12-2-2023

Enhancing Project-based Learning through Data-driven Analysis and Visualisation: A Case Study

Dejan Simic
University of Hamburg, Germany, dejan.simic@studium.uni-hamburg.de

Stephan Leible
University of Hamburg, Germany, stephan.leible@uni-hamburg.de

Dennis Schmitz
University of Hamburg, Germany, dennis.schmitz@uni-hamburg.de

Gian-Luca Gücük
University of Hamburg, Germany, gian-luca.guecuek@studium.uni-hamburg.de

Emir Kučević
Universität Hamburg, emir.kucevic@uni-hamburg.de

Enhancing Project-based Learning through Data-driven Analysis and Visualisation: A Case Study

Full research paper

Dejan Simic
Department of Computer Science
University of Hamburg
Hamburg, Germany
Email: dejan.simic@studium.uni-hamburg.de

Stephan Leible
Department of Computer Science
University of Hamburg
Hamburg, Germany
Email: stephan.leible@uni-hamburg.de

Dennis Schmitz
Department of Computer Science
University of Hamburg
Hamburg, Germany
Email: dennis.schmitz@uni-hamburg.de

Gian-Luca Gücük
Department of Computer Science
University of Hamburg
Hamburg, Germany
Email: gian-luca.guecuek@studium.uni-hamburg.de

Emir Kučević
Department of Computer Science
University of Hamburg
Hamburg, Germany
Email: emir.kucevic@uni-hamburg.de

Abstract

This paper investigates the application of data-driven analysis and visualisation to enhance project-based courses within a higher education setting. The research focuses on a digital course where students utilise digital tools such as Jira and Confluence, which generate event logs capturing students' actions. These event logs were leveraged in conjunction with process mining and business intelligence (BI) techniques to collect and analyse the data, visualised through the iterative development and evaluation of an artefact in the form of BI dashboards following the design science research paradigm. The dashboards provide lecturers with insights into student behaviour and progress, enabling them to derive actionable suggestions for adapting student behaviour. The findings demonstrate that incorporating data-driven approaches positively impacted student engagement and improved learning outcomes. This case study contributes to the fields of learning analytics and educational data mining, offering insights into utilising data-driven approaches to enhance project-based learning experiences.

Keywords Learning analytics, educational data mining, data analytics, project-based courses, design science research

1 Introduction

Nowadays, digitalisation has rendered data collection and analysis pervasive across diverse domains in society (Lammi & Pantzar 2019), for example, customer behaviour in online stores (Khade 2016) or the utilisation of collected data to generate weather forecasts, as shown in a meta-analysis by Fathi et al. (2022). Within the realm of business, data analytics has emerged as a beneficial tool for making informed decisions (Webber & Zheng 2020). For example, the potential for process optimisation and resource allocation can also be applied in the educational sector (Romero & Ventura 2020). In this context, students can be viewed as customers who are not seeking to purchase a product but rather to obtain knowledge, achieve academic success, and be prepared for the tasks ahead (Matuga 2009).

In recent years, project-based courses have become a prominent pedagogical approach in modern (information systems) education to facilitate experiential learning and promote active student participation (Du & Han 2016). As these courses often serve as students' initial exposure to practice-oriented teaching, acquiring new competencies – including effective team communication, strong organisational skills, and subject matter proficiency – is crucial to achieving individual project goals collaboratively (Hussein 2021). As a result, students encounter new challenges as they must grapple with a host of factors, including time constraints, resource limitations, and the need to navigate interpersonal dynamics within the project teams (Grant 2011, Hussein 2021). As such, these elements can amplify the demands placed on students in addition to the inherent challenges associated with the project's content (Du & Han 2016). In this context, lecturers are critical in guiding and supporting students as they undertake (complex) project assignments and develop important skills and knowledge in a collaborative learning environment (Owens & Hite 2022). However, this task is not without its challenges.

In this paper, we address lecturers' support mechanisms for students by utilising data analysis and visualisation via dashboards to improve, refine, and broaden their feedback. By leveraging appropriate tools, lecturers can procure critical insights into student behaviour and progress. This, in turn, equips them to propose behaviour adaptations and provide targeted guidance to students, thereby facilitating students in realising their project goals, achieving successful project outcomes and meeting their learning goals. Our research spotlighted a digital, project-based course encompassing 25 university bachelor's students, tailored to offer them an inaugural hands-on experience in process mining (PM) and business intelligence (BI). In this context, while the lecturers did conduct a retrospective analysis of data post-course completion based on student outcomes, they were stymied during the actual course duration. This was primarily due to the absence of timely data insights, hindering prompt intervention, especially when students demonstrated stagnation or required course correction to enhance learning efficacy. Consequently, this problem was pivotal in our quest to design, implement, and evaluate an IT artefact, specifically BI dashboards, following the design science research (DSR) paradigm (Hevner et al. 2004, Hevner 2007). These dashboards shed light on student progression and behavioural patterns, empowering lecturers with the knowledge to suggest behavioural adaptations to support students. In a subsequent iteration, these dashboards underwent evaluation when implemented in a comparable course with 34 students the following semester. This laid the groundwork for the derivation of design principles for building BI dashboards tailored for learning analytics (LA) in project-based courses.

The courses served as a comprehensive case study for demonstrating and evaluating the benefits of data analysis and visualisation in augmenting the educational experience and enhancing student outcomes and comprehension. Our methodology and subsequent adaptations are firmly anchored in the employment of digital tools—specifically, Jira, Confluence, and Mattermost. These platforms not only facilitate collaborative endeavours among students but also serve as robust conduits for data aggregation and analysis. Jira is a project management tool, Confluence was used to build a wiki for the course, and Mattermost served as a collaboration platform. These tools provided us with the student usage data that was our fundament for all analysis. The technical environment, which includes data collection and preparation, is not within the scope of this paper, as it was already present.

Our research is situated in the subject area of e-learning, especially educational data mining (EDM) (Mohamad & Tasir 2013) and LA (Clow 2013). Although the creation and analysis of dashboards is not a new concept and is widely used in various domains for enhancing decision-making, the educational approach of utilising data from the digital tools used in the course from students to improve lecturers' feedback and enhance student projects and learning success is relatively uncommon, as demonstrated by Viberg et al. (2018). Therefore, we evaluate the usefulness and suitability of the approach through our case study, complemented by creating design principles as prescriptive knowledge to consider when building BI dashboards for LA purposes. The research question (RQ) we address in this regard is:

To what extent can the student learning outcomes in project-based courses be enhanced by analysing usage data from digital tools via BI dashboards, and what should be considered when building them?

