interest.
- 568 3) **Challenges (MasteryGrids-PCLab).** Following the  
 569 pedagogical reasoning that examples are more effec-  
 570 tive when a learner solves a problem immediately  
 571 after the example [88], we presented a challenge  
 572 after each example. Each challenge shows a problem  
 573 similar to the one presented in the example, and  
 574 blank lines that need to be filled in by dragging and  
 575 dropping the pieces of code to the blank fields [89].
- 576 4) **Coding exercises (MasteryGrids-PCRS).** The Pro-  
 577 gramming Course Resource System [75], whose con-  
 578 tent server resides at the University of Toronto,  
 579 provides coding exercise with a problem description  
 580 and a baseline code. When learners submit their  
 581 code, the code is tested against a set of unit tests de-  
 582 veloped for that particular problem, and the learner  
 583 receives an immediate feedback on whether the tests  
 584 were passed or not.
- 585 5) **Course assignments (VLASP).** The ten individual  
 586 assignments learners solve in Eclipse IDE, as they  
 587 are able to test the code against a set of unit tests  
 588 developed by the instructor. Learners' Eclipse instal-  
 589 lation has been extended with a plug-in that collects  
 590 data from the learners' solutions.

591 All three systems keep a track of every click and store  
 592 data as logs with time stamps in the learning record store.  
 593 ProTuS collects data about *learners' actions* in the system.  
 594 The collected data for our study included the number of  
 595 actions in the system, the time spend in each session, what  
 596 topic a learner selected, and the level of difficulty of the  
 597 coding exercise. According to the level of difficulty, the cod-  
 598 ing exercises (PCRS) have been grouped in ProTuS in five

599 categories (e.g., novice, skillful, confident, proficient, and  
 600 expert). MasteryGrids collects *progress data* from learners' 601 interactions with the learning content. The generated data  
 602 included clicks on lines of explanations in the examples,  
 603 attempts to solve a challenge, coding exercises solved in the  
 604 first, second, or third attempt, distinct challenges seen, etc.  
 605 The challenges and the coding exercises could be attempted  
 606 multiple times, or until the learner is satisfied with his  
 607 or her performance. Finally, VLASP collects *activity data*  
 608 related to a programming assignment. In our study we  
 609 collected the number of submitted assignments, the number  
 610 of incorrect and incomplete submissions, and the number of  
 611 assignments not submitted at all. A full list of the generated  
 612 variables is presented in the Appendix A.

#### IV.2.3 Data processing

613 The data were extracted from the learning record store and  
 614 as .cvs file placed in R Studio, to extract features from  
 615 all three systems (ProTuS, MasteryGrids, and VLASP). In  
 616 total, 142 features were extracted from the three e-learning  
 617 systems, of which after removing the columns that had  
 618 SD = 0, the final data set included 55 features. Appendix  
 619 A includes a table with all 55 features plus explanation  
 620 for each. The data collection and the respective e-learning  
 621 systems abide by the European data privacy regulations that  
 622 allow data to be collected and anonymized before use.

#### IV.2.4 Variables

624 To answer the research questions, we selected learners' 625 performance to be our dependent variable. Learner perfor-  
 626 mance was computed from the score participants achieved  
 627 on the individual assignments they submitted, transformed  
 628 into a grade. The performance from the individual  
 629 assignments summarizes participants' development over the  
 630 course, considering the applied learning design, and there-  
 631 fore, it is timely, available during the course run-time, and  
 632 more granular and representative. The learner performance  
 633 includes values between 0 and 1000, which was discretized  
 634 into six levels (i.e., A to F) with the help of the instructor,  
 635 to resemble a grade that a learner would receive if the  
 636 instructor assigned grades based on the assignments (i.e.,  
 637 formative assessment perspective) rather than based on the  
 638 final exam (i.e., summative assessment). The rest of the  
 639 54 features that were all extracted using the R language,  
 640 were considered to be the experimental/predictor variables.  
 641 Appendix A provides a summary of all features.

#### IV.2.5 Data analysis

642 To demonstrate proof-of-concept for the feasibility of the  
 643 implemented cross-platform architecture, we applied both,  
 644

646 inference and prediction. A statistical model will help us  
 647 infer the relationship between the data variables to a degree  
 648 of statistical significance, and use prediction to identify the  
 649 best course of action. Our aim was to explore how can  
 650 we incorporate hybrid approaches that combine statistical  
 651 methods with machine learning techniques in education,  
 652 particularly when combining analytics across systems and  
 653 data sources.

654 First, to get an initial understanding of the measures, a  
 655 descriptive statistic was calculated and the Shapiro-Wilk test  
 656 was used to check for data normality. The Shapiro-Wilk test  
 657 showed that the data did not have a normal distribution (p  
 658 values were significant) but a highly skewed nature. How-  
 659 ever, because linear regression does not assume normality  
 660 for either the predictor or the outcome variable, the lack of  
 661 normal distribution of the collected data was not an obstacle  
 662 to perform a linear regression (for more information please  
 663 look at the GaussMarkov theorem) [90]. We also checked for  
 664 other assumptions important for linear regression to ensure  
 665 that the inferences are appropriate: 1) *multicollinearity*—there  
 666 was no perfect linear regression between two or more pre-  
 667 dictor variables. We calculated the variance inflation factor  
 668 (VIF) and following the rule of thumb, in order to consider  
 669 problems with collinearity the VIF value should exceed 10,  
 670 which was not the case in our data [91]; 2) *homoscedasticity*  
 671 or homogeneity of variance, refers to the constant variance  
 672 of the residuals [92]. We checked for homoscedasticity by  
 673 plotting the data and exploring the residuals vs fitted and  
 674 scale-location (or Spread-Location) diagnostic plots, as well  
 675 as running the Levene's test (p values were not significant)  
 676 [93]; 3) *normally distributed errors*—we checked if the residuals  
 677 in the model are normally distributed by generating the  
 678 quantile-quantile (Q-Q) plot. The Q-Q plot from our data  
 679 shows that each observation roughly falls on the straight  
 680 line, indicating that the residuals are roughly normally  
 681 distributed.

682 Second, we looked into several ways how variables can  
 683 be entered into a model, but because we were conducting  
 684 an exploratory study on all generated variables that does  
 685 not have  $SD=0$ , we decided to go with the stepwise method  
 686 (backward direction), which has a lower risk of making Type  
 687 II error (i.e., missing a predictor that does in fact predict  
 688 the outcome) due to suppressor effects [91]. The backward  
 689 method starts by placing all predictors in a model and based  
 690 on Akaike Information Criterion (AIC) the model removes  
 691 predictors that cause AIC value to increase. The stepwise  
 692 methods is usually used for exploratory model building  
 693 and when researchers do not know which predictors can  
 694 create the best model [94]. Thus, because there was no  
 695 previous research that we could consider and built upon  
 696 with respect to reported significant variables, we decided to  
 697 build the models on a purely mathematical criterion. Due  
 698 to the selection of the stepwise method, we performed a  
 699 10-fold cross validation.

700 Third, we evaluated the performance of each of the se-  
 701 lected features that are shown in Appendix A in predicting  
 702 learner performance (i.e., student grades), using Random  
 703 Forest [95]. Random Forest (RF) as a decision tree-based  
 704 algorithm, is suitable for large numbers of features that are  
 705 strongly correlated [96]. Moreover, RF offers easy extraction  
 706 of feature importance, and has been found to be a top

707 performing algorithm in a large comparative study [97]. To  
 708 build a predictive model (not a representative) when dealing  
 709 with an imbalance dataset (which is a common problem in  
 710 the education field) in a multiclass classification problem,  
 711 we first performed a stratified sampling with respect to the  
 712 majority class, i.e. grade B, used to control the sampling  
 713 process. This step was necessary to avoid creating a train  
 714 and test set with totally different data distributions. Then,  
 715 we divided the dataset into training (70% of the students)  
 716 and testing (30% of the students) sets, and applied a hybrid  
 717 re-sampling technique (i.e. SMOTE) to the training set [98].  
 718 Using SMOTE we down-sampled the majority classes and  
 719 synthesize new data points in the minority classes, using k-  
 720 nearest neighbours for the new data [98]. This was an impor-  
 721 tant step in the analysis, because RF algorithm is sensitive to  
 722 the proportions of the classes, tending to favor the majority  
 723 class. Finally, to remove the selection bias in the training set,  
 724 we used 70% of the data to train the model using a 10-fold  
 725 cross-validation. This reduces the variability and presents  
 726 more accurate estimates of learners' performance.

