Scaling up Learning Analytics in Blended Learning Scenarios

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Master Informatik
Vlatko Lukarov

aus Podaresh, Mazedonien

Berichter: Professor Dr.-Ing. Ulrik Schroeder
Professor Dr. Katrien Verbert

Tag der mündlichen Prüfung: 17.05.2019

ABSTRACT

In recent years, there is a prominent claim that learning analytics is a key transformative action that will radically transform education and its processes. This field draws its roots and methods from data analysis, statistics, data mining, business intelligence, computer science, and educational research and learning. Extensive research has been done to develop tools, prototypes, and analyze educational data to improve and innovate education, and this has advanced the research field of learning analytics. However, this has created a widening gap between what could the role of learning analytics be in education, and what learning analytics is actually doing in education. The research evidence shows that the use of learning analytics to improve learning and education is still in its infancy, and there is a lack of practical examples and implementations on scale and practical approaches of how to provide learning analytics services in education and put them into practice.

This dissertation focuses on the practical problem of scaling up learning analytics services in blended learning scenarios in a higher education institution in Germany. This dissertation presents the solution as a set of key principles for scaling up learning analytics in blended learning scenarios in higher education. These focus on five aspects: collecting correct requirements for the different stakeholder groups, preparing the legal and technical foundations of the higher education institution, continuously develop and improve the learning analytics services, and continuously evaluate the learning analytics services. These aspects were comprehensively investigated and realized by applying design-based research methods, software engineering methods, and evaluation methods from the Human-Computer Interaction field and the behavioral and cognitive sciences. The results and contributions from this research work are a verified end-to-end process for scaling up learning analytics in a higher education institution in Germany, a comprehensive set of requirements for the stakeholder groups, a categorized and comprehensive set of learning analytics indicators from the research and practitioners community, a sustainable learning analytics infrastructure with optimized analytics engine for scalability and performance and high fidelity prototypes, and a validated method for longitudinal studies for learning analytics impact evaluation. The basic idea is to enable another development team, or institutions from Germany to take this dissertation, its guidelines and results and use it to scale up learning analytics services at their higher education institutions.

ZUSAMMENFASSUNG

In den letzten Jahren gibt es die Behauptung, dass Learning Analytics eine zentrale transformative Maßnahme ist, die die Bildung und die Lernprozesse radikal verändern wird. Learning Analytics hat seine Wurzeln und Methoden aus den Bereichen Datenanalyse, Statistik, Data Mining, Business Intelligence, Informatik sowie Bildungsforschung und -lernen. Umfangreiche Forschungsarbeiten wurden durchgeführt, um Werkzeuge, Prototypen und Analysedaten zu entwickeln, um die Bildung zu verbessern und zu innovieren, und dies hat das Forschungsfeld der Learning Analytics vorangetrieben. Dies hat jedoch eine wachsende Diskrepanz zwischen der Rolle der Learning Analytics in der Bildung und der Rolle der Learning Analytics in der Bildung geschaffen. Die Forschungsergebnisse zeigen, dass die Nutzung der Lernanalytik zur Verbesserung von Lernen und Bildung noch in den Anfängen steckt, und es mangelt an praktischen Beispielen und Implementierungen in Bezug auf Umfang und praktische Ansätze, wie man Learning Analytics Dienste in der Bildung anbieten und in die Praxis umsetzen kann.

Diese Dissertation konzentriert sich auf das praktische Problem der Skalierung von Learning Analytics Dienste in Blended Learning Szenarien an einer Hochschule in Deutschland. Diese Dissertation stellt die Lösung als eine Sammlung von Grundprinzipien für die Skalierung der Learning Analytics in Blended Learning Szenarien in der Hochschulbildung dar. Diese Prinzipien konzentrieren sich auf fünf Aspekte: die Erhebung der korrekten Anforderungen für die verschiedenen Benutzergruppen, die Vorbereitung der rechtlichen und technischen Grundlagen der Hochschule, die kontinuierliche Entwicklung und Verbesserung der Learning Analytics Dienste und die kontinuierliche Bewertung der Learning Analytics Dienste. Diese Aspekte wurden durch den Einsatz von Design-based Forschungsmethoden, Software-Engineering-Methoden und Evaluierungsmethoden aus dem Bereich der Mensch-Computer-Interaktion sowie der Verhaltens- und Kognitionswissenschaften umfassend untersucht und realisiert. Die Ergebnisse und Beiträge dieser Dissertation sind ein verifizierter End-to-End-Prozess zur Erweiterung der Learning Analytics an einer Hochschule in Deutschland, ein umfassender Anforderungskatalog für die Stakeholdergruppen, eine nachhaltige Learning Analytics Infrastruktur mit High-Fidelity Prototypen und eine validierte Methode für Längsschnittstudien zur Wirkungsanalyse von Lernanalytik. Die Grundidee ist es, anderen Dienstleister oder Institutionen aus Deutschland die Möglichkeit zu geben, diese Dissertation, ihre Richtlinien und Ergebnisse anzunehmen und damit die Dienstleistungen der Learning Analytics an ihrer Hochschule zu erweitern.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Learning Analytics (LA) provides methods for detecting and analyzing patterns within educational data and uses them to support and improve the learning experience. In 2013 in the Horizon Report, this research field was accredited with high relevance and potential to develop. In the past years, the research field of learning analytics has been steadily growing and asserting itself as an independent field with practice-oriented effects (Consortium & Initiative, 2013). In 2016 the Horizon report on higher education predicted the time of adoption for Learning Analytics (LA) is one year or less. Higher Education Institutions (HEI) have access to a wide array of tools and versatile datasets to adapt and personalize the learning experience (NMC Horizon Report, 2016).

There are several important driving factors behind learning analytics. The first one is data availability of learning/learner data. Analytics in its essence is a data-driven approach. There are vast amounts of available educational data collected by learning platforms in higher education institutions. These learning platforms or learning management systems are most of the time used in so-called blended learning scenarios to augment the face-to-face teaching and learning methods, or in some specific cases, to support and enable distance learning. Moreover, the exponential growth of technology with regards to content creation tools have resulted in vast amounts of user-generated content across different learning environments, applications, and devices. Analyzing this data can shed light to unseen behavior, provide visibility to pieces of information and insight that could not be seen before, and would go unnoticed and be unactionable (A. L. Dyckhoff, Lukarov, Muslim, Chatti, & Schroeder, 2013). The second key factor is improving the teaching and learning experience. The introduction of online learning platforms supplies many benefits, for both parties, but online learning is not faultless. Technical problems, isolation, loss of motivation, constantly changing interfaces and technologies are among the common problems that students face. On the other hand, the teachers have difficulties assessing students' motivation and involvement in their learning, they do not know whether their pedagogical practices have positive influence on the students' learning, nor they receive any feedback about the quality and utility of the learning resources they provide (Lukarov, Chatti, & Schroeder, 2015).

Despite numerous and extensive advances in the research field of learning analytics, wide adoption, and successful implementations of learning analytics as a service is still not present. The added value of LA for learners and educators is clearly recognized and identified, but there has been little research done to provide conclusive evidence that the LA tools have desirable effects on the learning processes (Scheffel, Drachsler, Stoyanov, & Specht, 2014). Learning Analytics implementation as a service in Germany is a challenge. Strict data privacy laws make it exceedingly difficult to access even anonymized log data from the learning platforms and impossible to access highly personal data. Furthermore, the lack of institutional strategies and vision about using learning analytics to leverage the learning processes and experiences has an adverse effect on the learning analytics development and application in the same institutions. To conclude, every stakeholder group involved in the e-learning processes in higher education has its own expectations, goals, and understandings about what analytics is, and how should analytics be implemented. Therefore, this dissertation carefully investigates the issue of scaling up learning analytics for institutional adoption as an integral service to improve the learning processes. The idea is to supply directions and guidance about how learning analytics services can be designed, implemented, and integrated into higher educational practices.

Providing analytics as a service encompasses research and work in several areas that need to be investigated, prepared, and accomplished. Primarily, the institution needs to develop strengths and practices to provide support use of learning analytics; define the contexts in which analytics should be used and create and impose ethical standards including data and privacy protection. In turn, by creating such environment, the departments responsible for providing analytics as a service must involve the stakeholders during the entire process of development, so that the resulting analytics will have useful features and will be relevant to the context (Ferguson et al., 2016).

Second, the provided analytics results should be aligned with the applied pedagogical methods and practices within the institution. The analytics results should use and extend them and be aligned with the formative and summative assessment practices. Evaluation methodologies and quality assurance policies and practices should be developed to provide seal-of-approval and to ensure the validity and utility of the analytics tools (Ferguson et al., 2016).

Lastly, the institution and the departments responsible must build the knowledge, abilities, and infrastructure needed for learning analytics. To achieve this, they need to identify the necessary skills for analytics (technical, educational, promotional); provide training and instruction for the educators to use analytics in their day-to-day work; provide resources and possibilities for research and development in this field; build analytics ecosystem by providing enough technical resources and hardware infrastructures that can enable research and development of learning analytics on small and large scale (Ferguson et al., 2016).

This dissertation took upon these research and work areas and investigated how learning analytics services and tools can be implemented at large scale to be accepted and having proven (identified and confirmed) added value that is convincing and substantiated for the involved stakeholders while conforming to legal and institutional rules and regulations. The dissertation work covers the user involvement and collecting requirements relevant to learning analytics with well-established requirements acquisition and collection approaches. In this aspect, an innovation management technique, Outcome Driven Innovation, was also applied. Furthermore, it covers the effort and invested work for institutional preparation for the introduction of analytics in the learning processes from the legal, technical, and practical aspects. The technical aspects and implementation cover the design and development of an analytics infrastructure that uses learning generated data to supply analytics in every learning scenario supported by the learning environment and applications. This infrastructure also uses the same data to provide analytics and actionable intelligence to the administration and provide support in the decision-making processes regarding the e-learning initiatives and activities in the same higher education institution. The learning analytics infrastructure and prototypes were fielded and released in real-world learning scenarios and reached many users through two pilot phases and evaluation cycles, to provide an overview of the situation in those courses and learning scenarios and reach out to a wide audience. This helped in evaluating the analytics infrastructure and validating the application's integration and utility in the learning processes. This set-up provided invaluable feedback, knowledge, and conclusions to create proper design rules and guidelines for scaling up learning analytics on an institutional level.

This chapter goes as follows: Section 1.1 proceeds further with the motivation about providing learning analytics in higher education institutions. Section 1.2 summarizes the salient points of the research goals and objectives of this dissertation and defines the research questions. Section 1.3 provides an overview of the major technical and scientific contributions to the research community, and section 1.4 provides the skeleton and structure of this dissertation.