2 Related Work

Papamitsiou and Economides (2014) conducted a literature review of empirical evidence in LA and EDM and categorised approaches such as ours as student behaviour modelling, given that participants' behaviour in the project-based course is analysed and "modelled" using BI and PM techniques. Thereby, lecturers are critical in facilitating student success by providing effective feedback and ensuring that project outcomes align with the course and learning goals (Shibani et al. 2020). While AlQaheri and Panda (2022) adopt a similar approach to understanding students' learning behaviour, they rely solely on PM and a predefined dataset. They argue that LA based on PM techniques can help to reduce (course) dropout rates and has great potential in the context of the rising popularity of digital-only courses since the COVID-19 pandemic. Our approach combines BI and PM techniques to enhance project-based courses by enabling more tailored and actionable insights from past and current course instances. According to Macak et al. (2021) and Poncin et al. (2011), similar analysis can also be applied in domains such as software development, where analysing Git logs can provide valuable insights.

Bienkowski et al. (2012) also explore the concept of enhancing student engagement by analysing and modelling their behaviour, interpreting the data, and providing proposals for adaptation based on the insights gained. Their work contributes recommendations to lecturers and students, which is a key aspect of our paper as well. The insights obtained allow lecturers to understand their course situation and verify if proposing adaptations to the students has any effect. Macfadyen and Dawson (2010) demonstrated the effectiveness of PM techniques in learning management systems and observed a correlation between the students' use of the system and their final grades.

Paredes et al. (2020) provide valuable insights into LA and EDM, particularly regarding the importance of visualisation and presenting information in dashboards. They emphasise that information should be present according to the target group, considering log data, as well as usage data, such as time invested per learning unit and course progress. A crucial aspect is the establishment of trust on the part of lecturers in the data outputs of the system so that they incorporate it into their teaching (Paredes et al. 2020). Such dashboards can also have a didactic value for students, inspiring them to continue improving their project work by receiving clear and concise feedback. In relation to the design of dashboards, Ruoff et al. (2022) delineate principles specifically tailored for crisis situations, such as the COVID-19 pandemic. Their work highlights various approaches for navigating dashboards using natural language and presenting them to a broad audience.

As described by Kubiatko and Vaculová (2011), lecturers are particularly crucial in project-based courses. They possess the knowledge and expertise to provide students with domain-specific and collaborative skills. However, this often leads to a significant workload for lecturers, limiting their capacity to support students, as highlighted by Reimann et al. (2012) in the context of schools. This issue may also apply to universities to a certain degree. To address this problem, Reimann et al. (2012) propose an approach that enables teachers to conduct and apply (data-driven) analyses themselves, given that they are the domain experts with the necessary pedagogical background to disseminate knowledge effectively. Gücük et al. (2023) propose a data analytics stack to perform such analyses.

Ethical considerations are a further aspect that should not be overlooked in LA and EDM. The use of digital data can pose risks such as data misuse or violation of privacy, especially when dealing with sensitive information such as student behaviour and grades. According to the systematic analysis of Viberg et al. (2018), around 80% of researchers do not consider ethics in their LA studies. The analysis of courses in our work relies on a limited amount of pseudonymised data, and individual patterns or performances are not considered. The course and group levels are the data levels viewed; no data is processed at the individual student level.

In summary, the fields of LA and EDM are increasingly gaining attention, particularly in light of the current push toward digital concepts and the availability of data for these approaches (Lin et al. 2021). There is a multitude of techniques to analyse and enhance education using data (Muslim et al. 2020, Ukwuoma et al. 2019). This section has only provided a sample that has inspired our work. Nevertheless, the majority of the studies cited use data to improve future courses, whereas our approach also emphasises using data for ongoing courses.

3 Method

3.1 Course Setup

The project-based course we analysed is offered digitally at a university. The course provides bachelor students with knowledge and practical experience in PM and BI techniques while following the Scrum framework (Schwaber 1997) over a semester, starting from data understanding, collection, preprocessing, integration, and graphical processing. The attendees of this course come from the disciplines of Information Systems and Computer Science and are enrolled in either their fourth or sixth semesters. The course is divided into three phases, illustrated in Figure 1.

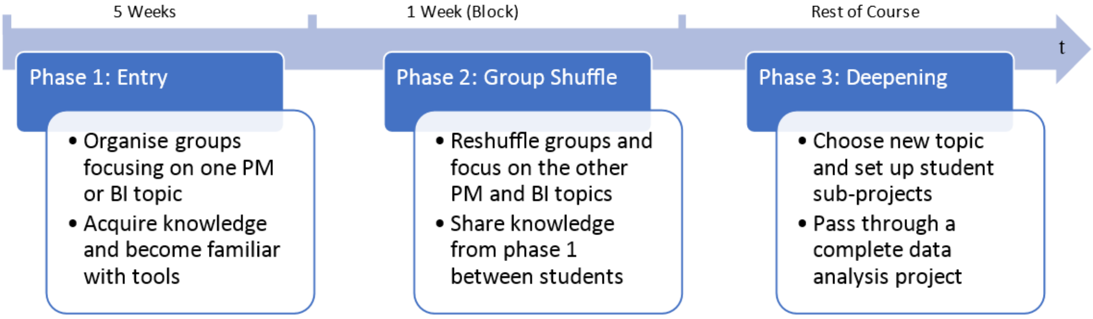

Figure 1: The Procedure of the Analysed Project-based Course

Phase 1 of the project-based course with 25 students spans five weeks, during which they are organised into (student) expert groups based on specific topics related to PM and BI for their individual projects. The goal of this phase is to acquire the knowledge needed and become familiar with the fundamental course-related tools, such as Jira and Confluence, and the project topic-related tools, such as Microsoft Power BI. For example, the BI group practices their skills with Microsoft Power BI by solving tasks assigned by the lecturers, organising their student project using Jira, Confluence, and Mattermost, and preparing intermediate results to present to lecturers and other students. In phase 2, a block session is held, in which students are reshuffled into new groups so that every group has one student expert on each PM and BI topic from phase 1. The same tasks assigned in phase 1 are repeated in the context of the new groups. Phase 2 enables students to acquire knowledge from the other topics, supported by the student experts who can guide their group since they already have expertise in their specific topic. Finally, phase 3 lasts until the end of the course, during which the groups of five to six students formed in phase 2 conduct data analytic processes in sub-projects and gain more practical experience.

Figure 2 depicts the flow and utilisation of data in the dashboards developed for the analysed course. Although the main focus of this work is the visualisation of structured data from the data warehouse, as described by Chaudhuri and Dayal (1997), it is essential to demonstrate the source of the data.