727 At the end, we used RF to measure the importance of  
 728 the individual features for learner performance (i.e., student  
 729 grades). While importance of individual classification  
 730 features might be calculated in many different ways [99], we  
 731 used Mean Decrease Gini (MDG) which is based on the  
 732 reduction in Gini impurity measure. Gini impurity measures  
 733 how often a randomly chosen data point from the data set  
 734 will be incorrectly labeled, which is essential for correctly  
 735 classifying new data points. Classification accuracy (ACC),  
 736 which is the ratio of the total number of correct predictions  
 737 and the total number of predictions, is a reliable measure but  
 738 it is not sufficient to evaluate machine learning classification  
 739 algorithms [100]. Hence, we employed precision, recall,  
 740 f-measure, and Cohen's kappa, as additional measures to  
 741 evaluate the robustness of the classifier. Precision is the  
 742 ratio between the true positives and all the positives (true  
 743 positives + false positives), and gives us the measure of  
 744 relevant data points; while recall shows the classifier's poten-  
 745 tial to find all the positive outcomes. Thus, we calculate  
 746 the average precision and recall, weighted by the number  
 747 of true instances for each label, to account for the label  
 748 imbalance. F-score aggregates precision and recall under  
 749 the concept of harmonic mean that summarizes the model  
 750 performance. Finally, Cohen's k shows how the classifier is  
 751 performing over the performance of a classifier that guesses  
 752 at random with respect to the frequency of each class.

## V RESULTS

753 Table 3 shows the results from the stepwise multiple regres-  
 754 sion (backward direction) in building the exploratory model  
 755 based on a purely mathematical criterion. In fact, we were  
 756 interested in identifying variables that have a scientifically  
 757 meaningful and statistically significant relationship with the  
 758 learner performance (i.e., the number of points 0-1000). This  
 759 step was required to explore if architecting analytics across  
 760 multiple systems can improve the explanation power over  
 761 the individual systems, and because there is no theoretical  
 762 grounding that can be used as a starting point for specific  
 763 predictors (i.e., features/variables) to create the best model.

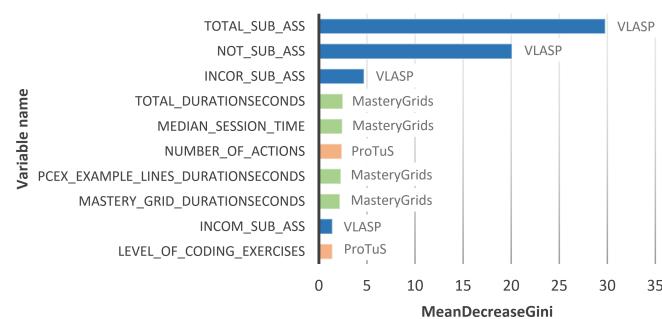


Fig. 3. Top ten feature importance - Random forest.

765 The model combination selected with  $AIC = 1088$  is significant  $(F(8,168) = 4600, p < 0.001)$  and explains 87% of the  
 766 variance in learners' performance.  
 767

768 The ProTuS model is not significant  $(F(3,173) = 0.688,$   
 769  $p = 0.561)$  and on its own explains only 12% of the variance,  
 770 while MasteryGrids model although not significant  
 771  $(F(46,130) = 1.22, p = 0.193)$ , accounts for 30%. The VLASP  
 772 model  $(F(3,173) = 1200, p < 0.001)$  is significant and with  
 773 total, incomplete, and incorrect submissions as predictors,  
 774 accounts for 80% of the variance in learners' performance.

775 Table 4 presents an overview of the main results, listing  
 776 the classifier's accuracy, Cohen's  $k$ , the average precision, re-  
 777 call, and the f-measure for RF. The features coming from the  
 778 separate systems (e.g., ProTuS, MasteryGrids, and VLASP)  
 779 have lower accuracy (and Cohen's  $k$ ) than the combined  
 780 features from the integrated system. The best classification  
 781 accuracy of 0.81 (95% CI[0.67, 0.91]) and Cohen's  $k$  0.79  
 782 comes from the classifier which considers the top 10 features  
 783 (i.e., RandomForest Top 10 features), and was obtained with  
 784 mtry value of 6. Mtry is a parameter of RF which shows  
 785 the number of variables randomly sampled as candidates at  
 786 each split. In our case, the mtry value of 6 means that each  
 787 decision tree took into account only 6 features out of the 54  
 788 features.

789 At the end, we performed a feature importance analysis  
 790 and we present the top ten most significant variables shown  
 791 in Fig. 3. We also present the top ten variables across the six  
 792 classes (i.e., grades A to F) as shown in Table 5.

## 793 VI DISCUSSION

### 794 VI.1 Interpretation of the Results with Respect to the 795 Research Questions

796 Since the nature of this study was exploratory, in which we  
 797 used both inference and prediction, we provide insights into  
 798 association relationships and not causality [100]. Considering  
 799 the results, we outline the positive findings from the  
 800 analyses, as a reinforcement toward the positive findings  
 801 reported by [82].

#### 802 VI.1.1 Insights derived from cross-platform learning analyt- 803 ics

804 With respect to the first research question, the regression  
 805 results presented in Table 3 show that learning analytics de-  
 806 rived from the separate systems, ProTuS and MasteryGrids,  
 807 are not significant and they explain less than 30% of the

808 variance, while learning analytics generated from VLASP  
 809 explain 80%. The analytics from VLASP are directly related  
 810 to student assessment outcomes, and as such, support the  
 811 previous research findings [7]–[9], which reported that stu-  
 812 dents often focus on assessment and technologies that can  
 813 maximize their chance to succeed (e.g., get a high grade  
 814 at the end of a course). Moreover, assessment was a major  
 815 focus of the learning design in the OOP course in which  
 816 our study was implemented. From past research [11], we  
 817 know that learning design and the instructional conditions  
 818 strongly affect what technologies and tools students use.  
 819 Thus, students' decision to focus on VLASP more than on  
 820 ProTuS or MasteryGrids, was based on their perception that  
 821 can help them to maximize their chances to succeed. In fact,  
 822 students were required to achieve more than 750 points on  
 823 the individual assignments to qualify for the final exam,  
 824 although this score from the individual assignments was  
 825 not counted in the final grade.

826 Nonetheless, we observed that combining data collected  
 827 across several distributed learning systems accounted for  
 828 an additional increase (i.e., 7%) in the explanation of the  
 829 variance of learner performance. The 7% increase is coming  
 830 from the following analytics: *the level of complexity of a chosen  
 831 coding exercise, the time students spend navigating in mastery  
 832 grids to monitor and reflect on their progress, the successful  
 833 attempts on challenges, and the distinct challenges successfully  
 834 solved*. Although the additional increase in the explanation  
 835 of the variance is not very large and is with an overall effect  
 836 of 9%, it is still a significant step (e.g., demonstrating proof  
 837 of concept) toward building learner models that can explain  
 838 higher portions of variation in the outcome (e.g. student's  
 839 grade performance) by combining analytics across different  
 840 platforms. Some of these analytics (i.e., *correct attempts to  
 841 problems, distinct problems attempted correctly, and time in  
 842 mastery grids navigation*) have also been found significant for  
 843 student engagement, usage, and attitude in the open social  
 844 student modeling (OSSM) compared to the open student  
 845 modeling (OSM) interface in technology-based learning [89].  
 846 The authors [89, p.459] have reported these three analyt-  
 847 ics/features and additional eleven, as "very attractive for  
 848 contexts where motivation and retention are critical, such as  
 849 modern MOOCs."

850 After describing the data using a statistical framework,  
 851 and characterizing the relationship between the variables  
 852 and the learner performance to a degree of statistical sig-  
 853 nificance, we utilized machine learning to build different  
 854 models with various features to predict a label, i.e., i.e.  
 855 student grade. RF used the associations between the pre-  
 856 dictors and the learner performance to validate the benefits  
 857 of cross-platform learning analytics in generating accurate  
 858 predictions for future outcomes. In addition, we also in-  
 859 vestigated if cross-platform learning analytics can build pre-  
 860 dictive models that carry an opportunity to reveal significant  
 861 elements for teaching practice.

862 The results reported in Table 4 are aligned with the  
 863 findings from the regression analysis, i.e., harnessing cross-  
 864 platform learning analytics can improve the classification  
 865 accuracy in predicting learner performance (i.e., student  
 866 grade). The baseline performance for the proposed learning  
 867 ecosystem that differentiates between 6 different classes (i.e.,  
 868 students' grades) is 1/6 or 16.7%. We argue that when

TABLE 3  
Stepwise Multiple Linear Regression Combining Features from all 3 Learning Systems

Model	adj. $R^2$	B	SE B	$\beta$	p
<b>Model Zero</b>					
(Intercept)		648	23		0.000
<b>Model Combination</b>	0.871				
(Intercept)		-4.10	5.31		0.446
level_of_coding_exercises		0.02	0.01	0.65	0.025
mastery_grids_durationseconds		1.21	0.71	0.42	0.033
pcex_ch_attempts_success		-9.66	6.34	-0.16	0.095
pcex_ch_success		9.95	6.56	0.16	0.090
total_sub_ass		99.95	0.56	0.99	0.000
incom_sub_ass		-33.44	2.50	-0.17	0.000
incor_sub_ass		-90.07	4.0	-0.32	0.000
not_sub_ass		1.19	0.85	0.41	0.095
<b>Model ProTuS</b>	0.117				
<b>Model MasteryGrids</b>	0.301				
<b>Model VLASP</b>	0.801				

TABLE 4  
Random Forest Classifier. Combo:Combining Features From all Three Learning Systems. Top10: Using Only the Ten Best Features. ProTuS, MasteryGrids, VLASP: Using Features Solely from one System