1.1 Motivation

Centralized learning platforms are the backbone for learning in institutions for higher education. They provide access for teaching staff and students to learning resources, information, and communication tools to support learning outside the university boundaries and face-to-face learning. These centralized learning platforms simplify the organization and communication between the teaching staff and the students; provide persistent storage and authoring tools to support independent and personalized learning; provide opportunities for cooperation, collaboration among students; enhance the quality, range, and diversity of learning resources provided by the teaching staff.

The underlying platform on which this dissertation work is based is the L²P Learning platform at RWTH Aachen University. The platform is developed in-house to support the students and teaching staff in different blended learning scenarios at RWTH Aachen University. Each semester around 3.000 courses are created and managed on the learning platform. On daily basis, there are from 18 – 32 000 unique clients (users), and 1.5 – 2.5 million requests, while on weekends the numbers drop to 8 – 10 000 unique clients, and 0.5 – 1.2 million requests. Around 25 percent of the usage comes from mobile devices. Based on these numbers, one can conclude that RWTH Aachen University has a successful implementation of blended learning scenarios. However, this is a general statement and it should not be taken at face value. Instead, concrete measures should be taken to get a better understanding of the educational and learning processes, especially for the two core activities: teaching and learning. They are dynamic activities in an ever-changing context and environment.

In blended learning scenarios, the teacher still holds the leading role which influences student learning and motivation. Teaching is an acquired mastery, and educators should use all means at their disposal to provide the most optimal learning setting and resources for their students. In other words, teachers should design proper pedagogical approaches and learning designs; create suitable and diversified learning resources; incorporate and carry out assessment within their pedagogical approaches; provide timely and appropriate feedback back to the students and be aware of student engagement in the learning process. They need to receive feedback and information about their teaching, analyze and reflect upon their work if they want to enact continuous enhancement and adaptation of their teaching. Furthermore, teachers need to be aware of how students perceive and behave in the learning setting on the learning platform. This way they can identify which learning resources work well; discover with which resources or assignments students struggle; identify and categorize learning patterns and strategies; understand which platform features are more effective within their pedagogical approach and setting; be able to identify and adapt materials to address the students' needs; and guide the students to become successful and better learners. These teachers' activities are well within the scope of teaching are crucial for providing high-quality education. Nevertheless, most learning platforms still do not supply such analytics features and support teachers to improve their teaching practices.

Student learners in higher education are a diverse group of young people. In the context of a single course, they have different age, gender, the field of study, (intrinsic) motivation, goals, cognition, skillset and earlier knowledge and experiences. These factors influence the ways students learn, grasp the presented material, and adopt new knowledge and concepts. In this learning environment, there is no teaching method that is equally well suited for all learners (Anna Lea Dyckhoff, 2014). However, all these factors and differences should not be considered as hindrances, but rather as opportunities. Experienced teachers will try to adapt their teaching methods and resources, adjust the course plan and learning design according to the target group. To do that, teachers must have a comprehensive insight into the specific behaviors of specific students (or student groups), they need to create informed conclusions about them, and act according to their conclusions. Still, the responsibility of student success does not lay only in the hands of the teachers. Students should be interested and intrinsically motivated to succeed in their tertiary education. They must take control of their own learning, and regulate it on their own pace, skills, and goals. Intermediate student goals and skills may vary across different student or student groups, but they share the goal of (successfully) finishing their studies and move on to the job market as trained professionals. Students need to have information about how they are progressing and what do they need to do to achieve their educational goals. In addition, they need guidance and feedback to become better learners, information about relevant learning materials and resources, and what effect these resources have on the matter at hand. Students also need to be aware of themselves in relation to the course, or peers, to assess and reflect on their current situation and help them make informed choices about achieving their goals.

Higher education institutions must support their teaching staff and their learners with providing the best conditions and prerequisites for educational success. Hence, the institutions need to make well-informed strategic decisions based on hard data to create a sustainable impact on their work and the e-learning scenarios. These institutional decisions are too important and critical to be based on intuitions, or presumptions. These decisions need facts, knowledge, and analytics. In theory, data and data analytics should be the main sources of information in decision making, standing for an improvement over intuition. Unfortunately, in practice, this is not the case. Often, decision making at higher level is usually based on intuition, presumption, and on accumulated experience, without any specific data or analysis (Campbell, DeBlois, & Oblinger, 2007). When it comes to institutional decisions and decision-making in higher education institutions concerning the learning and teaching processes, there are three main groups of stakeholders and decision-makers: students and faculty, and administrative bodies. In the administrative bodies there are students and faculty members that belong to different administrative bodies, but in this case, they have separate roles. At RWTH Aachen University, this group includes (but is not limited to) the blended learning coordinators[1] (one for each faculty), the committee for e-learning services (consisting of all BL coordinators, plus the management leadership from the institutes responsible for e-learning processes and implementation), the rectorate and dean(s) responsible for teaching, academic affairs, planning, development, control and institutional research, and the center for teaching and learning (including the developers' team) (Lukarov & Schroeder, 2017b).

Each of these institution stakeholders has different responsibilities, perspectives, and goals concerning their provision and support for the e-learning process and scenarios. Since most universities and higher education institutions (HEI) in Germany are public institutions, this means there should be a certain level of transparency and accountability to the public not only of the decisions but also of the decision-making processes. The funding and support they receive should be well-managed and distributed appropriately to guarantee a reach to a broad segment of students

[1] Blended Learning Coordinators are responsible for implementing the blended learning initiative at RWTH Aachen University. They are part of a rectorate appointed committee whose goal is to oversee the university-wide implementation of blended learning concepts.

and teaching staff. Since resources are finite, these decisions and activities should work towards optimizing resources on large scale, while containing and reducing costs by improving the productivity of the teaching staff and faculties. They should also demonstrate tangible effectiveness and status to the general public by ensuring that students graduate on time by having better course-level performance, which in turn can attract more new students. Hence, such position affords the higher education institutions to evaluate, revise, and improve the tools and assets they use for their decision making in order to stay on top and be competitive with regards to other universities (Yanosky & Arroway, 2015).

The research field of learning analytics undertakes to address these opportunities and demands in the educational processes. If one closely investigates different definitions of learning analytics, all of them emphasize on converting available educational data into useful actions to foster learning (M. A. Chatti, Dyckhoff, Schroeder, & Thüs, 2012). For example, Siemens et al. (Siemens, 2010) describe analytics as: "*use of intelligent data, learner- intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning.*" The Society for Learning Analytics Research (SoLAR) defines Learning Analytics in a similar fashion: "*Learning analytics is the measurement, collection, analysis, and reporting of data about learners and their context, for purposes of understanding and optimizing learning and the environments in which it occurs*." (Lockyer & Dawson, 2011) The Horizon Report of 2016 describes that learning analytics is an educational application of "**big data**" […] and that education is embarking on a pursuit to identify new ways to improve student engagement and provide a high-quality, personalized experience for learners (NMC Horizon Report, 2016).

The already conducted research work in the LA field provides vast opportunities and different analytics, metrics, and indicators. This variety of choices and possibility is overwhelming for the normal users and not all LA methods and visualizations are proper, meaningful, and understandable for all stakeholder groups. Moreover, aside from the research community and its most active members and practitioners, the bulk of the normal users and intended recipients of learning analytics do not actively participate much in providing feedback, knowledge, their expectations from learning analytics, or actively request such services. There is a well-known concept, or rule of the thumb called "90-9-1 Rule for Participation Inequality" from the Human-Computer Interaction field that can also be applied in the current situation (Nielsen, 2006; Whittaker, Terveen, Hill, & Cherny, 1998). Based on a study at Bell laboratories in the '90s (Whittaker, Terveen, Hill, & Cherny, 1998), Nielsen (2006) claimed that "*in most online communities, 90% of the users are lurkers who never contribute, 9% of the users contribute a little, and 1% of the users account for almost all the action.*". The aim of learning analytics tools and services is to provide insights, answers, and useful information for the users in supporting their work, questions, inquiries, and goals in their given context and the users should be steered towards the knowledge represented in these tools. However, when only using feedback and information from one particular group (in this case, researchers, early adopters and practitioners), this information is not representative of the average users for whom these analytics tools are intended. Nielsen (2006) asserts that if one looks for web postings for customer feedback, this person would be getting an unrepresentative sample, and one would almost always hear from the 1% of users who usually certainly differ from the 90% users (from which one almost never hears anything). Normally, it is the job of the service providers to identify, collect, analyze these questions, transform the collected data into clear and understandable data representations and visualizations, and provide them as answers to the respective users. The intention is through the work of this dissertation to build a strategy that covers the need of the majority, while building institutional and technical mechanisms for viable solutions for the more 'advanced' users.

1.2 Objectives

Learning Analytics as a service is still not available in higher education institutions across Europe which use blended learning scenarios within their educational goals, despite its potentials and perceived usefulness. This dissertation is a research project which aims to provide an end-to-end process of implementing learning analytics as an integral service for blended learning scenarios in higher education. Additionally, the research project will develop evaluation strategies to provide empirical evidence and data that learning analytics, in fact, has effects and impact when applied in higher education scenarios. The central research questions of this dissertation are enclosed around the theme of what kind of effects and influence learning (visual) analytics has on the stakeholders in higher education. The concrete question around which the work is centered is: *How can learning analytics services be provided on scale in blended learning scenarios?* The work in the dissertation was split into several integral parts, and each part covered a different aspect of implementing learning analytics on an institutional level. Each part was also enclosed around a specific research question, and all of them together are closely related to the central research question.

The parts/sub-questions that I have identified are the following:

- How can one capture the needs of the stakeholders, in our case, students, teaching staff, and the administration, who are directly involved in the learning processes?
- Where and how should analytics tools, mechanisms, and results be provided in the learning processes?
- What kind of empirical tools and research methods are most suitable for evaluating Learning Analytics in terms of usability, usefulness, utility, and most importantly, effectiveness and impact?
- What kind of practical data management strategies are necessary to provide analytics as a service, while conforming to data privacy laws and regulations?
- What kind of acceptance strategies are necessary to introduce Learning Analytics as an integral service for the stakeholders in higher education learning scenarios?

The basic idea of this dissertation is to provide a well-argued and comprehensive answer to all of them by applying different research methods and proven technical strategies and provide a verified practical process for scaling up learning analytics in blended learning scenarios in higher education institution in Germany.