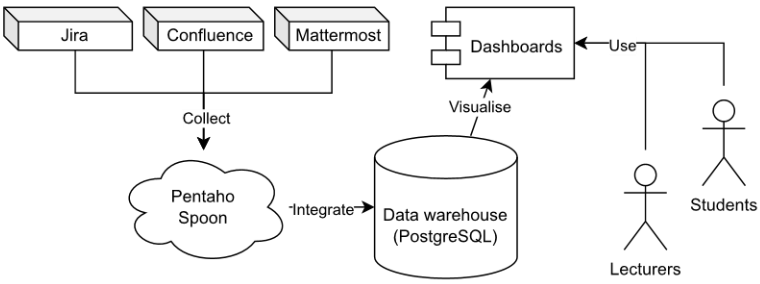

Figure 2: Data Flow and Utilisation in the Analysed Project-based Course

In this course, communication and project management rely on the tools Jira, Confluence, and Mattermost, which generate event logs when certain activities occur, such as when a Jira ticket is created. These logs store metadata such as the creator, assignee, creation timestamp, and other relevant information about the students' behaviour and progress in the respective digital tools' databases. To collect this data, we use Pentaho Spoon, a low-code platform that enables the construction of Extract-Transform-Load (ETL) processes, as described in Salaki et al. (2016). This process can be expanded to accommodate additional data sources and integrate new digital tools into the existing tool stack. Once transformed, the data is uploaded to a data warehouse based on PostgreSQL. This technical infrastructure was already established before the research commenced and served as our starting point. We then utilised the data extracted from the data warehouse to develop the dashboards using Microsoft Power BI, with particular attention given to tailoring them to the intended audience. In our case, the primary audience was the lecturers, and the secondary audience was the student tutors. The lecturers used our dashboards once a week to prepare feedback through regular group meetings. Feedback on a more high-level perspective was also communicated to the students through various formats, such as presentations or blog posts, aiming to improve their behaviour and long-term learning success.

3.2 Design Science Research

We adhered to the DSR paradigm by Hevner et al. (2004), motivated by the practice-driven problem of inadequate data availability for lecturers supervising project-based courses within a higher education setting. Consequently, BI dashboards were conceived as an artefact and were iteratively developed and appraised over two cycles to address the problem. Drawing from both the existing literature and insights from our evaluations, we established design principles, presenting them as prescriptive knowledge tailored to constructing BI dashboards in analogous course environments. In alignment, we undertook eight research steps mapped to Hevner's (2007) three cycle view, as delineated in Figure 3.

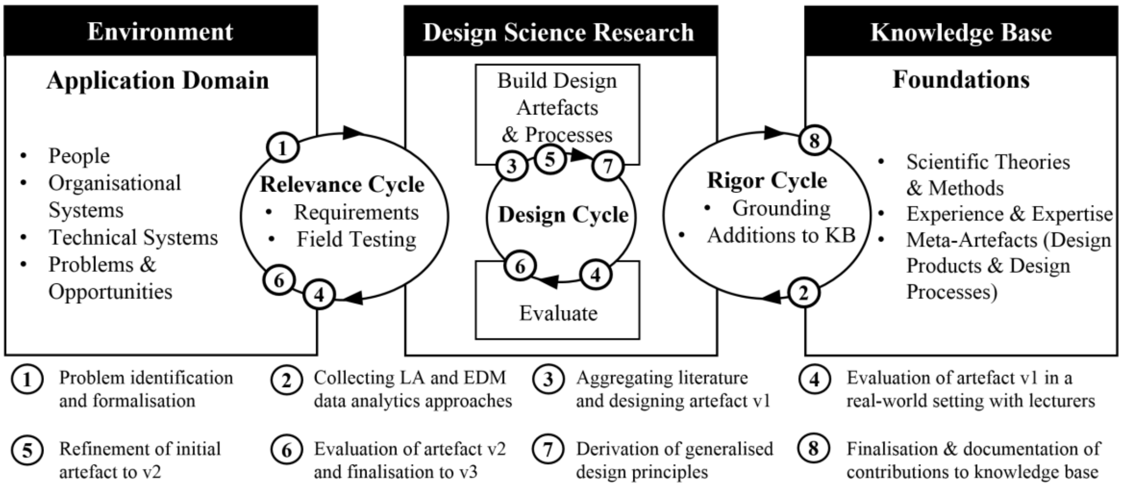

Figure 3: Design Science Research Steps according to the Three Cycle View of Hevner (2007)

In the initial *step 1*, problem identification and formalisation were performed (see Introduction). This includes recognising the problem faced by multiple lecturers overseeing project-based courses within the department of computer science. It included primarily the inability to sufficiently supervise students, particularly those who remained inconspicuous within the course or their respective groups. This could be due to their hesitance to share information or reluctance to seek assistance. Consequently, there was a need for a data-driven overview capturing the progress of the course and the various student groups therein, intending to enhance the support quality. Within this framework, supervising project-based courses using a data-driven approach represents a broader problem class.

We delved into LA and EDM literature during *step 2* to discern potential solutions, extracting data analysis methodologies (see Related Work). This exploration aimed to discern viable approaches and lay the groundwork. Stemming from this, a BI dashboard was conceptualised as an artefact in *step 3*. The premise for this was derived from student interactions with a myriad of digital tools, producing log data amenable to analysis. A pre-existing technical infrastructure facilitated data centralisation from diverse application programming interfaces, as illustrated in Figure 2. Incorporating insights from the literature and our understanding of the project-based course, step 3 also involved a situational assessment of extractable data. Consequently, the BI dashboard's inaugural version 1 was designed and implemented using Microsoft Power BI. Drawing upon Gregor and Hevner (2013), the artefact's contribution type is classified as a situational implementation at the first level. Meanwhile, the proposed design principles are classified as operational design knowledge at the second level. This classification leverages an exaptation strategy, wherein we repurpose dashboard visualisations from disparate sectors (for example, controlling and maintenance) and adapt them to an educational setting.

The implemented dashboard underwent rigorous testing and evaluation during *step 4* (relevance and design cycle). This was facilitated through weekly unstructured interviews (Zhang & Wildemuth 2009) involving lecturers and student tutors during a project-based course as a field test with 25 participants. The focal points of these interviews spanned usability, graphical perception, the utility of the artefact, any discrepancies within the artefact, potential enhancements, and overall user satisfaction. Our primary aim was to harness data from various digital tools to proffer lecturers a comprehensive overview of their project-based course, thereby augmenting feedback quality and pinpointing areas for practical student behaviour adaptation suggestions. All feedback was gathered, documented, and subsequently analysed. This laid the groundwork for revising the artefact into its second version during *step 5*.

In subsequent *step 6*, under the relevance and design cycle umbrella, this refined dashboard was deployed within a new course environment with 34 participants, in which the same digital tools plus

Gitlab are used. We mirrored the evaluation protocols from the fourth phase, with the qualitative feedback from new lecturers and student tutors informing the development of the artefact's final iteration: Version 3. In *step 7*, informed by the insights accrued, we delineated nine overarching normative design principles to expand the knowledge base. These should be considered for building and utilising dashboards in project-based course environments for LA. This step is consistent with Drechsler and Hevner's (2018) procedure, which emphasises the enrichment of collective human knowledge through the use of project-specific insights. *Step 8* comprises a synthesis, formalisation, and documentation of our findings from our DSR project through this paper. This aims to juxtapose our findings in the context of addressing our RQ. The following sections address steps 3 to 8.