Classifier	ACC	Kappa	Precision	Recall	F-measure
RandomForest combo	0.79	0.69	0.90	0.77	0.87
RandomForest Top10	0.81	0.79	0.92	0.81	0.89
RandomForest ProTuS	0.23	0.01	0.14	0.23	0.37
RandomForest MasteryGrids	0.42	0.05	0.79	0.42	0.59
RandomForest VLASP	0.70	0.65	0.88	0.77	0.78

TABLE 5  
Feature Importance Across Classes

Variable	A	B	C	D	E	F
total_sub_ass	53.88	100	43.97	51	33	87.47
not_sub_ass	47.56	78.22	44.62	41.55	34.11	56.26
incor_sub_ass	19.58	60.73	32.38	36	5.21	25.51
total_durationseconds	7.15	20	10	8.50	12.30	15.25
median_session_time	9.51	23.10	14	20	4.31	10.12
number_of_actions	4.48	9	15	10	8.76	3.51
pcex_example_lines_durationsec.	8	30	11.43	15.67	4.22	5.72
mastery_grid_durationsec.	6.12	9	24.62	3.18	12.55	10
level_of_coding_exercises	6.10	2.10	2.50	3.46	7.20	20
incom_sub_ass	7.14	9.75	5	17.30	6.31	34

reporting machine learning results, a baseline performance should also be reported, due to the importance of discussing performance in relation to the complexity of the machine learning task [100]. Thus, the results from the analysis show that our learning ecosystem achieved accuracy of 79% (RandomForest combo) and 81% (RandomForest top 10), exhibiting an improvement in the baseline by a factor of 3.73 and 3.85 respectively. In other words, the RandomForest Top10 performance exhibits a 5-fold increase over the baseline. Also, looking at the f-measure for RandomForest combo and RandomForest Top10 classifiers, one can notice that these classifiers have demonstrated better robustness (do not miss a significant number of instances) and precision (how many instances it classifies correctly) measures than the rest of the classifiers.

In a study presented by [101], prediction models have shown that combination of mastery data (i.e., mastery score) and use intensity data (i.e., number of attempts, time on

task) from e-tutorial systems that students used to practice homework exercises, constitute a good second best information source (after assessment data) for predicting performance. Their findings [101] strongly support the integrative approach to learning analytics as advocated by [102]. Moreover, our findings also align with these previous findings, that combining analytics across systems in distributed learning environments can provide insights into what is happening in each and across the different systems, and thus, be used to predict performance more accurately.

By harnessing cross-platform learning analytics, our predictive models also disclosed the potential for building future models that can reveal significant elements for teaching practice, which can be utilized to further refine the design of learning activities and instructional methods. In fact, the advantages from analytics generated across various e-learning systems lie in the support that educators will have, to shape teaching and learning with data that is timely and available

905 during the learning process. We posit that in future, such  
 906 insights can assist educators to examine the effectiveness of  
 907 their learning designs and assessment practices in relation to  
 908 serving the intended educational objectives and pedagogical  
 909 intent, and maximize the learning opportunities in digital  
 910 education.

911 *VI.1.2 Cross-platform learning analytics to support learning  
 912 design*

913 Compared to previous studies on predictive modeling that  
 914 have investigated academic success [20], [39], [103], our  
 915 intention was not to build a high performing prediction  
 916 models that outperform other machine learning models,  
 917 but to explore how predictive models can be constituted  
 918 with practical value for educators, to inform teaching and  
 919 learning practices as a “diagnostic” tool, pertinent to the  
 920 optimization of various technologies.

921 Based on the positive findings reported in Table 3 and  
 922 in particular Table 4, we posit that although the best pre-  
 923 dictor for performance is performance itself, there are other  
 924 features also relevant for learner performance (but not all of  
 925 the features) that can be extracted from behavior log data.  
 926 Our findings are supported from previous research [104]–  
 927 [108], which demonstrated that not all analytics that can be  
 928 collected in a learning environment are equally relevant for  
 929 learning, nor the same learning analytics are relevant for  
 930 every student. If we look at Fig. 3, which displays the top  
 931 ten features by importance (generated with RandomForest  
 932 Top10 features classifier), we can observe that in addition to  
 933 the activity data generated from the VLASP system (directly  
 934 related to the assessment outcomes), other analytics at dif-  
 935 ferent granular levels are also significant for learner per-  
 936 formance. Therefore, in this paper we present the value of more  
 937 granular data to monitor and assess learner progress, which  
 938 can be utilized to develop interpretable predictive models  
 939 based on cross-platform learning analytics. Such models  
 940 can reveal significant elements from the learning designs  
 941 for understanding behavior and progress in distributed  
 942 settings, in addition to data generated from summative  
 943 assessment or LMS, and instructor’s tacit knowledge, that  
 944 can be harnessed to identify best course of action in making  
 945 reliable and informed decisions.

946 Table 5 shows the top ten features across the six classes  
 947 (i.e., students’ grades). For example, if we look at grade  
 948 B, we can notice that although the assignment submissions  
 949 are very important (we explained the role of VLASP in the  
 950 learning design), the time spend on practicing learning tasks  
 951 and the time spend on reading examples before practicing  
 952 exercises, can also be important indicators for a learner per-  
 953 formance. Examples with explanations for the code are com-  
 954 monly used learning resources in learning programming  
 955 that help students grasp various programming structures  
 956 and concepts [109]. To optimize the support for learning  
 957 from examples, instructors can benefit from insights derived  
 958 from interpretable models as presented in this study, to  
 959 guide students to access the right example at the right time  
 960 [87], [110]. This is an important decision that instructors  
 961 can make, because past studies demonstrated that the effect  
 962 from worked examples is stronger in the early stages of  
 963 learning, and declines gradually as students’ knowledge  
 964 grows [111].

965 For grade C, another important indicator can be the time  
 966 a learner spend navigating in mastery grids to monitor  
 967 and reflect on their progress. In other words, if learners  
 968 spend time monitoring and reflecting on their progress, such  
 969 information can assist the instructor to help those students  
 970 improve their self-regulation skills and the decisions they  
 971 make. From a self-regulated learning perspective, learners  
 972 are considered to be active participants in the learning pro-  
 973 cess, who construct their own meanings and goals, and can  
 974 potentially monitor and regulate certain aspects of their own  
 975 metacognition, motivation, and behavior, from the informa-  
 976 tion available to them [112]. Thus, more granular data from  
 977 interpretable models can assist instructors to work toward  
 978 development of personalized feedback to support students’  
 979 self-regulated learning skills, thereby helping their students  
 980 to become independent professionals, who can shape their  
 981 own learning. In addition, sharing learning analytics from  
 982 such interpretable models with students, can enable them  
 983 to understand the state of their knowledge and use this  
 984 knowledge to plan their learning [113].

985 Finally, the last example for grade F, shows that other  
 986 important features for learner performance can be the level  
 987 of complexity students choose in the coding exercises and  
 988 the submission of incomplete assignments. One explanation  
 989 can be the potential association between these two, sug-  
 990 gesting that learners who get grade F, might have trouble  
 991 selecting assignments that match their current knowledge,  
 992 so they failed to learn meaningfully. Students’ lack of knowl-  
 993 edge (and potential development of misconceptions) is later  
 994 demonstrated in the submission of incomplete assignments.  
 995 In fact, these learners need an intervention through proper  
 996 scaffolding, to guide them gradually to master skills and  
 997 learn concepts, by aligning the complexity of the assign-  
 998 ments with their current knowledge proficiency. This is an  
 999 action, much in line with the existing research in adaptive  
 1000 learning and intelligent tutoring systems [114].

1001 In this study, the insights generated from cross-platform  
 1002 analytics through feature importance, depict a different  
 1003 approach where more granular data can offer additional  
 1004 information (not easily observable in digital settings) on  
 1005 top of the information educators have from the LMS in  
 1006 use, their tacit knowledge, or the summative and formative  
 1007 assessment data. Such additional information still does not  
 1008 reveal the whole picture how students learn, but discloses  
 1009 significant elements for teaching and learning practice about  
 1010 how students use the opportunities as given in the learning  
 1011 design, which can assist educators to further refine the  
 1012 design of learning activities and instructional methods in  
 1013 digital education.

1014 In sum, we argue that our approach can overcome some  
 1015 of the ongoing issues (e.g., one-sided learning analytics mea-  
 1016 sures, strong focus on summative assessment) in learning  
 1017 analytics by collecting, integrating, and harmonizing data  
 1018 from several learning systems and at different granular  
 1019 levels. This approach can generate data that represent a  
 1020 larger proportion of the learning process and the activities  
 1021 students engage with. Thus, educators can make effective  
 1022 and meaningful refinements in the learning designs that  
 1023 can encourage, enable, and advance learning. At last, al-  
 1024 though the technological advancements increased the inter-  
 1025 est for performance-based, formative assessment [115] and

1026 e-learning systems that can effectively support that [116],  
 1027 the biggest challenge is that there are many aspects (e.g.,  
 1028 reliability, validity) of assessment in online settings that are  
 1029 yet to be comprehend in relation to serving the intended  
 1030 learning purposes [117].