1.3 Contributions

The contributions of this dissertation to the field of learning analytics are the following:

- *Verified end-to-end process for scaling up learning analytics in a higher education institution in Germany*. Another development team, or institution from Germany, or the European Union can take this dissertation, its guidelines, and results and use it to scale up learning analytics services at their higher education institution.
- *A validated and comprehensive set of learning analytics requirements for the four identified stakeholder groups of learning analytics*. The high-quality requirements were collected with methods from the software engineering field and can be used both by researchers and practitioners for implementing new learning analytics tools and indicators to further advance the research field of learning analytics. Many institutions or researchers have learning scenarios which call for specific didactical questions and learning analytics indicators for their research or application. The requirements catalog can help them find the right ones.

- *Sustainable learning analytics infrastructure*. The infrastructure is highly modular and developed with goals and standards to be used in a productive environment. This infrastructure can be taken as-is and deployed on productive infrastructure and servers and be activated in every course room on the learning platform at RWTH Aachen University. It was developed and tested for both scalability, stability, performance, and maintainability, with the goal of providing it on system scale. The development of the infrastructure encompassed the collection and definition of a comprehensive set of learning analytics indicators regarding their intended users and necessary data for implementations. A number of these collected indicators were used in the development and deployment of the "Insights" learning analytics prototype, as part of the learning analytics infrastructure.
- *A longitudinal study for impact evaluation*. A qualitative evaluation strategy with multiple data triangulation strategies was designed and conducted twice (once with 53 courses, and once with 400 courses) to evaluate and validate the development and provision of sustainable learning analytics service. The method was designed in a way to be easily reproduced with other tools and scenarios, and with appropriate methodological and research rigor to achieve solid evaluation results.

1.4 Dissertation outline

This dissertation is organized into nine chapters. Chapter one introduces the reader to the topic and provides the research questions.

Chapter two presents past experiences about institutions which have provided learning analytics at scale and their initiatives for supplying learning analytics tools to their students and teaching staff. The second part of the chapter shows experiences about collecting requirements, prototyping, and development techniques and evaluating learning analytics tools.

Chapter three presents the concept of design-based research method to achieve the goals of this dissertation and answer the research questions defined in section 1.2.

Chapter four presents the scientific foundations and terminology which will be used in this dissertation. I introduce concepts and results concerning blended learning, data privacy in higher education, analytics in education, and the research field of learning analytics and its evaluation. The second part of the chapter introduces the concept of analytics and its relevant building blocks in the context of higher education and it includes definitions of learning analytics, its processes, the concepts of learning dashboards and indicators, and the basic concepts of evaluation strategies for learning analytics.

Chapter five provides the three necessary elements for scaling up learning analytics. The first element is the comprehensive engineering of the requirements, the second part is the institutional preparation for scaling up learning analytics, and the third part is the outline for the technical implementation of a sustainable learning analytics infrastructure which includes a data warehouse, an analytics engine, and user interfaces visualizing the analytics results. The results of the three elements and the outline of the sustainable analytics infrastructure are used as the basis for discussions and the findings presented in the later chapters.

Chapter six presents the practical approaches for the implementation of the sustainable infrastructure, the management of learners' data, and the user interface evolution through prototyping. Moreover, this chapter also sketches out an implementation strategy for transferring the infrastructure towards other learning platforms.

Chapter seven presents two longitudinal case studies for evaluating the most completed prototype in a real deployment scenario, to assess its effectiveness in the real-world application in blended learning scenarios with the teaching staff stakeholder group. The presented results and findings of the two case studies showed that the deployed learning analytics prototype inspired awareness, reflection and encouraged some forms of action, and proved to be widely suitable for deploying as a learning analytics tool at scale in blended learning scenarios in higher education.

Chapter eight presents the important principles for scaling up learning analytics in blended learning scenarios and provides summarized answers to the research questions defined in section 1.2. Furthermore, it outlines the five aspects of scaling up learning analytics in blended learning scenarios: collecting the correct requirements, preparing the legal and technical foundations on an institutional level, continuously develop and improve the learning analytics services, and continuously evaluate the learning analytics services.

Chapter nine concludes the research and practical work in this dissertation and outlines the future outlooks and challenges for learning analytics in blended learning scenarios.

2 RELATED WORK

The goal of this dissertation is to investigate the scaling up of learning analytics for institutional adoption as an integral service to improve the learning processes. The idea is to supply directions and guidance about how learning analytics services can be designed, implemented, and integrated into higher educational practice. The following chapter focuses on salient points of selected research work from the learning analytics research field. In chapter 1, I acknowledged several work areas for research in the context of providing analytics as a service. This chapter's structure is based on these work areas, and each section has summaries of carefully chosen research projects and published experiences.

The first section focuses on successful implementations of learning analytics as a service in higher education institutions. Additionally, I explain the difference and experiences concerning large-scale analytics and small-scale analytics. The second section will outline how these institutions prepared themselves for adopting learning analytics. The institutional preparation has two main aspects. The first aspect of the preparation describes institutional decisions and legal undertakings to incorporate learning analytics as part of the learning processes in these higher education institutions. The second aspect outlines the practical preparations for learning analytics, such as building the technical infrastructure, building knowledge bases, and building accountable teams and institutional bodies. The third section covers research work and experiences concerning collecting requirements for learning analytics tools. These tools have different target groups, such as students, teachers, researchers, instructional designers, and other persons involved in the learning processes. It is essential to understand how their needs are captured, analyzed, and translated into LA tools. The fourth section covers practical research projects with prototyping and development of learning analytics tools. The focus here is to understand and show the methodologies and practices developers and researchers use to design and implement learning analytics prototypes and tools. The last section outlines research work about the evaluation of learning analytics tools. LA tools (as most interfaces) are generative artifacts. They do not have value in themselves, but rather they produce results in a context. This means that an LA tool is used by a specific user (e.g. student, teacher, mentor, researcher), on a particular data set, and for a particular reason. Hence, evaluating such tool is complicated and difficult. The evaluation research work will be observed from two perspectives: usability and user experience (UX) evaluation, and usefulness and utility evaluation (A. M. Chatti et al., 2014).

2.1 Successful implementations of learning analytics services

This section starts with presenting several Learning Analytics implementations at scale from higher education institutions across the world. In the second part, I provide different arguments about the differences between large-scale analytics and small-scale analytics.

2.1.1 Learning analytics practices: institutional at scale

Course Signals is an early intervention solution for higher education faculty members and teaching staff, allowing instructors the opportunity to use analytics to provide real-time feedback to a student. The development team had closely tracked the student experience from the introduction of the tool (the pilot phase), and at the end of every additional semester. Course Signals utilized all of the available data at Purdue University, including collecting data from the student interactions with the learning tools, to determine in real-time which students might be at risk, by assessing their effort with all of the available data (Arnold et al., 2012). The goals were to identify which students are falling behind, and thus provide "actionable intelligence" and guide them to appropriate help resources and explaining how to use them. These outcomes were delivered to the students in the form of an email from the faculty member, coupled with a specific color on a traffic light to show how each student is doing. The teaching staff has access to Course Signals through the faculty dashboard. There they have detailed information and feedback for every student and can use the dashboard to provide action-oriented feedback and helpful resources and guidance earlier in the semester to the students (Arnold et al., 2012). This way, the students can receive concrete steps to improve their learning behavior or actions to be successful. Course Signals have a user base of more than 40,000 students across 111 courses, including 140 instructors who actively use the system. Furthermore, they conducted an anonymous student survey to collect feedback at the end of each semester and had held focus groups. In general, students reported positive experiences with the feedback they received from Course Signals. The students felt supported and the feedback provided by the system was labeled as motivating (Arnold et al., 2012). Some students had concerns that the system did over penetration (e-mails, text messages, LMS messages) all of them conveying the same message. In general, the faculty and instructors had a positive response but still approached it with caution. The main points the development team extracted from the faculty feedback were that the students might create a dependency on the system, instead of developing their own learning traits. Furthermore, the evaluation discovered that both teachers and students lacked best practices how to use Course Signals. The most important point here was that this tool with its evaluation provided an actual impact on teaching and learning (Arnold et al., 2012).

Georgia State University (GSU) has taken a distributed approach for implementing learning analytics at scale. They did not develop a single application for analytics at scale but built a suitable environment on which they based their distributed programs utilizing analytics for increasing retention rate, student success, and graduation rates (Kurzweil & Stevens, 2015). The Institutional Research team of Georgia State has built a comprehensive data warehouse specifically designed for analysis and reporting. This data held the complete transactional student data from the year 2000 onwards, the complete student records and student activities on the educational systems at GSU. The administration developed a systematic problem-solving approach by analyzing students' academic pathways, problems, and barriers to their success. They developed interventions based on the data to address the identified problems and scaled them up to demonstrate evidence of effectiveness. They tried tackling the barriers among the student body from different perspectives, and with different approaches, thus creating a distributed learning analytics setting for intervention (Kurzweil & Stevens, 2015). This distributed setting also created a cross-functional organizational structure where different departments were organized into a single unit. They pulled together (with the help and support from the university leadership) registrars, student advising, admissions, financial aid, and student accounts into one, in order to facilitate the coordination and enactment of the student interventions. The distributed learning analytics setting encompassed the following student success initiatives. The Institutional Research team developed and deployed a graduation progression system which provided a dashboard that displays real-time analyses of student academic progress (Kurzweil & Stevens, 2015). This system raised alerts calling for intervention and prompted student advisor meetings, whenever it predicted that a student might have made a decision that implicates negatively on the student success progress. Another initiative was providing juniors and seniors with financial relief based on their academic success. They helped students who had small outstanding balances on their tuition and fees but were academically on-track to graduate from GSU. This small financial incentive initiative (coupled with academic and financial counseling) helped 4,200 students to successfully graduate from GSU (Kurzweil & Stevens, 2015). The third initiative was the complete redesign of gate-keeper introductory math courses (algebra, statistics, and pre-calculus). The courses were redesigned into a blended learning scenario with a hybrid approach. They used face-to-face instruction from the teaching staff, and machine-guided instruction. With this approach, they managed to decrease the failure rate from 43 percent to 19 percent and reached out to 7,500 students. GSU introduced supplemental instruction to students who had problems with specific subjects. The university provided work-study opportunities for tutors successful in these courses, in order to provide instruction and additional help to students for these courses. This way, the qualified tutors received financial benefit for providing tutelage, and the students received help and feedback from students who excelled in these courses. These are just four examples of initiatives at Georgia State University that contribute to excellent results through systematically accumulating smaller victories with distributed initiatives for student success via learning analytics (Kurzweil & Stevens, 2015).