4 Results

4.1 Artefact v1

The initial artefact (v1) analysed the project-based course, focusing on generic insights obtained from Jira, Confluence, and Mattermost using BI and PM techniques. Specifically, we addressed aspects such as the number of events generated per user and the distribution of event generation throughout the week. These insights allowed us to understand how students' behaviour progresses from different perspectives, especially in terms of the work done per user and their worktime preferences. Additionally, the analysis of the Jira data focused on how appropriately the tickets were used during the students' work. This data provided insights into how many tickets were assigned to a person and how long it took the students to log their work after completing it. After the establishment of artefact v1 in the first weeks, a wealth of insights was gained about the student's use of the digital tools, leading to recommendations for adapting their behaviour. Initial feedback from lecturers and student tutors indicated that the dashboard appeared disorganised and challenging to read, particularly for certain chart types. We addressed errors and feedback to improve the artefact's quality, ultimately resulting in a second version. Due to space limitations, a visualisation of our PM approach can be found in the appendix.

4.2 Artefact v2

The main differences between the artefacts v1 and v2 pertain to design and layout, error corrections, and the inclusion of more information. After a few weeks into the course, artefact v2, illustrated in Figure 4, replaced v1 and was also the foundation for transferring it into a new course environment.

Figure 4: Artefact v2 Dashboard of Jira Worklog

This artefact addresses errors identified in v1, including timestamp discrepancies resulting from different time zones, which led to the oversight or exclusion of certain data from the analysis. To resolve this, we changed the filters in the analyses from CEST to UTC. Another problem was that the data in the students' work logs differed from those in Jira. Many students recorded their times on Epics instead of

User Stories or Issues at the beginning of the course, but work logs for Epics were not included in the Pentaho Spoon transformation since it was not a common practice in the course. As a result, the data was incomplete, and the dashboards did not accurately reflect the work status. To resolve this, we shifted the work logs to the related Issues and instructed students to adapt their working habits accordingly.

Finally, we established more charts to satisfy the new requirements, covering further detailed aspects of the student groups of the project-based course. Particularly relevant are aspects that analyse the assignment of tickets, working time by project, the distribution of work types (e.g., setup or implementation) among all groups, and how much time each group spent on specific work types. This way, the lecturers can determine each group's effort and identify whether they are struggling in a particular work phase, such as setting up the technical environment. Furthermore, the Confluence dashboard was expanded with charts dealing with active events, such as creating or updating pages. This enables tracking of how well the students document their work and decisions. With these insights, lecturers were able to create more relevant proposals for the students to adapt their behaviour positively.

4.3 Artefact v3

We expanded artefact v2, resulting in the final version, artefact v3, developed in this research project. This was based on the insights we gained from phase 3 of the first course and the utilisation of our artefact in another course (see Evaluation). As the student groups were assigned new topics for the final project phase, we adapted the dashboards accordingly. This included updating the Confluence pages and Jira boards used. Additionally, a lecturer noticed irregularities in the Jira work logs. Some students had multiple logs for the same time slot, resulting in overlapping bookings. The lecturers investigated further, leading to more occurrences being identified. Students were instructed to clean up their bookings, and the course material was adjusted to avoid this issue in the future.

The lecturers wished to gain insights into the complete development of all digital tools used in the project-based course through a single chart, and a line chart was deemed helpful for such use cases, as highlighted by Archambault et al. (2015). As a result, a summary dashboard was created, illustrated in Figure 5. The x-axis represents the course weeks, and the two y-axes indicate the number of events collected for that week. Two y-axes were necessary as different events were collected for Mattermost and Jira. This way, the chart stays organised and allows comparison between the tools.

The final version of the artefact enables tracking of how students utilised the digital tools throughout the course. For example, there was a decrease in usage during weeks seven and eight. This is because of public holidays during that time, and the course was not conducted, which resulted in students barely continuing their work. The final version of the artefact met all the requirements of the lecturers and student tutors. It facilitates the extraction of pertinent insights about the analysed course, leading to informed recommendations for adapting student behaviour. In addition, through the use of the dashboards, areas for improvement in the course materials were identified that contained unclear instructions and led to unwanted behaviour.

Figure 5: Summary Dashboard for Insights into the Complete Use of Digital Tools of the First Course

5 Evaluation

This section provides a comparison between our instance of the analysed project-based course using the developed artefact with 25 students and the same course in a previous semester with 30 students where the artefact (i.e., dashboards) was not utilised. With this framework, we examined a specific metric as an example of the analysed course instance and compared all metrics with data from the previous instance to evaluate the impact of the proposed adaptations on student behaviour. Additionally, we present a case study of a similar course, albeit not identical, to test the artefact's transferability.

To represent the impact compactly, we exemplarily focus on the single metric, "*Time from In Progress to Done*", analysed using the PM analysis for the Jira status. This metric describes the median duration of how long a ticket remained *In Progress* before it was transitioned to *Done*. As with any task, the duration to complete it depends on its complexity. In our case, we observed unrealistic values in weeks 2 and 3, where some tasks were completed in mere seconds, indicating that students worked on their tasks while the Jira status was set to *To-do* and quickly transitioned them to *In Progress* and then to *Done*. We communicated this issue to students in weeks 2 and 3, emphasising the importance of proper ticket status transitions. Consequently, the median increased to about 70 minutes in week 4. We view this as a successful illustration of how data-driven recommendations from lecturers can improve student behaviour. However, some metrics, such as those for Mattermost, have not performed satisfactorily, and proposals for adaptations have had minimal impact on tool usage, which has only slightly increased.

Several differences can be observed when comparing the course's current instance to the previous one's data. One point to note is that the number of overall actions was higher in the previous instance due to the participation of more students. While the previous instance had a strong start with more generated data, it decreased over time, reaching an all-time low in week 8. In contrast, as shown in Figure 5, the current course instance had less data generated in the initial weeks, but it showed a better development trend over time. There was a consistent rising trend until phase 3, except for Mattermost. Notably, in the previous instance, the students and lecturers had no dashboards. In summary, the current course instance showed a progression that is preferable to the previous one because it aligns with the course's learning goals by emphasising the proper handling of the utilised digital tools.

In addition, we evaluated the transferability of our artefact by applying it to a different project-based course with 34 students, which revealed certain limitations. The students also utilised the version control system Gitlab in this course, besides Jira, Confluence, and Mattermost. Furthermore, the course had a distinct project structure where the student groups worked cross-topic. As a result, the one-to-one relationship between the topic and group was dissolved, presenting a fundamental challenge for all group-related diagrams. Thus, significant adjustments would be required to transfer group-related data to courses with different project structures, as in this case. However, since three of the same digital tools were used in both courses, the other diagrams unrelated to student groups were transferable with minor modifications, such as adjusting the status names. The lecturers integrated the dashboards into their regular meetings with student groups, where they discussed the current project situation and provided proposals for adaptations. With the feedback from the lecturers, we noted that their data literacy was essential in facilitating their comprehension of the dashboards and deriving adaptation suggestions for student behaviour. Nevertheless, the dashboards showed weekly improvements in the course, enabling the lecturers to enhance course results and compare the impact of measures in future course instances.