## 1031 VI.2 Theoretical and Practical Implications

1032 The presented study provides useful insights for learning  
 1033 technology researchers, designers, and developers, by intro-  
 1034 ducing the concept of cross-platform analytics architecture  
 1035 that could measure the effectiveness and fine-tune learning  
 1036 designs, to maximize learning opportunities in distributed  
 1037 settings. The findings support the importance of harnessing  
 1038 data across various learning systems by emphasizing the  
 1039 potential of leaving the exclusive focus on single source  
 1040 data. By quantifying the usefulness of cross-platform learn-  
 1041 ing analytics, we would like to invite learning technology  
 1042 designers to focus on the development of valuable intercon-  
 1043 nected functionalities, affordances, and resources.

1044 One of the most important implications of this paper  
 1045 is related to how learning technology and user experience  
 1046 researchers and practitioners can employ analytics across  
 1047 platforms and build cross-platform methodologies to make  
 1048 sense of the requirements that stem from different learning  
 1049 designs, as well as take design decisions for various learner  
 1050 groups. The 21st-century learning systems are expected to  
 1051 become more interconnected and personalized (e.g., Khan  
 1052 Academy, Udacity), and incorporate smart and adaptive  
 1053 behavior (e.g., Adaptemy, Dreambox, SmartSparrow). How-  
 1054 ever, there is a lack of the state-of-the-art empirical ap-  
 1055 proaches that can combine and identify what analytics can  
 1056 measure the effectiveness of learning designs, and how vari-  
 1057 ous stakeholders can benefit from those combinations. Tak-  
 1058 ing a cross-platform analytics approach provides a unique  
 1059 opportunity to enrich the contemporary capacities of the  
 1060 current learning systems, by using statistical and machine  
 1061 learning techniques as a “diagnostic” practice that educators  
 1062 can utilize it, to improve the quality of the instructional  
 1063 conditions. This will allow contemporary learning ecosys-  
 1064 tems to leverage the capacities of their learning analytics  
 1065 and maximize their innovation potential.

1066 On the practical side, we managed to propose and im-  
 1067 plement in practice a cross-platform architecture that inte-  
 1068 grates and interconnects analytics capabilities, and enhances  
 1069 the present analytics capacities of ProTuS. As elaborated  
 1070 in the related work, currently there are many conceptual  
 1071 frameworks and software architectures that emphasize the  
 1072 need for a cross-platform methodologies; however, none at  
 1073 present completely solves the problem of collecting, inte-  
 1074 grating, and harmonizing learning-related behavioral log  
 1075 data from several distributed environments. The proposed  
 1076 architecture presents the minimum technical architecture  
 1077 requirements and provides solution for data format inter-  
 1078 operability and integration issues.

1079 Despite the limitations of this study, we obtained pos-  
 1080 itive and encouraging results, that developing cross-platform  
 1081 architecture and combining data across several learning  
 1082 systems can advance the state-of-the-art in developing an  
 1083 ecosystems of “dynamic, interconnected, and ever-evolving  
 1084 community of learners, instructors, tools, and content” [4],

1085 as well as toward predictive models that can provoke reflec-  
 1086 tion and action among learners and instructors. The humble  
 1087 analysis approach braces the proof-of-concept in furthering  
 1088 the understanding of how cross-platform analytics can add  
 1089 value to enrich the contemporary learner models and lever-  
 1090 age the capacities of their analytics. Finally, one of the most  
 1091 significant contributions of this study is the demonstrated  
 1092 feasibility of the defined concept, where the learner model  
 1093 is gradually built based on integration of data from three  
 1094 e-learning systems.

## 1095 VI.3 Limitations and Future Work

1096 One of the limitations in our analysis is related to the size  
 1097 of the data set. Although we have 153 students using the  
 1098 integrated system, not all of them are using the system  
 1099 frequently. Another limitation is the lack of comprehensive  
 1100 set of feature extraction, especially the features that can  
 1101 be extracted from the main tasks or activities related to a  
 1102 programming exercises. For example, how many times a  
 1103 student has run an individual test, how many times the  
 1104 code has been compiled, the number of errors and warnings  
 1105 resulting from the compiler’s analysis of the code, etc. These  
 1106 features can lead to improvements in designing program-  
 1107 ming instructions, assignments, and scaffolds, and reveal  
 1108 directions for future research on curriculum design and  
 1109 analytics in computing education. Third, the interpretation  
 1110 of the importance and significance of the results for learning  
 1111 design for researchers and practitioners (e.g., instructors) is  
 1112 limited and difficult to estimate, because it is mainly based  
 1113 on our understanding and knowledge in learning design  
 1114 and learning analytics. Therefore, in the future these find-  
 1115 ings need to be investigated with instructors who would uti-  
 1116 lize the learning ecosystem in their course. Finally, because  
 1117 all of our participants are coming from a single university  
 1118 with a particular pedagogical and instructional approach,  
 1119 the results from the classification algorithms might have  
 1120 effect on the generalizability of our findings. Thus, in our  
 1121 future work, we are planning to extend the content by  
 1122 developing a programming course for Python. We also plan  
 1123 to implement the integrated ecosystem in collaboration with  
 1124 a other universities that offer introduction courses in Java  
 1125 and Python, to increase the generalization power of our  
 1126 analyses, to further validate our findings, and to account for  
 1127 other important features that might have been overlooked  
 1128 in this analysis.

## 1129 VII CONCLUSION

1130 To demonstrate and validate real-life examples of how and  
 1131 when learning is taking place, educators and researchers  
 1132 need to embrace the complexity of the learning process and  
 1133 its distributed nature across various learning settings and  
 1134 contexts. In that regard, we tried to capture and explore  
 1135 authentic learner-generated behavior log data coming from  
 1136 three different e-learning systems (each system resides on  
 1137 different server at different university). Our objective was  
 1138 to integrate analytics across e-learning systems with the aim  
 1139 to explore and understand how to create and measure the  
 1140 effectiveness of learning designs that can maximize learning  
 1141 opportunities in distributed learning environments. Conse-  
 1142 quently, we proposed and implemented a cross-platform

1143 architecture for interactive courses and analytics support.  
 1144 While most of the previous work handles data from one  
 1145 source, this study aims to present a cross-platform archi-  
 1146 tecture for simple automatic integration and ease of data  
 1147 collection from four different data sources. To that end,  
 1148 this study takes a humble approach to analysis, comparing  
 1149 learning analytics metrics across three e-learning systems,  
 1150 using both inference and prediction. The proof-of-concept is  
 1151 envisioned to be the first step toward utilizing the potential  
 1152 of cross-platform learning analytics as an added value in  
 1153 (re)designing and evaluating learning and teaching activi-  
 1154 ties in distributed learning environments. This approach  
 1155 should aid users (e.g., educators, learners, instruction de-  
 1156 signers, and researchers) to engage in informed decision-  
 1157 making, considering relevant metrics that align with their  
 1158 goals and needs, and toward personalized and scaled feed-  
 1159 back practices in digital education.

## 1160 APPENDIX 1161 EXPLANATION OF ALL OF THE 55 FEATURES

1162 The Appendix A contains a full list of the generated vari-  
 1163 ables. It includes a table with all 55 features and an expla-  
 1164 nation for each.

## 1165 ACKNOWLEDGMENT

1166 This work was supported by the Research Council of Nor-  
 1167 way under the project FUTURE LEARNING (255129/H20).  
 1168 In addition, the authors are extremely grateful to the asso-  
 1169 ciate editor and the reviewers for their constructive com-  
 1170 ments and useful insights, which significantly improved the  
 1171 paper.