Another university from the United States of America that has implemented learning analytics is the University of Michigan. Like Georgia State University, they also have several projects that implement learning analytics at scale. To foster different analytics projects, they have built their own data warehouse. In this data warehouse, they store student records, information about recruiting and admissions, the financial aid and student finances, and the college resources from the learning systems. The data from the learning systems include instructor and course data, all activities from the CTools and Canvas learning platform, all the activity data from the Kaltura video platform, and live lecture capture data. The highlighted projects that use the warehouse data are Student Explorer (Aguilar, Lonn, & Teasley, 2014), E^2Coach (McKay, Miller, & Tritz, 2012), and GradeCraft (Holman, Fishman, & Aguilar, 2013) and the next three paragraphs outline a summary of the three analytics applications. Overall, these three projects have influenced thousands of students in hundreds of introductory and intermediate courses at the University of Michigan.

Student Explorer is an early warning system designed for academic advisors who are novices in data-driven decisions and interventions. The tool is designed to track student effort and performance in STEM (Science, Technology, Engineering, Mathematics) learning communities (Aguilar et al., 2014; Krumm, Waddington, Teasley, & Lonn, 2014). The tool has an analytics dashboard available to the advisors that visualize data about student engagement and effort during the semester. The analytics dashboard provides weekly updates for every student in a quick and understandable way whether the student should be encouraged, explored in more details, or that the advisor needs to engage in an intervention because the student is lagging behind his peers (Aguilar et al., 2014; Krumm et al., 2014). The advisors' primary goal for using this tool is to provide timely prescriptive guidance for students. The conducted evaluation confirmed that the advisors used the Student Explorer during the meetings with their students, although the intention of the developers was that the advisors should use it before the meetings with students. The ultimate goal of the tool is to reduce time and effort in the "feedback loops" between learners, teachers, and academic advisors (Aguilar et al., 2014; Krumm et al., 2014).

The second tool developed at the University of Michigan is the E^2Coach (Expert Electronic Coaching system) intended for students in high-enrollment introductory science courses (Huberth, Chen, Tritz, & McKay, 2015). It also offers features for teachers, but the accent here is on supporting students by providing personalized feedback to the students directly in their online learning environment. The E^2Coach aims to provide personalized student support in a large environment STEM courses using their own Michigan Tailoring System. E^2Coach is a web application that uses tailored communication, course data, and experience, student data and analytics to deliver customized supportive messages to the students (Huberth et al., 2015). In the backend, the application develops student success predictive models based on feedback and experience from the coaching team and student information from the data warehouse. Each student has his own personalized dashboard upon registration for these courses, and they receive an overview and summary of their profile for the course, and a suggested To-Do list of suggested class activities. These activities are drawn from earlier experiences from previous students whose profile matches the current student. During the semester, the student receives tailored messages and visualized feedback about the coursework, activities, and their predicted grade. The tool provides evidence and support for thousands of students by personalizing their learning experience and improving their performance in the introductory STEM courses (Huberth et al., 2015).

The third analytics application is a learning platform that enhances "game-like" experience for students by leveraging it with analytics and providing features to explore coursework and plan individual learning pathways to success in a course (Holman et al., 2015, 2013). The teaching staff defines grading as a set of rubrics, categories and scoring guidelines while segmenting the coursework into different quantifiable parts that contribute to successful course completion. Additionally, each part could be completed with multiple sets of assignments, thus providing multiple routes and paths for course success. The students then have a choice to personalize their own learning experience by selecting and putting individual engagement, decisions, and weight effort to each of these rubrics, assignments, and coursework (Holman et al., 2015, 2013). GradeCraft is a web-based comprehensive dashboard that allows students to see their course performance in a single view. The dashboard shows up-to-date information and a summary of their performance, in relation to the course goals and criteria. Furthermore, there is a decomposed view of their performance and progress of each individual assignment and activity and have the possibility to model and alter their progress by selecting and modeling their engagement, and which effect would have these actions on their success in the course (Holman et al., 2015, 2013). These student interactions are compared with the course overarching goals, and students can compare their progress to these course goals. This affords the students to concentrate more on intermediate goals, skills, and mastery of skills, than just on grades. However, since final grades are awarded at the end of each course, the students can also see their predicted grade based on their work, engagement, and intermediate results. This modeling of student engagement is also important for the teaching staff and course designers to identify and understand how students design their learning experiences and value their learning process and decisions. This type of feedback focuses more on learning and learning processes because it focuses on how students design their learning processes, rather than reporting on their outcomes (Holman et al., 2015, 2013).

Nottingham Trent University in the UK developed, piloted, and deployed for all undergraduate students a student dashboard. The student dashboard attempts to help students in understanding how the students are engaging in their studies, compared with other students-peers of the same course (Ferguson et al., 2016). The dashboard uses engagement data from library use, attendance, use of online learning environments, use of university facilities, and academic performance (grades). Students and student advisors (tutors) can use the dashboard to quickly understand how well each student is engaging with their studies. The advisors can aid and give direction to help the students to progress and succeed in their courses. The dashboard gives a live update that the students can use to check their progress (Ferguson et al., 2016). The students remain in charge of their studies, but they can see and assess the risks and benefits of their engagement with their studies. The dashboard generates a composite engagement score and displays this score graphically, in relation with the average for everyone on the course, and provides a rating of high, good, average, or low. The system sends automatic alerts to the tutors for triggers of intervention whenever it detects inactivity, or low rating to check whether everything is all right. The goal of the university is to provide support for the development of skills and competencies for the subject knowledge and time management, organization, and taking responsibility for personal development. The student dashboard helps the student to take responsibility for their studies and developing skills for their professional engagement after the studies (Ferguson et al., 2016).

2.1.2 Learning analytics: Large-scale vs. Small-scale

The selected examples in the previous section are concerned with learning analytics on a large scale. However, there is learning analytics research that considers also the "small" scale implementation. These "small" scale implementations of learning analytics are important for the development of the field. Such analytics scenarios could be extended and adapted for large-scale implementation and adoption.

One major difference between small-scale and large-scale analytics is the support for different learning scenarios. Learning analytics on scale supports the well-established learning scenarios and course design on a course and faculty level. In these scenarios, the teaching staff is the driving factor of how a course develops, and the student activities and engagement are strongly correlated and dependent on the teaching activities. On the other hand, small-scale learning analytics scenarios are implemented in smaller courses that use innovative and experimental learning scenarios and tools. In such scenarios, the students are the driving factor with intrinsic motivation and their engagement is crucial for the implementation of such scenarios.

Another difference is the number of affected people by analytics implementations. For small-scale implementations, only a few selected courses with a small number of students or teachers are supported within their specific learning scenarios. On the other hand, by learning analytics at scale, thousands of students, faculty, and courses are engaged by analytics implementations. The scale at which has an influence on the design and implementation of learning analytics. Small-scale analytics implementation has intrinsic research perspective, while analytics at scale is considered a service. The research perspective gives more leeway when it comes to data collection and pre-processing. This means that a more versatile and varied data collection could be employed, which usually has more privacy intrusion of the students and the teaching staff. This is crucial for the process of discovering and identifying new knowledge and evidence about learning analytics, discovering new analytics indicators, and its effects on the e-learning processes. Such implementation on large scale, albeit technically possible, in most cases is not compliant with current data privacy laws. On the other hand, learning analytics at scale is more conservative in the approach of data collection but is data privacy conformant. In turn, this influences the organization of the collected data within the data storage (or data warehouse). The data models and analysis methods for small-scale analytics must be extensible and user-oriented, which usually has profound adverse effects on performance and scalability. However, these two properties are critical for large-scale implementations. Learning analytics services should be fast, responsive, and able to scale, aggregate, and analyze large amounts of data on daily basis.

2.2 Preparing institutions for learning analytics

Learning Analytics must scale and become institutionalized at multiple levels throughout the educational system. So far, most of the research results are course-based, small-scale, or tool-centric. In order to scale up analytics, institutions must invest significant resources and develop system-wide strategies in order to grow and foster institutional capacities and analytics services (Arnold et al., 2014). The underlying subject in this section revolves around which institutional capacities are essential for successful implementation and adoption of learning analytics. Analytics as a service in higher education is a complex topic which encapsulates social and cultural aspects, on one side, and technological infrastructure on the other side. The preparation of an institution for learning analytics can be divided into three parts (which could be addressed simultaneously): 1) technological infrastructure, analytics tools, and applications; 2) policies, processes, practices, and workflows; 3) analytics skills and values. In the following sections, I recap what institutions have used as underlying strategies to prepare themselves for learning analytics. Afterward, I provide examples from higher education institutions that worked on developing the foundations for learning analytics introduction.

2.2.1 Infrastructure, analytics tools and applications

Technical infrastructure is the core of providing analytics as a service. The infrastructure provides methods and procedures for data collection, storage, processing, and access to raw and processed data (Arnold et al., 2014; Ferguson, Macfadyen, et al., 2014). The technical infrastructure must follow the appropriate IT rules, regulations and protocols of the given institution, and with all the local, state, and federal legislation concerning with data security and privacy. The legislative rules and regulations usually have a restrictive character that can hamper the development and implementation of analytics services, but in this case, compliance is more important than the desire to explore and innovate new analytics solutions. Additionally, when building the technical infrastructure, the institution should build auditing processes that oversee and control the data intake, the data analysis, and transformation methods, and the level of access for raw data, processed data, derived data, and analytics (Arnold et al., 2014). Furthermore, there should be built-in oversight procedures for domain experts who can inspect the entire analytics process, inspect data integrity, and ensure legal compliance of the analytics infrastructure and services. The technical infrastructure is the foundation on which analytics tools are built. The analytics capabilities and data analysis are built as a core feature of the infrastructure, the tools are the ones that deliver the analytics result to the end users (Arnold et al., 2014). Moreover, the analytics tools can extend the analytics capabilities and provide additional features to the end users/stakeholders. For large scale adoption, one must develop an extensive set of requirements, and define evaluation and validation processes. In some cases, one tool fits all approach might be a poor choice, and instead, institutions can support and create a various set of analytics tools to support the learning processes (Arnold et al., 2014; Ferguson, Macfadyen, et al., 2014).