6 Design Principles

As described by Bienkowski et al. (2012), it is crucial to generalise the results in one context to be adapted and usable for other contexts as well. To generalise the outcomes of this paper, we derived design principles to consider as normative high-level guidelines when building and utilising BI dashboards for LA in the context of project-based courses in higher education. As Drechsler and Hevner (2018) note, the contributed knowledge from DSR projects should be in a transferable form that allows to address similar problems of the same problem class. By applying our design principles, we support BI dashboards' planning, creation, and usage for LA use cases, particularly for project-based courses. Table 1 shows the nine design principles we derived from the literature and the insights of our case study.

The first design principle underscores the significance of *clarity and simplicity*. Bienkowski et al. (2012) highlight the necessity of exploring alternative methods of data visualisation, with an emphasis on simplicity to mitigate cognitive load. Our case study affirms the paramountcy of utilising straightforward visualisations, such as bar or line charts, over more intricate displays like 3D diagrams. Users may struggle to comprehend the dashboard if a chart exhibits undue complexity or conveys unintended or erroneous messages. Adhering to a selection of two to four chart types streamlines the user experience, sparing them the challenge of interpreting an array of diverse charts. Paredes et al. (2020) corroborate the importance of opting for the most appropriate chart visualisation, as users must glean insights autonomously. This design principle's genesis is multifaceted and influenced by factors such as granularity and aggregation level, as Bennet and Folley (2020) posited.

Design principle	Description
(1) Clarity and simplicity	Opt for simplistic chart designs, ensuring that the selected chart type is congruent with the nature of the data being visualised. Refrain from intricate charts to mitigate cognitive strain and facilitate an overview.
(2) User-centric design	Elements, including dashboard layout, intricacy, and lexicon, should be meticulously calibrated to align with the intended audience's expertise, proficiencies, and requirements.
(3) Customisability and interactivity	Dashboards should be constructed with modularity and interactivity in mind. Features, such as temporal filters, should be integrated, allowing the display to be tailored to meet the users' requirements.
(4) Privacy and ethical considerations	Privacy and ethical standards should be rigorously upheld when designing dashboards. User data must be anonymised and utilised only with informed consent, ensuring transparency and trustworthiness in analytics processes.
(5) Documentation	Thoroughly documenting decisions, functionalities, and dashboard utilisation protocols to facilitate customisation by subsequent users should be made.
(6) Actionable insights	Dashboards should be designed to deliver precise, actionable insights. This ensures that lecturers can swiftly identify and execute relevant interventions.
(7) Historical comparison	Progress over time should be monitored by incorporating historical data, such as prior courses, enabling benchmarking and comparisons.
(8) Data relevance	Given the limited space on the dashboards, the relevance of the data must be carefully considered. The chosen data should link with the learning goals of the project-based course.
(9) Data topicality	Dashboards should be designed to ensure data is delivered promptly, in real-time or based on specific triggers so that the highlighted prevalent issues assist lecturers in their feedback to students.

Table 1. Derived Design Principles for BI Dashboards in Project-based LA Use Cases

The second design principle underscores a *user-centric design* approach for BI dashboards. As stated by Paredes et al. (2020), the design must resonate with its intended audience. This encompasses aesthetic considerations such as adhering to the corporate design. Beyond visual consistency, the linguistic content of BI dashboards warrants equal attention. It is imperative to use terminology that is familiar to the target demographic. Emerging from our experiences and the literature is the design principle emphasising *customisability and interactivity*. Adopting a modular approach in BI dashboards is pivotal, especially when user-specific adaptations, such as the replacement of particular diagrams, become necessary. Verbert et al. (2020) advance the argument that future iterations of dashboards for LA purposes will need enhanced configurability, heightened interactivity, and more user-oriented customisability. Charleer et al. (2016) reaffirm this sentiment, noting that interactive dashboards empower users to delve deeper into the data, facilitating richer insights and catering more effectively to individualised requirements.

Another design principle that surfaced prominently from our case study and the broader literature is the imperative of *privacy and ethical considerations*. While the sanctity of user data is paramount, privacy should not act as a debilitating constraint to insightful data analysis. A growing body of literature, including findings by Kaliisa et al. (2023) and Viberg et al. (2018), indicates that the areas of privacy and ethics have often been overshadowed or inadequately addressed in data-driven endeavours. As such, the incorporation of ethical and privacy-conscious methodologies into dashboard construction is not just advisable but imperative. For example, data anonymisation is a crucial step to ensure the protection of individual identities. At the same time, it is important to exercise caution with visualisations that might inadvertently reveal specific patterns or associations that could lead back to individual users or specific demographics like gender or age. Bienkowski et al. (2012) further emphasise the dual responsibility of ensuring impeccable data anonymisation while progressing in LA. However, it is crucial to strike a balance. While privacy and ethics are essential, the design of dashboards should not be dominated by these concerns that the quality and utility of insights are compromised.

Comprehensive *documentation* represents another design principle. According to Paredes et al. (2020), when architecting visualisation systems, such as dashboards, it becomes essential to record the rationale behind design choices, as well as any alternative approaches considered but not pursued. Beyond the foundational design aspects, it is equally important to provide detailed documentation on the operational facet of the dashboard. This encompasses guidelines on its usage, maintenance, potential expansion approaches, and a thorough breakdown of any intricate algorithms employed. Such rigorous documentation serves as a foundational guide for current users and a valuable resource for those who might inherit, modify, or build upon the dashboard. The subsequent design principle emphasises the imperative of producing *actionable insights*. Kaliisa et al. (2023) underscore the importance of deriving actionable intelligence from dashboards tailored to lecturers. Such insights should be directly translatable into pragmatic measures, for example, issuing adaptive announcements that lecturers can effectively convey to students. Furthermore, Paredes et al. (2020) highlight the pivotal role of trust in assimilating insights into pedagogical strategies. The capacity to navigate a dashboard and subsequently deduce actionable steps inherently bolsters such trust. This sentiment was palpably reinforced through the observations in our specific case study, for example, the incorrect handling of Jira issues.

The following design principle is *historical comparison*. Jivet et al. (2017) point out that while historical data remains pivotal, many dashboards delineated in the literature primarily focus on the current status of learners. As exemplified in our research, incorporating historical data facilitates a robust comparison and benchmarking. By designing dashboards with capabilities such as capturing temporal snapshots or filtering through past time windows, lecturers are better equipped to monitor longitudinal progress and discern emergent trends throughout the courses.