## 1172 REFERENCES

- 1173 [1] R. C. Clark and R. E. Mayer, *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons, 2016. doi: [:10.1002/9781119239086](https://doi.org/10.1002/9781119239086)
- 1174 [2] K. Mangaroska and M. N. Giannakos, "Learning analytics for  
 1175 learning design: A systematic literature review of analytics-  
 1176 driven design to enhance learning," *IEEE Trans. on Learn. Techn.*,  
 1177 2018.
- 1178 [3] C. Stephanidis, G. Salvendy, M. Antona, J. Y. Chen, J. Dong, V. G.  
 1179 Duffy, X. Fang, C. Fidopiastis, G. Fragomeni, L. P. Fu *et al.*, "Seven  
 1180 hci grand challenges," *Int. Journal of Human-Computer Interaction*,  
 1181 vol. 35, no. 14, pp. 1229–1269, 2019.
- 1182 [4] M. Brown, J. Dehoney, and N. Millichap, "The next generation  
 1183 digital learning environment," *A Report on Research. ELI Paper*,  
 1184 *Louisville, CO: Educause April*,
- 1185 [5] D. Suthers and D. Rosen, "A unified framework for multi-  
 1186 level analysis of distributed learning," in *Proc. 1st Int. Conf.*  
 1187 *on Learn. Analytics & Knowledge*. ACM, 2011, pp. 64–74. doi:  
 1188 [10.1145/2090116.2090124](https://doi.org/10.1145/2090116.2090124)
- 1189 [6] D. Gašević, N. Mirriahi, S. Dawson, and S. Joksimović, "Effects  
 1190 of instructional conditions and experience on the adoption of a  
 1191 learning tool," *Comp. in Human Behavior*, vol. 67, pp. 207–220,  
 1192 2017.
- 1193 [7] P. H. Winne, "A cognitive and metacognitive analysis of self-  
 1194 regulated learning: Faculty of education, simon fraser university,  
 1195 burnaby, canada," in *Handbook of self-regulation of Learn. and*  
 1196 *performance*. Routledge, 2011, pp. 29–46.
- 1197 [8] D. Boud, "Assessment and learning: contradictory or comple-  
 1198 mentary," *Assessment for Learn. in Higher Educ.*, pp. 35–48,
- 1199 [9] B. W. Wormald, S. Schoeman, A. Somasunderam, and M. Penn,  
 1200 "Assessment drives learning: an unavoidable truth?" *Anatomical*  
 1201 *Sciences Educ.*, vol. 2, no. 5, pp. 199–204, 2009.
- 1202 [10] W. Matcha, D. Gašević, N. A. Uzir, J. Jovanović, and A. Pardo,  
 1203 "Analytics of learning strategies: associations with academic per-  
 1204 formance and feedback," in *Proc. 9th Int. Conf. on Learn. Analytics*  
 1205 & *Knowledge*, 2019, pp. 461–470. doi: [10.1145/3303772.3303787](https://doi.org/10.1145/3303772.3303787)
- 1206 [11] B. Rienties and L. Toetenel, "The impact of learning design  
 1207 on student behaviour, satisfaction and performance: A cross-  
 1208 institutional comparison across 151 modules," *Comp. in Human*  
 1209 *Behavior*, vol. 60, pp. 333–341,
- 1210 [12] B. Rienties, L. Toetenel, and A. Bryan, "Scaling up learning  
 1211 design: impact of learning design activities on lms behavior and  
 1212 performance," in *Proc. Fifth Int. Conf. on Learn. Analytics & Knowl-*  
 1213 *edge*. ACM, 2015, pp. 315–319. doi: [10.1145/2723576.2723600](https://doi.org/10.1145/2723576.2723600)
- 1214 [13] G. Lust, J. Elen, and G. Clarebout, "Regulation of tool-use within  
 1215 a blended course: Student differences and performance effects,"  
 1216 *Comp. & Educ.*, vol. 60, no. 1, pp. 385–395, 2013.
- 1217 [14] K. Mangaroska, R. Martinez-Maldonado, B. Vesin, and D. Gašević,  
 1218 "Challenges and opportunities of multimodal data in human  
 1219 learning: The computer science students' perspective," *Journal of*  
 1220 *Comp. Assisted Learn.*, 2021.
- 1221 [15] K. Vassallo, L. Garg, V. Prakash, and K. Ramesh, "Contemporary  
 1222 technologies and methods for cross-platform application de-  
 1223 velopment," *Journal of Computational and Theoretical Nanoscience*,  
 1224 vol. 16, no. 9, pp. 3854–3859, 2019.
- 1225 [16] J. M. Dodero, E. J. González-Conejero, G. Gutiérrez-Herrera,  
 1226 S. Peinado, J. T. Tocino, and I. Ruiz-Rube, "Trade-off between  
 1227 interoperability and data collection performance when designing  
 1228 an architecture for learning analytics," *Future Generation Comp.*  
 1229 *Systems*, vol. 68, pp. 31–37, 2017.
- 1230 [17] S. K. Shankar, L. P. Prieto, M. J. Rodríguez-Triana, and A. Ruiz-  
 1231 Calleja, "A review of multimodal learning analytics architec-  
 1232 tures," in *2018 IEEE 18th Int. Conf. on Advanced Learn. Techn.*  
 1233 (*ICALT*). IEEE, 2018, pp. 212–214. doi: [10.1109/ICALT.2018.000057](https://doi.org/10.1109/ICALT.2018.000057)
- 1234 [18] L. P. Prieto, Y. Dimitriadis, C.-K. Looi *et al.*, "Orchestration in  
 1235 learning technology research: evaluation of a conceptual frame-  
 1236 work," *Research in Learning Technology*, 2015.
- 1237 [19] J. Roschelle, Y. Dimitriadis, and U. Hoppe, "Classroom orchestra-  
 1238 tion: synthesis," *Comp. & Educ.*, vol. 69, pp. 523–526, 2013.
- 1239 [20] D. Gašević, S. Dawson, T. Rogers, and D. Gašević, "Learning  
 1240 analytics should not promote one size fits all: The effects of  
 1241 instructional conditions in predicting academic success," *The*  
 1242 *Internet and Higher Educ.*, vol. 28, pp. 68–84, 2016.
- 1243 [21] R. Mazza and V. Dimitrova, "Coursevis: A graphical student  
 1244 monitoring tool for supporting instructors in web-based distance  
 1245 courses," *Int. Journal of Human-Computer Studies*, vol. 65, no. 2,  
 1246 pp. 125–139, 2007.
- 1247 [22] G. Siemens and P. Long, "Penetrating the fog: Analytics in  
 1248 learning and education," *EDUCAUSE review*, vol. 46, no. 5, p. 30,
- 1249 [23] P. Blikstein and M. Worsley, "Multimodal learning analytics  
 1250 and education data mining: using computational technologies  
 1251 to measure complex learning tasks," *Journal of Learn. Analytics*,  
 1252 vol. 3, no. 2, pp. 220–238, 2016.
- 1253 [24] X. Ochoa, C. Lang, and G. Siemens, "Multimodal learning ana-  
 1254 lytics," *The Handbook of Learn. Analytics*, vol. 1, pp. 129–141, 2017.
- 1255 [25] F. Domínguez and K. Chiluiza, "Towards a distributed frame-  
 1256 work to analyze multimodal data," in *Proc. Workshop Cross-LAK-*  
 1257 *held at LAK '16*, 2016, pp. 52–57.
- 1258 [26] R. Martinez-Maldonado, Y. Dimitriadis, A. Clayphan, J. A.  
 1259 Muñoz-Cristóbal, L. P. Prieto, M. J. Rodríguez-Triana, and J. Kay,  
 1260 "Integrating orchestration of ubiquitous and pervasive learning  
 1261 environments," in *Proc. 25th Australian Computer-Human Inter-  
 1262 action Conf.: Augmentation, Application, Innovation, Collaboration*.  
 1263 ACM, 2013, pp. 189–192. doi: [10.1145/2541016.2541076](https://doi.org/10.1145/2541016.2541076)
- 1264 [27] Z. Jeremić, J. Jovanović, and D. Gašević, "Personal learning  
 1265 environments on the social semantic web," *Semantic Web*, vol. 4,  
 1266 no. 1, pp. 23–51, 2013.
- 1267 [28] M. Worsley and R. Martinez-Maldonado, "Multimodal learning  
 1268 analytics past, present, and, potential futures," in *Proc. CEUR*  
 1269 *Workshop*, vol. 2163, 2018.
- 1270 [29] D. Di Mitri, J. Schneider, R. Klemke, M. Specht, and H. Drachsler,  
 1271 "Read between the lines: An annotation tool for multimodal data  
 1272 for learning," in *Proc. 9th Int. Conf. on Learn. Analytics & Knowledge*  
 1273 (*LAK'19*). ACM, 2019. doi: [10.1145/3303772.3303776](https://doi.org/10.1145/3303772.3303776)
- 1274 [30] J. L. Santos, K. Verbert, J. Klerkx, E. Duval, S. Charleer, and  
 1275 S. Ternier, "Tracking data in open learning environments," *Journal*  
 1276 *of Universal Comp. Science*, 2015.
- 1277 [31] J. A. Muñoz-Cristóbal, M. J. Rodríguez-Triana, V. Gallego-Lema,  
 1278 H. F. Arribas-Cubero, J. I. Asensio-Pérez, and A. Martinez-Monés,  
 1279 2020.

1281 "Toward the integration of monitoring in the orchestration of  
1282 across-spaces learning situations," in *CrossLAK*, 2016, pp. 15–21.

1283 [32] C. E. Kassko and R. Sarig, "System and method for performing  
1284 cross-platform big data analytics," U.S. Patent 10515386B2, Feb.  
1285 2020.

1286 [33] J. A. Quiane Ruiz, S. Kruse, Z. Kaoudi, and S. Chawla, "Apparatus,  
1287 system, and method for cross-platform data processing," U.S. Patent  
1288 20190347261A1.

1289 [34] E. C. Kassko and R. Raz, "System and method for providing  
1290 big data analytics on dynamically-changing data models," U.S.  
1291 Patent 20150213109A1.

1292 [35] G. Siemens and R. Baker, "Learning analytics and educational  
1293 data mining: towards communication and collaboration," in *Proc.  
1294 2nd Int. Conf. on Learn. Analytics & Knowledge*. ACM, 2012, pp.  
1295 252–254.

1296 [36] C. Brooks and C. Thompson, "Predictive modelling in teaching  
1297 and learning," *Handbook of Learn. analytics*, pp. 61–68, 2017.