2.2.2 Building analytics policies, processes, practices, and workflows

The institutional and legal basis of building the technical infrastructure is provided by the governing bodies of the higher education institutions. The governing bodies must develop learning analytic policies and legal basis, provide financial and logistical support for building practices, processes, and project management to foster institutional change and to build sustainable learning analytics services (Arnold et al., 2014; Ferguson, Macfadyen, et al., 2014). Primarily, the universities need to deal with the question of data access and ownership. Learning Analytics can exist on an institutional level, if and only if, this question is resolved. The already existing data within the different systems and services in a higher education institution are already covered by existing legislative and internal institutional policies, but these policies and regulations in most cases are not conclusive, nor comprehensive. Learning analytics uses student/teacher data, but also with its processes generates (or derives) new data and knowledge based on the raw data (Arnold et al., 2014). The governing bodies need to develop new policies, or at least adapt the current policies, to take analytics into consideration. These policies should be restrictive enough, to protect the rights and interests of the stakeholders, and permissive enough to foster analytics implementation and innovation within the learning processes. At any given university, there are many levels of governance, but basic rules and policies about analytics data stewardship and management can be the framework within the technical infrastructure is built, and institutionally strengthens the learning analytics services (Arnold et al., 2014).

Providing analytics as a service is a complex process. Building the technical infrastructure and the analytics tools require a lot of work from people with different skills, perspectives, and responsibilities. Furthermore, they need to cooperate with the people from the governing bodies of the university to ensure that the work is in line with the goals, expectations, and vision of the institution. This means that task completion is crucial and on time; integration and compatibility of the analytics services with the already existing university has to be guaranteed; the different responsible bodies and teams need to build channels of communication and workflows for providing milestone deliverables on time and within budget (Arnold et al., 2014; Ferguson, Macfadyen, et al., 2014).

2.2.3 Building a team with analytics skills and values

Building analytics services require highly skilled people. They need to have data analytics expertise, research, and evaluation competencies and be able to provide support and qualification for the involved stakeholders. One of the major obstacles to overcome is the lack of common understanding of analytics and data, and the definitions associated with raw user data, and derived analytics data (Arnold et al., 2014). This can be alleviated by people who understand e-learning and learning processes and understand data. Institutions that want to implement analytics services need domain experts who understand how data collection and analysis work, how data is stored, and most importantly, how can this data be interpreted, and which conclusions and information can be inferred from it (Arnold et al., 2014). Furthermore, these experts must understand the structure of the learning platforms, the didactical approaches and implemented learning scenarios, and all additional tools and techniques used within the courses. This ensures that the data analysis, calculations, visualizations, and other interfaces for delivering analytics data and interventions are appropriate and aligned with the established learning processes at the higher education institution (Arnold et al., 2014; Ferguson, Macfadyen, et al., 2014). It is imperative that the institution provides support for the creation of a repository and evaluation of the applied learning processes at course level, department level, faculty level, and institutional level. Such repository can reduce the effort of scaling up analytics by providing requirements for analytics, evaluation metrics for analytics, and provide opportunities for additional (new) data collection efforts which can provide new insight and new information about the learning processes (Arnold et al., 2014).

Key pedagogical questions and didactical approaches must be the driving force for analytics in higher education. The learning analytics infrastructure and tools are just the means to provide answers. Incorporating analytics in the day-to-day activities of the teaching staff and students is the ultimate success factor for scaling up analytics in higher education institutions (Arnold et al., 2014; Ferguson, Macfadyen, et al., 2014). The teaching staff and students are no analytics experts, nor data experts, and they should not be treated as such. However, for step-by-step introduction to learning analytics, teaching staff and students who are willing to experiment with learning data, statistics, (self-)reflect on their teaching based on data, and use knowledge from data for improvement, should have early access to the analytics services and tools (Arnold et al., 2014). Furthermore, their feedback needs to be taken into consideration for iterative tool improvement. They also need support in interpreting the data and information presented to them to avoid misinterpretation, or more severely, base inadequate or wrong interventions based on the data (Dyckhoff et al., 2012). On the other hand, students have the right to know how their data is being used, as well as, the benefits and risks that can arise due to data usage. In conclusion, data literacy is a major skill that both teaching staff and students need to master with the help of the experts (Arnold et al., 2014).

2.2.4 Institutions with learning analytics initiatives

The following section introduces initiatives from universities that have committed themselves to provide learning analytics as a core service at the institutional level.

The Open University in the UK is a distance-based learning institution that serves around 250,000 students with different study programs. The university tries to leverage its learning models with learning analytics to extend and improve the curriculum development and improve the learning experience for the students (MacNeill & Powell, 2012; Siemens, Dawson, & Lynch, 2013). The OU has a dedicated team of data experts within the institution, with a special mission to develop innovative use of technology to support open and distance learning. A university-wide analytics strategy group oversees and steers the different analytics prototypes, projects, and activities of the so-called “data wranglers”, requiring a clear pipeline for successful innovations to be mainstreamed into operations. The university automatically and routinely collects various amounts of data through its online systems (MacNeill & Powell, 2012; Siemens et al., 2013). Furthermore, the data teams combine this data with other data sources and generated information from the course design, assessment activities, and learning reports, and exit surveys at each end of a learning module. The data team built a data warehouse that brings together the different data types and sources, so they can be analyzed, processed, and inspected in a way that they provide different viewpoints and pieces of information of the same student experience (MacNeill & Powell, 2012; Siemens et al., 2013). The main points of improvement through analytics have been identified as follow:

- Facilitate more informed discussions about strategic direction and resource allocation for faculties;
- Identify good teaching and course design practices and share them;
- Use data to influence course design processes and identify which are the most useful design steps and activities within course design, so additional data sources for analytics could be identified.

The OU succeeded in providing faculties and institutes with retrospective analytics and is working on real-time predictive modeling, and closing the learning design/analytics loop (MacNeill & Powell, 2012; Siemens et al., 2013).

The University of Wisconsin is a large distributed university with 180,000 students in 26 different campuses spread across the state. The university has the vision to use learning analytics as a system to help improve the effectiveness of the university and the overall quality of the learning processes (MacNeill & Powell, 2012). The university leaders have supported projects that build sustainable and scalable institutional strengths. They have started building the capacity building for working with data and grounded organizational strategies to act on analytics. They commissioned a group of technical staff to develop a learning analytics readiness instrument to assess the organizational maturity of the institution and find and fill organizational gaps in the context of analytics (MacNeill & Powell, 2012). One key goal of the university is to build organizational capacity around the use of learner data to track and improve the learning and teaching processes. For this reason, they started several smaller analytics projects which work on a common dataset from the university’s records and their learning platform. Furthermore, they built a learning analytics community within the university, consisting of 80 members, who actively worked on scaling up analytics, by developing processes and tools that use contextual data to better target learners at risk, and personalized the learning experience in order to give the students more control over their learning (MacNeill & Powell, 2012).

The University of New South Wales developed a strategic plan to support the teaching staff to better interpret collected institutional data, in order to improve the staffs’ knowledge, and ultimately the services they provide to their students (MacNeill & Powell, 2012). The university developed an institutional data warehouse to store all the key data necessary to track student achievements and predict the at-risk students. This was done by developing user-friendly and flexible tools to analyze and interpret data and produce reports for the Student Services Team, which then can act upon these reports. Hence, the university effectively connected the data team with the academic services, and the teaching staff. This way the university built a systemic approach that identifies students at academic risk in the early stages in their courses, thus allowing for potential and opportunities for early advisement and timely intervention (MacNeill & Powell, 2012).

The University of Technology, Sydney in 2011 set a goal and a vision to be data intensive university (Ferguson, Clow, et al., 2014). For this purpose, the university administration developed an operational definition of how they are planning to be a data-intensive university. They pointed out that at a data-intensive university, the staff and students understand data, they can use it and reuse it regardless of its size, volume, and diversity, and they can store it, and develop analytical tools and methods to interpret it. In essence, this institutional strategy outlines using data to understand the learning environment, learning processes, identify, analyze, and solve issues and challenges, and most of all, provide opportunities to develop new knowledge (Ferguson, Clow, et al., 2014). This opens the way for the university to use learning analytics to improve the student learning and their learning experiences. Additionally, such a data-rich environment allows the teaching and research staff to access and manipulate data with ease, which in turn can enable them to think and act differently when designing their teaching and research practices. In such an environment the administration and the university governing bodies can identify opportunities and data to support their decision-making capabilities and improve business outcomes. The entire initiative was put in motion by the senior executives of the university, which gave the project the required administrative and financial support to bring together different people with different skills and expertise to work together on scaling up data and analytics (Ferguson, Clow, et al., 2014).

Sydney University of Technology established a strategy for implementation to develop institutional analytics (Ferguson, Clow, et al., 2014). They pledged to support institutional culture for analytics to ensure engagement from all administrative bodies and stakeholders through communication and governance. Additionally, they provided leadership and engaged the institutional leaders to provide funding and financial support for pilot projects in analytics; provide funding for building analytics expertise either by in-house development or through recruitment of critical staff; invest in the technical infrastructure of tools and applications that can host and provide the analytics services. The university identified that analytics literacy is crucial for scaling up analytics. Therefore, they developed courses and training to help the stakeholders develop capacities for understanding data, make judgments based on the presented data and its meaning, thus, to engage them in evidence-based decision making (Ferguson, Clow, et al., 2014).

This data-rich environment started many prototypes and pilot projects that addressed different learning analytics goals. The analytics staff developed a project that provides information and data that is used for interventions to decrease student attrition. Another project provided a very detailed understanding of factors that occur in subjects that have low pass rates (or high failure rates over time). Another ongoing project of the team is the development of a student personalized dashboard that provides information about their own studying and engagement in learning (Ferguson, Clow, et al., 2014).

The last example in this section is the University of Michigan’s strategy for learning analytics (Arnold et al., 2014). The senior staff and leadership designated learning analytics as an important organizational initiative to aid the university in coping with current and future challenges. For this purpose, the provost built a Learning Analytics Task Force to advise the senior leadership and administration of how to utilize learning analytics to improve the learning processes and overall university productivity. This task force included staff and representatives from different faculties, schools, and colleges under the roof of Michigan University. The task force has goals to explore information technology environment and systems on campus, fund and foster the best and most prospective analytics projects, and review the most effective metrics for learning and teaching (Arnold et al., 2014). As a base for these analytics projects, the university built a data warehouse which collects various kinds of learning and learner data in one place and provided this data to different learning analytics projects. The learning analytics network and the task force has provided many grants and support for research projects, created an official communications channel for organizational discussion of learning analytics, and accumulated a community that assists researchers, faculty, and others who would like to utilize analytics in teaching and learning (Arnold et al., 2014).