Emphasising *data relevance* is paramount. The mere abundance of data is not inherently valuable; it must directly address the queries that lecturers wish to resolve, evaluating its pertinence and utility to mirror their pedagogical practices (Verbert et al. 2013). Consequently, the inquiries addressed by the dashboards, and more pertinently, the data presented, must be aligned with the learning goals of the course, ensuring support and guidance for its intended purpose. The last design principle emphasises the temporal relevance of the *data topicality* that dashboards can provide. Shimada et al. (2018) delineate feedback frequencies into annual, weekly, and real-time categories. Our dashboards accommodate these temporal modalities, given that data refreshing is facilitated on-demand by a button as a trigger. This versatility ensures that lecturers can retrospectively analyse information, address challenges in collective sessions, or promptly react to incidents during active project work.

7 Discussion and Conclusion

In our DSR project, we leveraged the existing technical infrastructure to develop and evaluate an artefact in the form of BI dashboards. These dashboards were designed to provide lecturers with valuable insights into student group behaviour in a project-based course of 25 students, enabling them to identify improvement areas for positively adapting student behaviour. Additionally, we derived nine design principles for researchers and practitioners for building BI dashboards for LA purposes and expanding the prescriptive knowledge base. Based on the evaluation and normative guidelines, we assert that our RQ has been addressed. Lecturers of our case study can now provide more sophisticated feedback, fostering students to enhance their course performance and achieve learning goals more effectively.

Our approach showed transferability to another course with 34 students and similar technical infrastructure, requiring only minor modifications for data collection and analysis. However, when incorporating new digital tools, customisation of the ETL process is necessary to accommodate the tool stack before dashboards can be created to analyse the data and derive adaptation suggestions from lecturers (Gücük et al. 2023). This customisation process demands a significant investment of time and is best suited for courses held regularly rather than one-time ones. Our approach and design principles can serve as a template, source of inspiration, and guidelines for analysing various data types in different project-based courses, designing BI dashboards, and supporting lecturers and student tutors.

7.1 Theoretical Implications

From a theoretical standpoint, it is evident that lecturers play a critical role in incorporating data-driven analysis, as highlighted by scholars such as Çakiroğlu and Erdemir (2019). As noted by Bienkowski et al. (2012) and Drechsler and Hevner (2018), a generalisation of results is needed in order to learn from one context and integrate the results into another, for example, a problem of the same problem class. Therefore, the design principles have been derived, which present normative guidelines to researchers and practitioners when dealing with BI dashboards for LA purposes. In addition to imparting knowledge to students, lecturers must possess the expertise and adequate data literacy to utilise data effectively and derive informed suggestions for adapting student behaviour. Moreover, for the dashboards to have an impact, lecturers must embrace this approach and integrate the analysis into regular meetings with student groups, as Joshi et al. (2020) propose. By providing comprehensive feedback before or during group discussions based on the dashboard data, lecturers can consistently challenge students to reflect on their performance. This requires a foundation that aligns the dashboards to the target audience, focusing on clarity and simplicity as well as supporting deriving actionable insights.

Adopting a similar approach, Ruoff et al. (2022) formulate design principles for conversational dashboards tailored for crisis response. They incorporated natural language interactions alongside conventional navigation mechanisms. While our second design principle touches upon this, Ruoff et al. (2022) present more specific principles adapted to their particular use case, enriching the discourse on user-centric design. Stemming from their principle of self-teaching dashboards, the context of crisis response is distinct. In our study, as well as in many educational settings, a lecturer typically introduces and guides the utilisation of dashboards. Therefore, pivotal questions need to be addressed in the design of dashboards, such as the suitability of different types of navigation or the need for a guiding person.

Our work highlights, in line with Paredes et al. (2020), the potential to enhance project-based courses by employing analysis and visualisation of data extracted from digital tools used by students. In doing so, we emphasise the involvement of both lecturers and students throughout the development, aiming to create a more beneficial learning environment. We assume from our findings that adopting data-driven approaches can further enhance students' (extrinsic) motivation to utilise the analysed digital tools, which are integral to achieving the desired course outcomes. This aligns with the findings of Macfadyen and Dawson (2010) regarding the motivational impact of online learning platforms. As Mattern (2005) demonstrates, motivation is pivotal in facilitating successful learning experiences and should be prioritised in educational contexts. This implies a bidirectional impact, as such analysis can lead to motivational benefits and reveal errors in course materials, further benefiting the lecturers.

7.2 Practical Implications

From a practical perspective, our work presents an exemplary approach applicable to project-based courses where student groups utilise digital tools that provide usage data for analysis. However, careful preparations are needed for transferability. For example, data at the individual student level, especially in small cohorts like our use case, can be highly sensitive and require measures such as pseudonymisation or anonymisation to protect privacy (Viberg et al. 2018), as we integrated into our design principles. Consequently, preprocessing efforts are essential to ensure students cannot identify each other in the dashboards. In addition, we emphasise transparently communicating with students about data collection and tracking practices to foster a deeper understanding and heightened awareness of the purpose and implications of data-driven analysis in their learning process. Documenting the creation process of the dashboards, as suggested in this work, helps increase transparency as students and lecturers are able to understand how their data is processed, ultimately gaining trust.

7.3 Limitations and Future Research

Nevertheless, this work is not without limitations that should be acknowledged. It is important to recognise that the behavioural adaptation proposals derived from this analysis are subject to the lecturers' subjectivity and are specifically tailored to the analysed courses. As the dashboards were developed based on two project-based courses, they are highly adapted to the stack of digital tools used, making it more challenging to adapt the same dashboards to entirely different project-based courses. Thus, design principles are derived to enable the use of the results across different contexts and use cases. Another limitation of this work is the representative nature of the literature review conducted. While efforts were made to cover pertinent research, some relevant work may have been missed.

In future research, exploring methods to enhance understandability, usage behaviour, and motivation through improved data visualisation techniques could yield more profound insights. Additionally, incorporating practices such as "data vomit" and refining raw data into more stable formats, as Patil (2012) demonstrated, could enhance the artefact's effectiveness. Moreover, we believe there is promise in investigating the relationships across different digital tools, as Figure 5 illustrates a starting point. This can enable more in-depth analyses, leading to meaningful findings and identifying behavioural patterns across multiple digital tools. Finally, further research can be done on advancing the developed design principles for building BI dashboards for LA purposes, particularly in project-based courses.

In summary, the fields of LA and EDM are gaining relevance within the context of e-learning, particularly in relation to the growing number of project-based courses that have demonstrated effective learning outcomes for students. These courses are vital in equipping students with essential skills for their future careers (Power 2016). Our analysis focused on a digital course mirrored the ongoing evolution in many domains, such as software development, where decentralised teams collaborate remotely (Marek et al. 2021). In this context, data-driven approaches in education hold tremendous potential for enhancing teaching practices and fostering meaningful and impactful learning experiences.

8 References

AlQaheri, H. & Panda, M. (2022): An Education Process Mining Framework: Unveiling Meaningful Information for Understanding Students' Learning Behavior and Improving Teaching Quality. In: *Information* 13:1, (doi: 10.3390/info13010029).