1298 [37] K. E. Arnold and M. D. Pistilli, "Course signals at purdue: Using  
1299 learning analytics to increase student success," in *Proc. 2nd Int.  
1300 Conf. on Learn. Analytics & Knowledge*. ACM, 2012, pp. 267–270.  
1301 doi: [10.1145/2330601.2330666](https://doi.org/10.1145/2330601.2330666)

1302 [38] R. Barber and M. Sharkey, "Course correction: Using  
1303 analytics to predict course success," in *Proc. 2nd Int. Conf. on  
1304 Learn. Analytics & Knowledge*. ACM, 2012, pp. 259–262. doi:  
1305 [10.1145/2330601.2330664](https://doi.org/10.1145/2330601.2330664)

1306 [39] C. Romero, M.-I. López, J.-M. Luna, and S. Ventura, "Predicting  
1307 students' final performance from participation in on-line discus-  
1308 sion forums," *Comp. & Educ.*, vol. 68, pp. 458–472, 2013.

1309 [40] Á. F. Agudo-Peregrina, S. Iglesias-Pradas, M. Á. Conde-  
1310 González, and Á. Hernández-García, "Can we predict success  
1311 from log data in vles? classification of interactions for learning  
1312 analytics and their relation with performance in vle-supported  
1313 f2f and online learning," *Comp. in human behavior*, vol. 31, pp.  
1314 542–550, 2014.

1315 [41] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and  
1316 comparison of supervised data mining techniques for student  
1317 exam performance prediction," *Comp. & Educ.*, vol. 143, p. 103676,  
1318 2020.

1319 [42] A. P. Cavalcanti, A. Diego, R. F. Mello, K. Mangaroska, A. Nasci-  
1320 mento, F. Freitas, and D. Gašević, "How good is my feed-  
1321 back? a content analysis of written feedback," in *Proc. Tenth Int.  
1322 Conf. on Learn. Analytics & Knowledge*, 2020, pp. 428–437. doi:  
1323 [10.1145/3375462.3375477](https://doi.org/10.1145/3375462.3375477)

1324 [43] J. Heo, H. Lim, S. B. Yun, S. Ju, S. Park, and R. Lee, "Descriptive  
1325 and predictive modeling of student achievement, satisfaction,  
1326 and mental health for data-driven smart connected campus life  
1327 service," in *Proc. 9th Int. Conf. on Learn. Analytics & Knowledge*,  
1328 2019, pp. 531–538. doi: [10.1145/3303772.3303792](https://doi.org/10.1145/3303772.3303792)

1329 [44] G. Shmueli *et al.*, "To explain or to predict?" *Statistical science*,  
1330 vol. 25, no. 3, pp. 289–310, 2010.

1331 [45] E. Archer and P. Prinsloo, "Speaking the unspoken in learning  
1332 analytics: troubling the defaults," *Assessment & Evaluation in  
1333 Higher Educ.*, pp. 1–13, 2019.

1334 [46] Z. Tanes, K. E. Arnold, A. S. King, and M. A. Remnet, "Using  
1335 signals for appropriate feedback: Perceptions and practices,"  
1336 *Comp. & Educ.*, vol. 57, no. 4, pp. 2414–2422, 2011.

1337 [47] A. Essa and H. Ayad, "Student success system: risk analytics and  
1338 data visualization using ensembles of predictive models," in *Proc.  
1339 2nd Int. Conf. on Learn. Analytics & Knowledge*, 2012, pp. 158–161.  
1340 doi: [10.1145/2330601.2330641](https://doi.org/10.1145/2330601.2330641)

1341 [48] A. Pardo, N. Mirriahi, R. Martínez-Maldonado, J. Jovanovic,  
1342 S. Dawson, and D. Gašević, "Generating actionable predic-  
1343 tive models of academic performance," in *Proc. sixth Int.  
1344 Conf. on Learn. Analytics & Knowledge*, 2016, pp. 474–478. doi:  
1345 [10.1145/2883851.2883870](https://doi.org/10.1145/2883851.2883870)

1346 [49] A. F. Wise, "Designing pedagogical interventions to sup-  
1347 port student use of learning analytics," in *Proc. fourth Int.  
1348 Conf. on Learn. Analytics & Knowledge*, 2014, pp. 203–211. doi:  
1349 [10.1145/2567574.2567588](https://doi.org/10.1145/2567574.2567588)

1350 [50] A. Bianco, M. De Marsico, and M. Temperini, "Standards for e-  
1351 learning," *The TISIP Foundation*, Trondheim, Norway,

1352 [51] L. De-La-Fuente-Valentín, M. Pérez-Sanagustín, D. Hernández-  
1353 Leo, A. Pardo, J. Blat, and C. D. Kloos, "Technological support  
1354 for the enactment of collaborative scripted learning activities  
1355 across multiple spatial locations," *Future Generation Comp. Sys-*  
1356 tems

1357 [52] G. Siemens, D. Gasevic, C. Haythornthwaite, S. P. Dawson,  
1358 S. Shum, R. Ferguson, E. Duval, K. Verbert, R. Baker *et al.*, "Open  
1359 learning analytics: an integrated & modularized platform," *Canan-  
1360 dian Initiative for Distance Education Research*,

1361 [53] T. Rabelo, M. Lama, R. R. Amorim, and J. C. Vidal, "Smartlak: A  
1362 big data architecture for supporting learning analytics services,"  
1363 in *Frontiers in Educ. Conf. (FIE)*, 2015 IEEE. IEEE, 2015, pp. 1–5.  
1364 doi: [10.1109/FIE.2015.7344147](https://doi.org/10.1109/FIE.2015.7344147)

1365 [54] N. Sclater, A. Berg, and M. Webb, "Developing an open architec-  
1366 ture for learning analytics," *EUNIS Journal of Higher Educ.*,

1367 [55] A. T. Team, "Advanced distributed learning shareable content  
1368 object reference model." <http://www.adlnet.org/Scorm/scorm.cfm>

1369 [56] I. IMS Global Learning Consortium, "Ims learning  
1370 resource meta-data specification." <http://www.imsproject.org/metadata/index.html>

1371 [57] I. L. T. S. Committee, "Learning object meta-data." <http://ltsc.ieee.org/wg12/index.html>

1372 [58] C. Qu and W. Nejdl, "Towards interoperability and reusability  
1373 of learning resources: A scorm-conformant courseware for com-  
1374 puter science education," in *Proc. 2nd IEEE Int. Conf. on Advanced  
1375 Learn. Techn. (IEEE ICALT 2002)*, Kazan, Tatarstan, Russia, 2002.

1376 [59] C. Severance, T. Hanss, and J. Hardin, "Ims learning tools in-  
1377 teroperability: Enabling a mash-up approach to teaching and  
1378 learning tools," *Technology, Instruction, Cognition and Learn.*, vol. 7,  
1379 no. 3-4, pp. 245–262,

1380 [60] J. M. Kevan and P. R. Ryan, "Experience api: Flexible, decentral-  
1381 ized and activity-centric data collection," *Technology, knowledge  
1382 and Learn.*, vol. 21, no. 1, pp. 143–149, 2016.

1383 [61] A. Bakharia, K. Kitto, A. Pardo, D. Gašević, and S. Dawson,  
1384 "Recipe for success: lessons learnt from using xapi within the  
1385 connected learning analytics toolkit," in *Proc. sixth Int. Conf. on  
1386 Learn. Analytics & Knowledge*. ACM, 2016, pp. 378–382. doi:  
1387 [10.1145/2883851.2883882](https://doi.org/10.1145/2883851.2883882)

1388 [62] A. Berg, M. Scheffel, H. Drachsler, S. Ternier, and M. Specht,  
1389 "The dutch xapi experience," in *Proc. Sixth Int. Conf. on  
1390 Learn. Analytics & Knowledge*. ACM, 2016, pp. 544–545. doi:  
1391 [10.1145/2883851.2883968](https://doi.org/10.1145/2883851.2883968)

1392 [63] T. Murray, "An overview of intelligent tutoring system authoring  
1393 tools: Updated analysis of the state of the art," in *Authoring tools  
1394 for advanced techn. Learn. environments*. Springer, 2003, pp. 491–  
1395 544. doi: [10.1007/978-94-017-0819-7\\_17](https://doi.org/10.1007/978-94-017-0819-7_17)

1396 [64] K. A. Kerr, J. A. Marsh, G. S. Ikemoto, H. Darilek, and H. Barney,  
1397 "Strategies to promote data use for instructional improvement:  
1398 Actions, outcomes, and lessons from three urban districts," *Ameri-  
1399 canian Journal of Educ.*, vol. 112, no. 4, pp. 496–520, 2006.

1400 [65] M. Rey-López, P. Brusilovsky, M. Meccawy, R. Díaz-Redondo,  
1401 A. Fernández-Vilas, and H. Ashman, "Resolving the problem of  
1402 intelligent learning content in learning management systems,"  
1403 *Int. Journal on E-Learning*, vol. 7, no. 3, pp. 363–381,

1404 [66] L. Bass, P. Clements, and R. Kazman, *Software architecture in  
1405 practice*. Addison-Wesley Professional, 2003.

1406 [67] M. Derntl, "The person-centered e-learning pattern repository:  
1407 Design for reuse and extensibility," in *EdMedia: World Conf. on  
1408 Educ. Media and Technology*. Association for the Advancement of  
1409 Computing in Educ. (AACE), 2004, pp. 3856–3861.