Preparing institutions for adopting analytics as an integral service is a multi-layered and complex undertaking. The lessons and experiences that can be learned from the selected examples can be summarized in the three categories. The informed institutional leadership has to use pedagogical questions as drivers and vision for analytics and provide comprehensive support for developing the key institutional analytics policies, processes, and workflows that can serve as a legal basis for scaling up analytics and providing it as a service. Additionally, they need to provide the necessary funding to build an analytics team with skills and competencies and give enough resources to build the data strategy, and design and develop the required analytics tools and services.

2.3 Collecting requirements for learning analytics

In the previous sections, I have presented a comprehensive summary of successful analytics initiatives and learning analytics projects at scale in higher education institutions. At the core of these initiatives and projects stand practical implementations of software tools and solutions, which support every technical aspect of them. In the next sections, I focus on experiences about the practical implementations of such tools and initiatives. I divided these experiences into three parts: the first part summarizes the set of techniques used to elicit, collect, and analyze requirements for building learning analytics applications and services; the second part summarizes experiences about applied techniques for prototyping and development of analytics tools, and the last part covers evaluation practices and experiences when evaluating and validating learning analytics tools and applications.

It is difficult to build an analytics solution if the development team does not know the requirements, understandings, expectations, and goals of the stakeholders. In software engineering, there are well-established techniques and methodologies for eliciting and gathering requirements, such as interviews, facilitated sessions and focus groups, joint application development, questionnaires and surveys, use cases and prototyping, and brainstorming sessions. Each of them has advantages in certain situation or scenario, but usually, multiple techniques are applied to get the needs and goals from a diverse set of stakeholders and clients. Researchers and developers from the research field of learning analytics have used different combinations of these techniques and have inferred conclusions and decisions for the design and implementations of their respective tools. In the following paragraph, I summarize different experiences concerning requirements elicitation and collection conducted by researchers and developers in the context of learning analytics.

One of the most used technique for requirements elicitation about the stakeholder needs is a literature review. In 2011 Dyckhoff (Anna L Dyckhoff, 2011) conducted a meta-analysis on different case studies publications on e-learning, looking for research questions by teachers who want to improve technology-enhanced learning, and what kind of methodologies these teachers apply to answer these research question. The author analyzed many publications with the goal to draw requirements for learning analytics and shape future learning analytics tools with easy-to-use and understandable interfaces (Anna L Dyckhoff, 2011). The authors collected 86 questions and statements and divided them into six categories according to the analytics methods and tools that are necessary to answer them. The author discovered that teachers already had various questions and concerns about their learning design, the usage of different learning offerings, the student behavior and engagement, and they wanted to see a correlation between teaching activities and student activities. The main result of the meta-analysis was a taxonomy of requirements which they used as insight for the development of analytics processes and tools aimed to facilitate awareness and reflection for improving the teaching and learning practices (Anna L Dyckhoff, 2011).

Surveys were also used to provide insight into learning analytics, level of understanding, goals, and expectations of the stakeholders. Drachsler and Greller (Drachsler, Greller, Greller Wolfgang, & Drachsler Hendrik, 2012) conducted a survey among higher education practitioners and researchers to support their theoretical framework for learning analytics based on a literature review. The design and implementation of the survey for collecting requirements, goals, and visions were designed with partiality to support the theoretical framework, rather than collect actual requirements for learning analytics implementations. Furthermore, the authors used the term "learning analytics" directly without explanations, which led to many subjective interpretations among the survey participants. Despite this, useful insights and knowledge about learning analytics were collected (Drachsler et al., 2012).

Researchers in a Mid-Atlantic university in the United States have conducted a user-centered, design-based qualitative study with students and instructors who engage in a first-year engineering program to investigate how they describe the potential of learning analytics approaches to contribute to student success (Knight, Brozina, & Novoselich, 2016). The researchers have conducted a three-hour session with eight students. The session included a two-hour semi-structured interview about student data usage and learning analytics, followed up with a brainstorming session and idea generation about data visualization, data streams, variables, and how they can be merged to create a dashboard to satisfy different users. The students were guided by the researchers, who also observed the design session, took notes and developed a protocol (Knight et al., 2016). To collect data and feedback from the instructors, the developer team conducted individual semi-structured interviews. The gist of the interviews was how the teaching staff could use data to analyze and reflect upon their teaching practices and students. The result of the study was a set of concrete perspectives, requirements, and statements with a focused disciplinary context for learning analytics. The researchers stated that user-focused approach can produce important insights regarding designing analytics solutions and data usage (Knight et al., 2016).

In a project in the UK, a group of researchers used a combination of surveys, brainstorming, and discussions on a workshop indented for designing a learning analytics apps for the students (Sclater, 2015). This workshop included staff from higher education institutions and some motivated students who were interested in learning analytics research and implementations. In advance of the workshop, the researchers had collected potential requirements in an anonymous survey, so that they could be used as a starting point for the workshop, and with discussions iterated and improved. This is a novel approach that actively uses student input and feedback for developing learning analytics solutions (Sclater, 2015). The collected potential requirements were divided into three groups by the researchers. The first area of requirements was based on information provision, the second group of requirements was concerned with action, and the third group was concerned with various issues of analytics such as the user interface, data ethics, and privacy. After the initial composition, grouping, and aggregation, another questionnaire was again run to the students to find out what they really want to see in an analytics app. This is a classic example of keeping the user in the loop from the initial stages of development of a learning analytics tool (Sclater, 2015).

Researchers at the University of Minnesota have used informal discussions, surveys, and focus groups to develop a data-browsing tool that can facilitate faculty implementations of evidence-based course design (Dunbar, Dingel, & Prat-Resina, 2014). They actively involved the faculty in the entire process, because user involvement was a prerequisite from the institutional policy, and to explore the types of questions faculty members were interested in exploring, and what kind of data they would need to explore those questions about course design. The survey and the focus group helped in facilitating interactions between faculty members (non-experts) and provided a sensible perspective on the goals for developing the analytics tool. Within the focus group, the faculty members presented pedagogical questions relevant for designing and evaluating their courses and coupled them with associated metrics and indicators (Dunbar et al., 2014). Additionally, within the focus group was discussed the physical location and accessibility of the data for each indicator. In the end, the faculty members produced a list of questions with the associated indicators and pointed out mechanisms that could allow them to statistically analyze and explore these questions with the data. In the end, they prioritized the questions, the associated indicators, the relevant data sources and the statistical tools and mechanism to build the first version of the learning analytics tool. The researchers started with the research questions, and then sought to identify how to browse, analyze, and employ the data in a way that allowed faculty to draw useful conclusions (Dunbar et al., 2014).

2.4 Prototyping and development of learning analytics tools

Learning Analytics tools are, in most cases, interactive systems with users who try to complete their tasks and goals within a given context. In the following section, I recap the different prototyping and development methodologies used by the researchers and developers to develop learning analytics solutions. When applied, these methods transform the requirements into a working design, which is a specific choice among many possible design options. In general, the design follows the collected requirements, but during the design process some of them must be questioned and changed, because, design is all about making choices. Researchers and developers create prototypes when working on a design, or a set of designs for an interactive system. Prototypes are a concrete representation of one part, or an entire interactive system. They are tangible artifacts, which unlike abstract descriptions, do not need interpretations. Prototypes serve different purposes and can have different forms, distinct levels of precision, different extents of interactivity, and lifecycle. As such, different prototypes and techniques are applied at various stages of the development of an interactive system. The same is valid for prototyping and development of learning analytics tools/applications/systems, at least to the parts of them who are visible to the users. Designing and implementing learning analytics applications is a multifaceted process that needs software engineering skills, knowledge, and experience with human-computer interaction, application of information visualization techniques, psychology, and educational design. This leads to different approaches and experiences in implementing learning analytics.

One approach for designing, implementing, and deploying learning analytics is to develop a framework that encompasses the interdisciplinary skills and techniques, which was based on the LATUX workflow (**L**earning **A**wareness **T**ools – **U**ser e**X**perience) developed at the University of Sydney (Martinez-Maldonado et al., 2015; Martinez-Maldonaldo et al., 2015). The workflow is a five-stage workflow to design, validate, and deploy awareness interfaces in technology-enhanced learning scenarios, based on well-established design processes for building user interfaces. The workflow tries to provide a holistic approach for the development of analytics by splitting the entire process into five stages: stage one - Problem identification; stage two - Low-fidelity prototyping; stage three - Higher-fidelity prototyping; stage four – Pilot studies: Classroom use; and stage five – Validation in-the-wild: Classroom use. In this section, I cover stage two and three: Low-fidelity and higher-fidelity prototyping. The researchers have defined a set of guiding questions that need to be considered and answered for each of these phases to streamline the completion of the stage. For the low-fidelity prototyping stage, the researchers are trying to find out which visualizations should be generated, how to generate the mock-ups and sketches, which data is needed for the actual visualizations, and how to evaluate these early sketches and mock-ups so that real usage is mimicked. The researchers sketch out low-fidelity prototypes with wireframe tools to test a wide range of visualizations that could be part of the analytics tool; they already conduct initial evaluation of the prototypes including sense-making of the data; and talk to the users before beginning of the technical implementation of the analytics tool (Martinez-Maldonado et al., 2015; Martinez-Maldonaldo et al., 2015). The authors state that low-fidelity prototyping can focus on either on the evaluation of the user interface or to the actual information conveyed in the visualizations (data-oriented prototypes). The main benefits of creating a low-fidelity prototype are experimenting and testing the usability and sense-making in the early stages of the development of learning analytics, testing with multiple people and multiple prototypes are possible (and in most cases advisable), and they are cheap (easy to create, and easy to discard). However, low-fidelity prototypes also have limitations. These prototypes cannot catch the process of intervention or change in learning behavior after feedback has been provided to the user. Insights into the real user experience and interactivity are limited because the prototype is outside of the real learning environment. The researchers also propose another way of collecting rapid feedback about user experience and interactivity through conducting the Wizard of Oz Studies (Martinez-Maldonado et al., 2015; Martinez-Maldonaldo et al., 2015).