Archambault, S. G., Helouvry, J., Strohl, B. & Williams, G. (2015): Data visualization as a communication tool. In: *Library Hi Tech News* 32:2, pp. 1-9, (doi: 10.1108/LHTN-10-2014-0098).

Bennett, L. & Folley, S. (2020): Four design principles for learner dashboards that support student agency and empowerment. In: *Journal of Applied Research in Higher Education* 12:1, pp. 15-26, (doi: 10.1108/JARHE-11-2018-0251).

Bienkowski, M., Feng, M. & Means, B. (2012): Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. U.S. Department of Education Available at: http://files.eric.ed.gov/fulltext/ED611199.pdf. Last accessed: 8/25/2023.

Çakiroğlu, Ü. & Erdemir, T. (2019): Online project based learning via cloud computing: exploring roles of instructor and students. In: *Interactive Learning Environments* 27:4, pp. 547-566, (doi: 10.1080/10494820.2018.1489855).

Charleer, S., et al. (2016): Creating Effective Learning Analytics Dashboards: Lessons Learnt. In: *European Conference on Technology Enhanced Learning*, Springer, Cham, pp. 42-56, (doi: 10.1007/978-3-319-45153-4_4).

Chaudhuri, S. & Dayal, U. (1997): An overview of data warehousing and OLAP technology. In: *ACM SIGMOD Record* 26:1, pp. 65-74, (doi: 10.1145/248603.248616).

Clow, D. (2013): An overview of learning analytics. In: *Teaching in Higher Education* 18:6, pp. 683-695, (doi: 10.1080/13562517.2013.827653).

Drechsler, A. & Hevner, A. R. (2018): Utilizing, Producing, and Contributing Design Knowledge in DSR Projects. In: *Proceedings of the International Conference on Design Science Research in Information Systems and Technology (DESRIST2018)*, pp. 82-97, (doi: 10.1007/978-3-319-91800-6_6).

Du, X. & Han, J. (2016): A Literature Review on the Definition and Process of Project-Based Learning and Other Relative Studies. In: *Creative Education* 7:7, pp. 1079-1083, (doi: 10.4236/ce.2016.77112).

Fathi, M., Kashani, M. H., Jameii, S. M. & Mahdipour, E. (2022): Big Data Analytics in Weather Forecasting: A Systematic Review. In: *Archives of Computational Methods in Engineering* 29, pp. 1247-1275, (doi: 10.1007/s11831-021-09616-4).

Grant, M. M. (2011): Learning, Beliefs, and Products: Students' Perspectives with Project-based Learning. In: *Interdisciplinary Journal of Problem-Based Learning* 5:2, (doi: 10.7771/1541-5015.1254).

Gregor, S. & Hevner, A. R. (2013): Positioning and Presenting Design Science Research for Maximum Impact. In: *MIS Quarterly* 37:2, pp. 337-355.

Gücük, G.-L., et al. (2023): Enhancing educational insights: A real-time data analytics stack for project-based learning. In: *Proceedings of the INFORMATIK 2023 - 53rd Jahrestagung der Gesellschaft für Informatik,* Berlin, Germany, pp. 1963-1971.

Hevner, A. R. (2007): A Three Cycle View of Design Science Research. In: *Scandinavian Journal of Information Systems* 19:2, pp. 87-92.

Hevner, A. R., March, S. T., Park, J. & Ram, S. (2004): Design Science in Information Systems Research. In: *MIS Quarterly* 28:1, pp. 75-105.

Hussein, B. (2021): Addressing Collaboration Challenges in Project-Based Learning: The Student's Perspective. In: *Education Sciences* 11:8, (doi: 10.3390/educsci11080434).

Jivet, I., Scheffel, M., Drachsler, H. & Specht, M. (2017): Awareness Is Not Enough: Pitfalls of Learning Analytics Dashboards in the Educational Practice. In: *European Conference on Technology Enhanced Learning*, Springer, Cham, pp. 82-96, (doi: 10.1007/978-3-319-66610-5_7).

Joshi, A., Desai, P. & Tewari, P. (2020): Learning Analytics framework for measuring students' performance and teachers' involvement through problem based learning in engineering education. In: *Procedia Computer Science* 172, pp. 954-959, (doi: 10.1016/j.procs.2020.05.138).

Kaliisa, R., Jivet, I. & Prinsloo, P. (2023): A checklist to guide the planning, designing, implementation, and evaluation of learning analytics dashboards. In: *International Journal of Educational Technology in Higher Education* 20:1, pp. 1-22, (doi: 10.1186/s41239-023-00394-6).

Khade, A. A. (2016): Performing Customer Behavior Analysis using Big Data Analytics. In: *Procedia Computer Science* 79, pp. 986-992, (doi: 10.1016/j.procs.2016.03.125).

Kubiatko, M. & Vaculová, I. (2011): Project-based learning: Characteristic and the experiences with application in the science subjects. In: *Energy Education Science and Technology Part B-Social and Educational Studies* 3:1, pp. 65-74.

Lammi, M. & Pantzar, M. (2019): The data economy: How technological change has altered the role of the citizen-consumer. In: *Technology in Society* 59, (doi: 10.1016/j.techsoc.2019.101157).

Lin, C.-L., et al. (2021): Factors Influence Students' Switching Behavior to Online Learning under COVID-19 Pandemic: A Push–Pull–Mooring Model Perspective. In: *The Asia-Pacific Education Researcher* 30:3, pp. 229-245, (doi: 10.1007/s40299-021-00570-0).

Macak, M., Kruzelova, D., Chren, S. & Buhnova, B. (2021): Using process mining for Git log analysis of projects in a software development course. In: *Education and Information Technologies* 26, pp. 5939-5969, (doi: 10.1007/s10639-021-10564-6).

Macfadyen, L. P. & Dawson, S. (2010): Mining LMS data to develop an "early warning system" for educators: A proof of concept. In: *Computers & Education* 54:2, pp. 588-599, (doi: 10.1016/j.compedu.2009.09.008).

Marek, K., Wińska, E. & Dąbrowski, W. (2021): The State of Agile Software Development Teams During the Covid-19 Pandemic. In: *Proceedings of the International Conference on Lean and Agile Software Development (LASD 2021)*, pp. 24-39, (doi: 10.1007/978-3-030-67084-9_2).

Mattern, R. A. (2005): College Students' Goal Orientations and Achievement. In: *International Journal of Teaching and Learning in Higher Education* 17:1.

Matuga, J. (2009): Self-Regulation, Goal Orientation, and Academic Achievement of Secondary Students in Online University Courses. In: *Educational Technology & Society* 12:3, pp. 4-11.

Mohamad, S. K. & Tasir, Z. (2013): Educational Data Mining: A Review. In: *Procedia - Social and Behavioral Sciences* 97, pp. 320-324, (doi: 10.1016/j.sbspro.2013.10.240).