1410 [68] S. Gilmore and M. Tribastone, "Evaluating the scalability of a web  
1411 service-based distributed e-learning and course management  
1412 system," in *Int. Workshop on Web Services and Formal Methods*.  
1413 Springer, 2006, pp. 214–226. doi: [10.1007/11841197\\_14](https://doi.org/10.1007/11841197_14)

1414 [69] J. F. Nunamaker Jr, M. Chen, and T. D. Purdin, "Systems devel-  
1415 opment in information systems research," *Journal of management  
1416 information systems*, vol. 7, no. 3, pp. 89–106, 1990.

1417 [70] L. Gilbert and V. Gale, *Principles of e-learning systems engineering*.  
1418 Elsevier, 2007.

1419 [71] P. B. Kruchten, "The 4+1 view model of architecture," *IEEE  
1420 software*, vol. 12, no. 6, pp. 42–50, 1995.

1421 [72] B. Vesin, K. Mangaroska, and M. Giannakos, "Learning in smart  
1422 environments: user-centered design and analytics of an adaptive  
1423 learning system," *Smart Learning Environments*, vol. 5, no. 1, p. 24,  
1424 2018.

1425 [73] J. Guerra, R. Hosseini, S. Somyurek, and P. Brusilovsky, "An intel-  
1426 ligent interface for learning content: Combining an open learner  
1427 model and social comparison to support self-regulated learning  
1428 and engagement," in *Proc. 21st Int. Conf. on Intelligent User Inter-  
1429 faces*. ACM, 2016, pp. 152–163. doi: [10.1145/2856767.2856784](https://doi.org/10.1145/2856767.2856784)

[74] J. Barria-Pineda and P. Brusilovsky, "Explaining educational recommendations through a concept-level knowledge visualization," in *Proc. 24th Int. Conf. on Intelligent User Interfaces: Companion*. ACM, 2019, pp. 103–104. doi: [10.1145/3308557.3308690](https://doi.org/10.1145/3308557.3308690)

[75] A. Petersen, "Programming course resource system (pcrs)," 2018. <https://mcs.utm.utoronto.ca/~pcrs/pcrs/>

[76] H. Trætteberg, A. Mavroudi, M. Giannakos, and J. Krogstie, "Adaptable learning and learning analytics: A case study in a programming course," in *European Conf. on Technology Enhanced Learning*. Springer, 2016, pp. 665–668. doi: [10.1007/978-3-319-45153-4\\_87](https://doi.org/10.1007/978-3-319-45153-4_87)

[77] K. Mangaroska, K. Sharma, M. Giannakos, H. Trætteberg, and P. Dillenbourg, "Gaze insights into debugging behavior using learner-centred analysis," in *Proc. 8th Int. Conf. on Learn. Analytics & Knowledge*. ACM, 2018, pp. 350–359.

[78] N. Herde, "The vent learning analytics dashboard-and vson container format for visualization data," Master's thesis, NTNU, 2018.

[79] F. Wang and M. J. Hannafin, "Design-based research and technology-enhanced learning environments," *Educ. technology research and development*, vol. 53, no. 4, pp. 5–23, 2005.

[80] S. McKenney and T. C. Reeves, "Systematic review of design-based research progress: Is a little knowledge a dangerous thing?" *Educ. Researcher*, vol. 42, no. 2, pp. 97–100, 2013.

[81] K. Mangaroska, B. Vesin, and M. Giannakos, "Elo-rating method: Towards adaptive assessment in e-learning," in *2019 IEEE 19th Int. Conf. on Advanced Learning Techn. (ICALT)*, vol. 2161. IEEE, 2019, pp. 380–382. doi: [10.1109/ICALT2019.00116](https://doi.org/10.1109/ICALT2019.00116)

[82] K. Mangaroska, B. Vesin, and M. Giannakos, "Cross-platform analytics: A step towards personalization and adaptation in education," in *Proc. 9th Int. Conf. on Learn. Analytics & Knowledge (LAK'19)*. ACM, 2019. doi: [10.1145/3303772.3303825](https://doi.org/10.1145/3303772.3303825)

[83] M. Cocea and G. D. Magoulas, "Participatory learner modelling design: A methodology for iterative learner models development," *Information Sciences*, vol. 321, pp. 48–70, 2015.

[84] S. Buckingham Shum, R. Ferguson, and R. Martinez-Maldonado, "Human-centred learning analytics," *Journal of Learn. Analytics*, vol. 6, no. 2, pp. 1–9, 2019.

[85] E. FitzGerald, N. Kucirkova, A. Jones, S. Cross, R. Ferguson, C. Herodotou, G. Hillaire, and E. Scanlon, "Dimensions of personalisation in technology-enhanced learning: A framework and implications for design," *British Journal of Educ. Technology*, vol. 49, no. 1, pp. 165–181, 2018.

[86] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," *Theory into practice*, vol. 41, no. 2, pp. 64–70, 2002.

[87] R. Hosseini, K. Akhuseyinoglu, A. Petersen, C. D. Schunn, and P. Brusilovsky, "Pcex: Interactive program construction examples for learning programming," in *Proc. 18th Koli Calling Int. Conf. on Computing Educ. Research*. ACM, 2018, p. 5. doi: [10.1145/3279720.3279726](https://doi.org/10.1145/3279720.3279726)

[88] K. J. Crippen and B. L. Earl, "The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy," *Comp. & Educ.*, vol. 49, no. 3, pp. 809–821, 2007.

[89] P. Brusilovsky, S. Somyurek, J. Guerra, R. Hosseini, V. Zadorozhny, and P. J. Durlach, "Open social student modeling for personalized learning," *IEEE Trans. on Emerging Topics in Computing*, vol. 4, no. 3, pp. 450–461, 2016.

[90] A. Gelman and J. Hill, *Data Analysis Using Regression and Multi-level Hierarchical Models*. Cambridge university press, 2006. doi: [10.2307/20461880](https://doi.org/10.2307/20461880)

[91] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.

[92] O. L. O. Astivia and B. D. Zumbo, "Heteroskedasticity in multiple regression analysis: What it is, how to detect it and how to solve it with applications in r and spss," *Practical Assessment, Research, and Evaluation*, vol. 24, no. 1, p. 1.

[93] H. Levene, "Robust tests for equality of variances," *Contributions to probability and statistics. Essays in honor of Harold Hotelling*, pp. 279–292,

[94] S. Lallé, C. Conati, and G. Carenini, "Prediction of individual learning curves across information visualizations," *User Modeling and User-Adapted Interaction*, vol. 26, no. 4, pp. 307–345, 2016.

[95] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22,

[96] K. Matsuki, V. Kuperman, and J. A. Van Dyke, "The random forests statistical technique: An examination of its value for the study of reading," *Scientific Studies of Reading*, vol. 20, no. 1, pp. 20–33, 2016.

[97] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *The Journal of Machine Learn. Research*, vol. 15, no. 1, pp. 3133–3181,

[98] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[99] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," in *Advances in neural information processing systems*, 2013, pp. 431–439.

[100] V. Kostakos and M. Musolesi, "Avoiding pitfalls when using machine learning in hci studies," *Interactions*, vol. 24, no. 4, pp. 34–37, 2017.

[101] D. T. Tempelaar, B. Rienties, and B. Giesbers, "In search for the most informative data for feedback generation: Learning analytics in a data-rich context," *Comp. in Human Behavior*, vol. 47, pp. 157–167, 2015.

[102] S. B. Shum and R. D. Crick, "Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics," in *Proceedings of the 2nd Int. Conf. on Learn. Analytics & Knowledge*. ACM, 2012, pp. 92–101. doi: [10.1145/2330601.2330629](https://doi.org/10.1145/2330601.2330629)

[103] L. P. Macfadyen and S. Dawson, "Numbers are not enough. why e-learning analytics failed to inform an institutional strategic plan," *Journal of Educ. Technology & Society*, vol. 15, no. 3, pp. 149–163,

[104] L. Macfadyen and S. Dawson, "Mining lms data to develop an "early warning system" for educators: A proof of concept," *Comp. & Educ.*, vol. 54, no. 2, pp. 588–599, 2010.

[105] S. Lonn, S. J. Aguilar, and S. D. Teasley, "Investigating student motivation in the context of a learning analytics intervention during a summer bridge program," *Comp. in Human Behavior*, vol. 47, pp. 90–97, 2015.

[106] A. F. Wise, J. M. Vytasek, S. Hausknecht, and Y. Zhao, "Developing learning analytics design knowledge in the" middle space": The student tuning model and align design framework for learning analytics use," *Online Learning*, vol. 20, no. 2, pp. 155–182, 2016.

[107] D. Gašević, S. Dawson, and G. Siemens, "Let's not forget: Learning analytics are about learning," *TechTrends*, vol. 59, no. 1, pp. 64–71, 2015.

[108] K. Mangaroska, K. Sharma, D. Gasevic, and M. Giannakos, "Multimodal learning analytics to inform learning design: Lessons learned from computing education," *Journal of Learning Analytics*, vol. 7, no. 3, pp. 79–97, 2020.