For the higher-fidelity prototyping stage, the researchers propose the creation of at least partial working analytics backend with meaningful learner data. The guiding questions for streamlining the development of this stage are concerned with dynamic generation of visualizations; how to simulate conditions of the setting; how different data can be collected and processed; how interventions can be simulated; how to support different interactions; and identify methods and techniques to evaluate the impact of the analytics tools. High-fidelity prototypes require more time and effort, but they are a more detailed and realistic representation of the real interactive system. Still, they are cheaper than developing the final tool, and they require less effort and offer more flexibility for experimentation (Martinez-Maldonado et al., 2015; Martinez-Maldonaldo et al., 2015). The strengths of these types of prototypes lay in the inclusion of the time factor within the visualization of real (or simulated) analyzed student data to examine and experiment with the decision-making process on the fly. For the teaching staff, this means what they would decide to do after looking at the visualization and data, for the students means what would they do after they have received feedback and information about their engagement. In addition, the aspects of the user interface and interaction design (especially the data exploration part) can be evaluated and explored in detail. These prototypes can also be evaluated with multiple people, and a deeper evaluation of the tools and the interactions could provide a lot of high-quality feedback. On the downside, with such prototypes, interventions still cannot be tested, and they are more expensive and time-consuming to build in comparison with the paper prototypes. The results of several iterations of the low-fidelity and higher-fidelity prototyping stage is an analytics tool that can be deployed in a pilot study (Martinez-Maldonado et al., 2015; Martinez-Maldonaldo et al., 2015).

Another project that was developed with incremental and iterative approach was the exploratory learning analytics toolkit (eLAT) at RWTH Aachen University (Anna Lea Dyckhoff et al., 2012). The researchers have developed the tool within two main stages that overlapped to meet the requirements of the stakeholders. In the first stage they developed the analytics backend framework, and in the second stage of the development, they worked on the design and evaluation of the user interface (UI). For the first stage of the development, they collected various kinds of learning data and implemented an initial set of indicators in a rapid application prototyping strategy. In this manner application prototypes with real data was shown to the teaching staff for informal evaluation, feedback, and improvement. The design and evaluation of the interface in this project were carried out in three iterations. The first two iterations were conducted with paper prototypes to collect feedback about which data should be collected, and how this data should be visualized and organized on the dashboard of the analytics tool (Anna Lea Dyckhoff et al., 2012). The third iteration had a functional user interface based on the previous stages, and this interface was tested for its interactivity and usability aspects. The layout and data presentation were designed with HCI heuristics for interactive systems, cognitive walkthrough and pluralistic walkthrough were also conducted to discover problems early in the development. The pluralistic walkthrough was conducted as a meeting of experts from different domains, and they discussed different elements of the interface prototype according to the view of users. In the third iteration, the researchers evaluated the interactions with a qualitative think-aloud study. This way the researchers identified interaction pitfalls of the users. After the iterations, the result was an analytics tool that was deployed in a pilot study (Anna Lea Dyckhoff et al., 2012).

One challenging aspect of prototyping data visualizations is coordinating the expertise of people who have different roles and skills when creating data visualizations guided by an educational theory informed goal (Hillaire, Rappolt-Schlichtmann, & Ducharme, 2016). For this purpose, a workflow for prototyping visual learning analytics guided by an educational theory informed goal was developed to rethink the interdisciplinary approach of creating analytics prototypes. This way, more concentration is put on the educational goals, instead of following well-established best practices. Hence, defining the educational goal is the first step of this approach for analytics prototyping. This goal can be leveraged through the process of prototyping and development to ensure that all efforts are in check towards achieving it. The second step of the process is to define the target users who are affected by the tools and solutions from the educational goal. The prototyping comes in the third place, with a notable accent on interdisciplinary paper prototyping process. The basic idea of this process is to engage in several rounds of the group and individual idea generation and synthesis. The results of these individual and common brainstorming sessions are discussed, rated, and summarized so that sketching different solutions on paper are possible (Hillaire et al., 2016). Then, the participants are asked to quickly sketch out high-level ideas, and variations of them. The point is not to delve into details but keep sketches clean and the focus is on high-level ideas. Another approach is collaborative white-board sketches that can help converge and refine the different ideas. The main driving factor for combining and converting the different ideas and sketches is the realistic expectations out of the data visualizations. Therefore, a basic introduction/understanding of the available data is preferred, so that all participants can provide potential shapes of the educational data. From these individual and group data sketches, more detailed storyboards with success scenarios are shaped. For each screen (part of the storyboard) individual person is responsible, so each participant annotates the functionality, interactions, and navigation within and between screens, and the states to simulate the animation (change over time). The result is a complete storyboard and paper prototype ready for formative testing, and it can serve as a basis for wireframe creation or a software prototype when extended with mockup educational data (Hillaire et al., 2016).

The next step of the process is the implementation of the analytics tool. The first problem that needs to be addressed is about the chart types. The paper prototype evaluation should have provided enough feedback and insight about the known strengths and weaknesses about the charts/chart types for an analytics visualization. The second key point that needs to be addressed is the scale and the axis of the graph that clearly communicate information about the established learning goal. Here, the researchers warn that the data and its form might not be suitable for the visualization and scale at hand. Therefore, they provide a suggestion that the data might be reworked and reprocessed to fit the scale and visualization method. The third key point the researchers advise on is the color selection and the design of the interactions. Both must support the goal of the visualization while reducing the cognitive load of the user. The last point from the researchers is accessibility, so that everybody is able to understand and access the visualizations.

Many learning analytics provide the analytics results in the form of visual learning dashboards. This type of analytics provision can support data exploration by both teachers and learner (Charleer, Klerkx, & Duval, 2014). Researchers from KU Leuven have applied a design-based iterative user-centered research that focuses on users and empirical measurement via designed and implemented prototypes. They evaluated their prototypes following user-centered rapid prototyping approach by producing paper prototypes for initial feedback gathering on ideas. In later stages, they gradually developed their more functional digital prototypes (also in rapid iteration cycles). They tested their prototypes in a realistic setting with real users in inquiry-based learning environments, where users (students) generate a large number of digital artifacts. For evaluating the prototypes, they have used semi-structured interviews, activity tracking, awareness, (perceived) usefulness, and effectiveness (Charleer, Klerkx, et al., 2014).

2.5 Learning analytics evaluation research and experiences

Learning Analytics should support the stakeholders to achieve their goals. This section summarizes what has been done in terms of evaluation research and experiences to confirm the added value Learning Analytics tools bring into the learning processes. The added value of LA for learners and educators is clearly recognized and identified, but there has been little research done to provide conclusive evidence that the LA tools have desirable effects on the learning processes (Scheffel et al., 2014). Gašević et. al (Gaševi, Dawson, & Siemens, 2015) in a publication in 2015 focused on critical goals, topics, and aspects that require immediate attention in order for learning analytics to make a sustainable impact on the research and practice of teaching and learning. In their work, they provide a summary of critical points and discuss a growing set of issues that need to be addressed and strongly points out that learning analytics are about learning. One point they provide is that LA resources should be aligned to well-established research on effective instructional practice. To argue this point they point out that observations and analyses suggested that instructors preferred tools and features which offered insights into the learning processes and identified student gaps, rather than simple performance measures (Gaševi et al., 2015).

On the other hand, LA tools (as most interfaces) are generative artifacts: They do not have value in themselves, but rather they produce results in a particular context. This means that an LA tool is used by a user (e.g. student, teacher, mentor, researcher), on a particular data set, and for a particular reason. Hence, making the evaluation of such tool very complicated and difficult (A. M. Chatti et al., 2014). Irrespectively, the research field of learning analytics conclusive evidence that LA tools reach and support the learners, teachers, and researchers in achieving their goals. This is a major issue for justifying the investments and efforts in scaling up analytics in higher education institutions. The focus of analytics evaluation should be on the effectiveness and the tool's value; the tool's impact on the teaching/learning experience, and whether the tools support the users in achieving their goals in the context of learning. However, the focus on most conducted evaluations was on the user interface. To show this, I provide three different case studies that demonstrate different evaluations carried out on analytics tools.

LOCO-Analyst is a learning analytics tool that was developed to provide educators with feedback on the learning activities and performance of students (Ali, Hatala, Gašević, & Jovanović, 2012). The researchers have done the evaluation of their tool in two iterations. The first iteration they conducted was a formative evaluation aimed to investigate how educators perceive the usefulness of such a tool to help them improve the content in their courses, and to which extend the user interface of the tool affected this perceived value. Additionally, they used the evaluation as a chance to elicit additional requirements for the improvement of the tool. The study design was implemented with a focus on the collection of quantitative data and perceived qualitative data from a larger sample of educators. During the study, 18 participants from different higher education received the tool and a questionnaire with guidelines on how to evaluate the tool. The researchers analyzed and coded the results in three distinct categories: Data Visualization, GUI, and Feedback. The results of the first evaluation of the tool influenced the enhancement of the tool's data visualization, user interface, and supported feedback types (Ali et al., 2012).

The second example is the evaluation of Course Signals is an early intervention solution for higher education faculty, allowing instructors the opportunity to use analytics to provide real-time feedback to a student (Arnold et al., 2012). The development team had closely tracked the student experience from the introduction of the tool (the pilot phase), and at the end of every additional semester. Furthermore, they conducted an anonymous student survey to collect feedback at the end of each semester and had held focus groups. In general, students reported a positive experience with the feedback they received from Course Signals. The students felt supported, and the feedback provided by the system was labeled as motivating. Some students had concerns that the system did over penetration (e-mails, text messages, LMS messages) all of them conveying the same message (Arnold et al., 2012). In general, the faculty and instructors had a positive response but still approached it with caution. The main points the development team extracted from the faculty feedback were that the students might create a dependency on the system, instead of developing their own learning traits. Furthermore, the evaluation discovered that there was an apparent lack of best practices on how to use Course Signals. The students also confirmed this. The most important point here was that this tool with its evaluation provided an actual impact on teaching and learning (Arnold et al., 2012).

Student Activity Meter (SAM) is an LA tool that visualizes collected data from learning environments (Govaerts & Duval, 2012; Govaerts, Verbert, & Duval, 2011). The researches incorporated the evaluation in the development of the tool, i.e. applied design-based research method which relied on rapid prototyping to evaluate ideas in frequent short cycles. They did four iterations over the course of 24 months. The results of each evaluation iteration were put into two major groups: positive and negative. The results and the provided feedback were incorporated to improve the tool. The method for the first iteration was task-based interviews coupled with think-aloud strategy, and usage of the System Usability Scale, and Microsoft Desirability Toolkit on Computer Science students. The negative results were the identification of usability issues and points for improvement. The positive results revealed that learnability was high, the error rate was low. The user satisfaction and usability were decent, and preliminary usefulness was regarded as positive. The study also revealed which analytics indicators were considered as most useful (Govaerts & Duval, 2012; Govaerts et al., 2011).