Muslim, A., Chatti, M. A. & Guesmi, M. (2020): Open Learning Analytics: A Systematic Literature Review and Future Perspectives. In: *Artificial Intelligence Supported Educational Technologies*, N. Pinkwart & S. Liu (eds.), Springer Cham, pp. 3-29.

Owens, A. D. & Hite, R. L. (2022): Enhancing student communication competencies in STEM using virtual global collaboration project based learning. In: *Research in Science & Technological Education* 40:1, pp. 76-102, (doi: 10.1080/02635143.2020.1778663).

Papamitsiou, Z. & Economides, A. A. (2014): Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. In: *Educational Technology & Society* 17:4, pp. 49-64.

Paredes, Y. V., Siegle, R. F., Hsiao, I.-H. & Craig, S. D. (2020): Educational Data Mining and Learning Analytics for Improving Online Learning Environments. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 64:1, pp. 500-504, (doi: 10.1177/1071181320641113).

Patil, D. J. (2012). *Data Jujitsu: The Art of Turning Data into Product*, O'Reilly Media.

Poncin, W., Serebrenik, A. & van den Brand, M. (2011): Process Mining Software Repositories. In: *Proceedings of the 15th European Conference on Software Maintenance and Reengineering*, (doi: 10.1109/CSMR.2011.5).

Power, D. J. (2016): Data science: supporting decision-making. In: *Journal of Decision Systems* 25:4, pp. 345-356, (doi: 10.1080/12460125.2016.1171610).

Reimann, P., et al. (2012): Supporting teachers in capturing and analyzing learning data in the technology-rich classroom. In: *Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012)*, pp. 33-40.

Romero, C. & Ventura, S. (2020): Educational data mining and learning analytics: An updated survey. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10:3, (doi: 10.1002/widm.1355).

Ruoff, M., Gnewuch, U., Maedche, A. & Scheibehenne, B. (2022): Designing Conversational Dashboards for Effective Use in Crisis Response. In: *Journal of the Association for Information Systems Preprints (Forthcoming)* 78.

Salaki, R. J., Waworuntu, J. & Tangkawarow, I. R. H. T. (2016): Extract transformation loading from OLTP to OLAP data using pentaho data integration. In: *Proceedings of the International Conference on Innovation in Engineering and Vocational Education*, (doi: 10.1088/1757-899X/128/1/012020).

Schwaber, K. (1997): SCRUM Development Process. In: *Business Object Design and Implementation*, J. Sutherland, et al. (eds.), Springer, London, pp. 117-134, (doi: 10.1007/978-1-4471-0947-1_11).

Shibani, A., Knight, S. & Shum, S. B. (2020): Educator Perspectives on Learning Analytics in Classroom Practice. In: *The Internet and Higher Education* 46, (doi: 10.1016/j.iheduc.2020.100730).

Shimada, A., Konomi, S. & Ogata, H. (2018): Real-time learning analytics system for improvement of on-site lectures. In: *Interactive Technology and Smart Education* 15:4, pp. 314-331, (doi: 10.1108/ITSE-05-2018-0026).

Ukwuoma, C. C., Bo, C., Chikwendu, I. A. & Bondzie-Selby, E. (2019): Performance Analysis of Students Based on Data Mining Techniques: A Literature Review. In: *Procceedings of the 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, (doi: 10.1109/TIMES-iCON47539.2019.9024396).

Verbert, K., et al. (2013): Learning Analytics Dashboard Applications. In: *American Behavioral Scientist* 57:10, pp. 1500-1509, (doi: 10.1177/0002764213479363).

Verbert, K., et al. (2020): Learning analytics dashboards: the past, the present and the future. In: *Proceedings of the 10th International Conference on Learning Analytics & Knowledge,* Frankfurt, Germany, pp. 35-40, (doi: 10.1145/3375462.3375504).

Viberg, O., Hatakka, M., Bälter, O. & Mavroudi, A. (2018): The current landscape of learning analytics in higher education. In: *Computers in Human Behavior* 89, pp. 98-110, (doi: 10.1016/J.CHB.2018.07.027).

Webber, K. L., and Zheng, H. Y. (2020). *Big Data on Campus*: *Data Analytics and Decision Making in Higher Education*, Johns Hopkins University Press.

Zhang, Y. & Wildemuth, B. M. (2009): Unstructured Interviews. In: *Applications of Social Research Methods to Questions in Information and Library Science*, B. M. Wildemuth (ed.), Libraries Unlimited, pp. 222-231.

Appendix 1

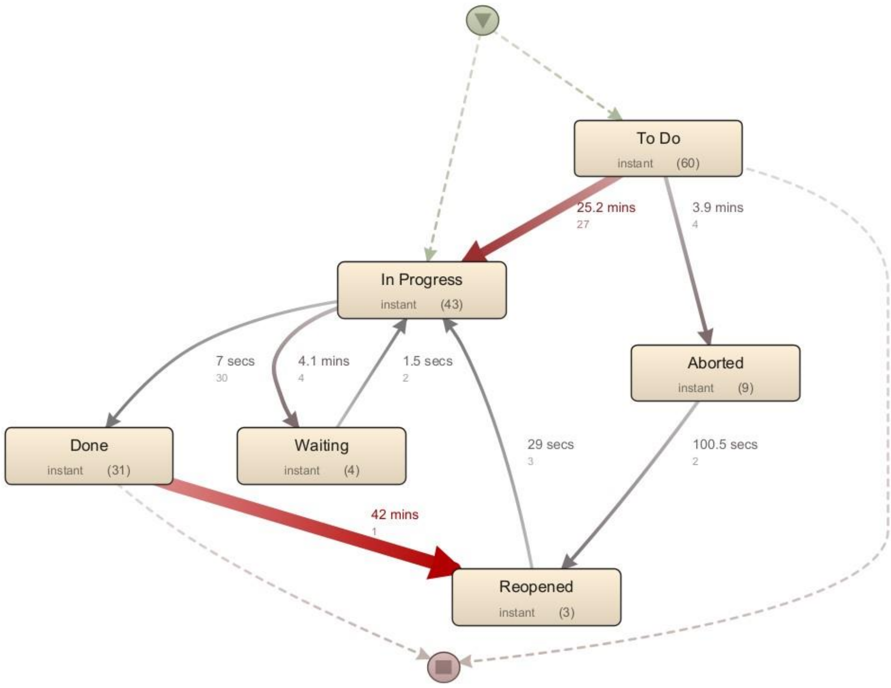

Figure 6: PM Graph of all Jira Ticket Transitions

Figure 6 exemplifies the PM graph created with the tool Disco by analysing all cases of Jira ticket transitions. The nodes represent the status used by the students during their work, while the edges show the direction, frequency, and duration of transitions. These graphs can offer valuable information to lecturers, including status utilisation effectiveness, transition duration plausibility, and potential bottlenecks in ticket creation and completion.

Copyright