[109] R. Hosseini, T. Sirkiä, J. Guerra, P. Brusilovsky, and L. Malmi, "Animated examples as practice content in a java programming course," in *Proceedings of the 47th ACM technical symposium on computing science Educ.*, 2016, pp. 540–545. doi: [10.1145/2839509.2844639](https://doi.org/10.1145/2839509.2844639)

[110] A. Davidovic, J. Warren, and E. Trichina, "Learning benefits of structural example-based adaptive tutoring systems," *IEEE Trans. on Educ.*, vol. 46, no. 2, pp. 241–251, 2003.

[111] J. Sweller, P. L. Ayres, S. Kalyuga, and P. Chandler, "The expertise reversal effect," *Educational Psychologist*, vol. 38, no. 1, pp. 23–31, 2003.

[112] P. R. Pintrich, "A conceptual framework for assessing motivation and self-regulated learning in college students," *Educ. psychology review*, vol. 16, no. 4, pp. 385–407, 2004.

[113] S. Bull and J. Kay, "Open learner models," in *Advances in intelligent tutoring systems*. Springer, 2010, pp. 301–322. doi: [10.1007/978-3-642-14363-2\\_15](https://doi.org/10.1007/978-3-642-14363-2_15)

[114] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," *Educ. Psychologist*, vol. 46, no. 4, pp. 197–221, 2011.

[115] K. Mangaroska, R. Tahir, M. Lorås, and A. Mavroudi, "What do we know about learner assessment in technology-rich environments? a systematic review of systematic reviews," in *2018 IEEE 18th Int. Conf. on Advanced Learn. Techn. (ICALT)*. IEEE, 2018, pp. 16–20. doi: [10.1109/ICALT2018.00010](https://doi.org/10.1109/ICALT2018.00010)

[116] N. Rushby and D. Surry, *The Wiley Handbook of Learning Technology*. John Wiley & Sons, 2016. doi: [10.1002/9781119173243](https://doi.org/10.1002/9781119173243)

[117] J. W. Gikandi, D. Morrow, and N. E. Davis, "Online formative assessment in higher education: A review of the literature," *Comp. & Educ.*, vol. 57, no. 4, pp. 2333–2351, 2011.

1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600

**Katerina Mangaroska** has recently received her Ph.D. from the Department of Computer Science at the Norwegian University of Science and Technology. Her primary research area is learning analytics and computing education research. She was part of the "Future Learning: Orchestrating 21st Century Learning Ecosystem using Analytics" project, that aims to develop new knowledge based on how analytics allow us to better orchestrate learning tools and practices, and design optimal technology-rich distributed learning environments. Her other research interests center around learning design, multimodal learning analytics, adaptive learning, eye-tracking, intelligent tutoring systems, and human-computer interaction. Mangaroska is a Fulbright scholar.

1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
16521601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613

**Michail Giannakos** is a professor of interaction design and learning technology at the Department of Computer Science, Norwegian University of Science and Technology and research director of the Centre for Excellent IT Education. His research interests center on making sense of user experiences and practices in order to redesign and optimize the education settings and systems. He has worked at several research projects funded by diverse sources like EC, Microsoft Research, Norwegian Research Council (NRC), NSF, DAAD and Cheng Endowment; Giannakos is a recipient of a Marie Curie fellowship, the Norwegian CAREER award and he is one of the twenty outstanding academic fellows of NTNU.

1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
16521614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624

**Boban Vesin** is an associate professor at the University of South-Eastern Norway, Vestfold, Norway. He received an MSc (in Computer science) from the University of Novi Sad in 2007, and PhD (Computer science) from the same University in 2014. His research is currently focused on the engineering of e-learning systems and learning analytics to assist teachers in understanding and improving teaching processes. Boban Vesin has published over 40 scientific papers in proceedings of journals and international conferences in personalized learning, software engineering, and recommender systems.

**Vassilis Kostakos** is a professor in Human-Computer Interaction at the University of Melbourne School of Computing and Information Systems. He is a Marie Curie Fellow, a Fellow in the Academy of Finland Distinguished Professor Program, and a Founding Editor of the PACM IMWUT journal. He holds a PhD in Computer Science from the University of Bath. His research interests include ubiquitous computing (Ubicomp), human-computer interaction (HCI), social computing, and Internet of Things.

1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638

**Peter Brusilovsky** received the B.S. and Ph.D. degrees from Moscow Lomonosov State University in 1983 and 1987, respectively. He received postdoctoral training at the University of Sussex, the University of Trier, and Carnegie Mellon University. He is currently a professor of information science and intelligent systems with the University of Pittsburgh, where he directs the Personalized Adaptive Web Systems lab. He has been involved in the field of adaptive educational systems, user modeling, and intelligent user interfaces for over 20 years. He published numerous papers and edited several books on adaptive hypermedia, adaptive educational systems, user modeling, and the adaptive Web.

## APPENDIX: EXPLANATION OF ALL OF THE 55 FEATURES

Features	Explanation	Learning system
number_of_actions	# actions in a session	ProTuS
number_of_visited_content	# visited content in a session	ProTuS
level_of_coding_exercises	difficulty level of the coding exercise	ProTuS
sessions_dist	# sessions with the system	MasteryGrids
median_sessions_time	median of time spent in a session (seconds)	MasteryGrids
median_sessions_act	median of number of activities in session	MasteryGrids
median_sessions_self_assesment	median of assessment activities done in a session	MasteryGrids
median_sessions_example_lines	median of example lines clicked in a session	MasteryGrids
topics_covered	# topics covered (i.e., at least one coding exercise was solved in that topic)	MasteryGrids
pcex_topics_covered	# topics covered for pcex (i.e., at least one challenge was solved in that topic)	MasteryGrids
question_attempts	total attempts on questions	MasteryGrids
question_attempts_success	successful attempts on questions	MasteryGrids
questions_dist	distinct number of questions attempted	MasteryGrids
questions_dist_success	distinct number of successfully attempted questions	MasteryGrids
questions_sucess_first_attempt	questions solved in the 1st attempt	MasteryGrids
example_lines_actions	total number of actions in examples	MasteryGrids
pcrs_attempts	total attempts on coding exercises	MasteryGrids
pcrs_attempts_success	total attempts on coding exercises where all tests were passed	MasteryGrids
pcrs_dist	distinct number of coding exercises attempted	MasteryGrids
pcrs_dist_success	distinct number of coding exercises attempted where all tests were passed	MasteryGrids
pcrs_success_first_attempt	# coding exercises solved in the 1st attempt	MasteryGrids
pcrs_success_second_attempt	# coding exercises solved in the 2nd attempt	MasteryGrids
pcrs_success_third_attempt	# coding exercises solved in the 3rd attempt	MasteryGrids
pcex_completed_set	# set completed (a set includes an example and all of its related challenges)	MasteryGrids
pcex_ex_dist_seen	# distinct examples seen	MasteryGrids
pcex_ch_attempts	# attempts on challenges	MasteryGrids
pcex_ch_attempts_success	# successful attempts on challenges	MasteryGrids
pcex_ch_dist	# distinct challenges seen	MasteryGrids
pcex_ch_success	# distinct challenges solved	MasteryGrids
pcex_success_first_attempt	# challenges solved in the 1st attempt	MasteryGrids
pcex_success_second_attempt	# challenges solved in the 2nd attempt	MasteryGrids
pcex_success_third_attempt	# challenges solved in the 3rd attempt	MasteryGrids
mg_total_loads	MasteryGrids loads	MasteryGrids
mg_topic_cell_clicks	MasteryGrids clicks on topic	MasteryGrids
mg_activity_cell_clicks	MasteryGrids click on content	MasteryGrids
total_durationseconds	total time spend in the system in one session	MasteryGrids
quizjet_durationseconds	total time spend on quiz	MasteryGrids
pcrs_durationseconds	total time spent on coding exercises (seconds)	MasteryGrids
pcrs_durationseconds_first_attempt	time spent on coding exercises during the 1st attempt (seconds)	MasteryGrids
pcrs_durationseconds_second_attempt	time spent on coding exercises during the 2nd attempt (seconds)	MasteryGrids
pcrs_durationseconds_third_attempt	time spent on coding exercises during the 3rd attempt (seconds)	MasteryGrids
pcex_example_durationseconds	total time spent on examples before clicking any line (seconds)	MasteryGrids
pcex_example_durationseconds_median	median of time spent on examples before clicking any line (seconds)	MasteryGrids
pcex_example_lines_durationseconds	total time spent on reading example lines (seconds)	MasteryGrids
pcex_challenge_durationseconds	total time spent on challenges (seconds)	MasteryGrids
pcex_challenge_durationseconds_median	median of time spent on challenges (seconds)	MasteryGrids
pcex_challenge_durationseconds_first_attempt	time spent on challenges during the 1st attempt (seconds)	MasteryGrids
pcex_challenge_durationseconds_second_attempt	time spent on challenges during the 2nd attempt (seconds)	MasteryGrids
pcex_challenge_durationseconds_third_attempt	time spent on challenges during the 3rd attempt (seconds)	MasteryGrids
mastery_grid_durationseconds	total time spend in mastery grids statistics	MasteryGrids
total_sub_ass	total number of submitted assignments	VLASP
incom_sub_ass	number of incomplete assignments submitted	VLASP
incor_sub_ass	number of incorrectly submitted assignments	VLASP
not_sub_ass	number of not submitted assignments	VLASP
Eclipse_score	total score from all submitted assignments	VLASP