The methodology for the second iteration was conducting an online survey with Likert items on teachers to assess and evaluate teacher needs, extract information about use and usefulness, and whether SAM can assist them in their everyday work (Govaerts & Duval, 2012). The most prominent result that was considered negative was that teachers did not find resource recommendation useful. On the positive side, the results showed that SAM provided awareness to the teachers, that all the indicators were useful, and that 90% of them wanted to continue using SAM. The third iteration was also an online survey with Likert items, but the demographics were LAK conference participants (teachers and visualization experts). The goal was also similar like in the second iteration, to assess the teacher needs, improve the use and usefulness, and enhanced to collect feedback from the experts in the field. The negative results of the evaluation were the failure to find which needs to be needed more attention. On the positive side, the results from the second iteration were strengthened, with the addition that resource recommendation could be useful. The fourth iteration was conducted with Computer Science teachers and teaching assistants. The methodology was conducting structured face-to-face interviews with tasks and open-ended questions with the goal to assess user satisfaction, the use, and usefulness of SAM (Govaerts & Duval, 2012).

The fourth iteration revealed that there are conflicting visions of what were students who were doing well, or what were students who were at risk. Furthermore, it revealed which indicators were good and useful, provided different insights from the teachers, and further use cases for SAM were discovered. Overall, the conducted evaluation discovered that the most important feature that SAM addresses were to help teachers supply feedback to the students. Another important provision was the methodology of the evaluation which could be applied/used from other researchers when creating a visualization tool (Govaerts & Duval, 2012).

Table 1. Learning analytics evaluation strategies

Evaluation Purpose	Evaluation Methods
Usability	Interviews System Usability Scale (SUS) Think aloud strategies Controlled trials Microsoft Desirability Toolkit Observation in the field Questionnaires
Usefulness/Perceived Usefulness	Controlled experiments Interviews Observation System Usability Scale (SUS) Focus groups Questionnaires Surveys
Impact/Potential Impact	Interviews Surveys Controlled trials Data analysis from the tool's usage logs Focus groups Observation in the field (iterative) System Usability Scale (SUS) Think aloud strategies
Exploration/Explorative study	Interviews Surveys Observation Controlled trials Focus groups Correlation of survey data with log data Exploration and insights searching Exploratory data analysis of the tool's usage logs Log data clustering Questionnaires System Usability Scale (SUS)
Support for Reflection	Interviews Surveys Controlled trials System Usability Scale (SUS)

The three different evaluation case studies show that there is no standardized approach on how to effectively evaluate learning analytics tools, and there are many reasons and purposes for conducting the evaluation. Measuring the impact and usefulness of LA tools is a particularly challenging task. Table 1 summarizes the different purposes for evaluating learning analytics tools, and what kind of evaluation methods were used in the research community to conduct the evaluations. This mapping of evaluation purpose and methodologies is based on a literature review from the proceedings of the LAK Conference on learning analytics and knowledge. Although there are different purposes evaluations, the evaluation methods are in-fact overlapping. This implores the question, whether these evaluation methods as such, provide valid results in different evaluation contexts, and whether the results and outcomes of the evaluations could be generalized. Overall, the most used evaluation methods for learning analytics are interviews which are present in all five subcategories, as can be seen from Table 1. For conducting *Exploration/Explorative study* the researchers have used interviews, surveys, and observations. The subcategory *Usefulness/Perceived Usefulness* include evaluation methods such as controlled experiments, interviews, observations, and System Usability Scale (SUS). The subcategories that contain evaluations concerning *Usability* and *Impact/Potential impact* contain very similar evaluation methods, such as interviews, System Usability Scale, controlled trials, which are usually associated with usability studies. Lastly, for the *Support for Reflection* subcategory, the evaluation methods are remarkably similar to the ones applied in the other subcategories, namely interviews, surveys, and controlled trials.

3 RESEARCH METHOD

I use design-based research as an underlying research method for this thesis. In this chapter, I present and explain the salient points of conducting design-based research on information systems in the context of technology-enhanced learning.

3.1 Design science

The leading goal of design science is to gain knowledge and understanding by creating new models and artifacts. Design-based research is conducted with a "build-and-evaluate" loop and each iteration results with better and improved models and artifacts. The main difference from routine design lies in the fact that it is not building on existing knowledge about problems, but it tries to address important unsolved problems in unique and innovative ways or provide more effective and efficient solutions to solved problems (Hevner & Chatterjee, 2010b, 2010a; Wang & Hannafin, 2005). Basically, design-based research refines both theory and practice, and its value of a theory is valued by the extent to which its principles inform and improve practice (Barab & Squire, 2004; Dyckhoff, 2014; Hevner & Chatterjee, 2010b, 2010a). I chose this research method because it applies well to the research problem described in chapter 1.

In the following paragraphs, I describe the main characteristics of design science problems: (1) context, (2) relationships between the problem and solution, (3) flexible design processes, and (4) cognitive and social abilities of the stakeholders to produce effective solutions within the context. The context of my research is higher education institutions with different blended learning scenarios.

1. The location of the research is in real-life settings where learning occurs and is by default complex because the learning scenarios depend on the people involved, their goals and expectations, and the available resources. Any of these variables change over time or every iteration. Additionally, scaling up analytics and introducing it as an integral part of the learning processes will change the learning context and the stakeholders' perceptions about learning. Therefore, it is essential to characterize the context and situation in all its complexity within the setting over a long period of time (Wang & Hannafin, 2005).
2. Design-based research methodologies involve multiple dependent context variables, outcome variables, and system variables, which are explicitly defined and influenced by the real-world context, and the complex interactions among the components of the problem and the solution. Additionally, the involved stakeholders often have different goals and interests which influence the artifacts' design. Each design decision influences these several other aspects of the overall situation, which in turn means that the LA solution is a reasonable compromise (Barab & Squire, 2004; Dyckhoff, 2014; Hevner & Chatterjee, 2010b).

3. Design-based research involves flexible design and artifacts revision based on their success in practice. Learning Analytics effects are difficult to predict on paper unless used and tested in real-world learning scenarios and environment. Also, real-world learning scenarios might alter the initial design processes and artifacts, because of its complexity, unpredictability. The newly acquired knowledge about the problem domain might even change the applied research methodology, thus leading to the development of knowledge that can be used in practice, and other practitioners and designers can learn from it (Barab & Squire, 2004; Dyckhoff, 2014; Hevner & Chatterjee, 2010b).
4. In design-based research, the participants are co-participants in both the design and sometimes, in the analysis. The analytics systems and tools are tools which are built to support people in learning scenarios in higher education. These people need to interpret, understand, and act upon the information provided by these analytics tools. Therefore, these tools are dependent on user involvement, and user-centered design approaches have to be used when designing and developing such tools and artifacts (Barab & Squire, 2004; Dyckhoff, 2014; Hevner & Chatterjee, 2010b).

3.2 Research guidelines

The seven research guidelines in Table 2 are based on the conjecture that knowledge and understanding of a design problem and its solution are acquired in the building and application of an artifact. In the following paragraphs, I explain the guidelines through the lens of the research problem and artifact created for this dissertation.

The outcome of a design-science project is a purposefully built IT artifact which can be applied to the problem domain. In the context of this work, the outcomes are three IT artifacts: The sustainable learning analytics infrastructure, the “Insights” analytics prototype as an integral feature (functionality) on the learning platform, and the institutional analytics tool AiX Analytics (see chapter 6). I provide a sustainable learning analytics infrastructure, empirically collected and tested goals and requirements for learning analytics in blended learning scenarios (see chapter 5), and practical approaches for scaling up analytics to an institutional level. All of them can be used and extended by future research in similar problem domains.

The problem relevance focuses on constructing innovative artifacts aimed at changing the occurring phenomena. The artifacts should solve specific problems within the problem domain. They need to alleviate the difference between a goal state and the current state of the context while being relevant for a specific target group. In this case, the context is blended learning scenarios in higher education, and the target group is the involved people (Hevner & Chatterjee, 2010b, 2010a). Based on the problem description in chapter 1, the main goal is to scale up learning analytics on an institutional level to support the applied blended learning processes and scenarios in a higher education institution in Germany.

Evaluation is a vital part of the research process. The quality, efficacy, and utility of the design artifact must be evaluated and validated with well-established evaluation methods. A design artifact is complete when it satisfies the requirements and constraints of the problems it solves and fulfills the evaluation metrics defined by the context people (Hevner & Chatterjee, 2010b, 2010a). The design evaluation methods can include observational, analytical, experimental, testing, and descriptive methods. These methodologies should be well matched with the evaluation goals and metrics, to receive proper and valid evaluation results. In the context of this work, I have used literature reviews, surveys, case studies, field studies, controlled experiments, functional black-box pilot studies, interviews, and large-scale prototype evaluations about usability, usefulness, and impact. I have evaluated each analytics prototype (design artifact) with a suitable combination of evaluation methodologies coupled with respective evaluation goals people (Hevner & Chatterjee, 2010b, 2010a).

The conducted research must provide a valid contribution to the problem area in terms of the design artifact, design foundations, and/or design methodologies. The contributions of this work are listed in chapter 1.

Research rigor is needed in both construction and evaluation of the designed artifact. With respect to construction, rigor must be assessed for the applicability and generalizability of the design artifact. The evaluation should be supported by thorough and comprehensive data collection and usage of analysis techniques in order to validate the results and relevance of the artifact to the problem domain. The core goal is to identify how well a solution works, rather than, why a solution works. I applied quantitative and qualitative methods to gain knowledge about the usability and effectiveness of the analytics implementation. I also conducted several phases of data collection and analysis to understand and tackle the complexity of the research problem and deliver an appropriate design artifact people (Hevner & Chatterjee, 2010b, 2010a).

Design-science is an inherently iterative method that tries to discover an effective solution to a problem. In the initial stages, prototypes start as simplified sub-problems, and over time evolve by eliminating their deficiencies. This process enables the researchers to gain a deeper understanding of the design space, acquire new and deeper knowledge about their goals, expectations, and restrictions of the problem context. Implementing iterative design with an evaluation cycle provides information about any subsequent steps of the research project and leads to finding a satisfactory solution to the problem at hand. By these iterative evaluations, real-world factors and constraints are identified in a timely fashion and addressed by subsequent iterations, which in turn reduces the complexity of the problem domain and helps in developing an appropriate model for the solution people (Hevner & Chatterjee, 2010b, 2010a).

Table 2. Design-Science Research Guidelines according to Henver (Hevner & Chatterjee, 2010a).

Guideline	Description
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

The last guideline advocates that research results should be communicated and shared within the research community. This ensures the validation and quality of the research work, especially when compared to related work within the same research area. Another important consequence of sharing research work is building a general knowledge base. These research studies have to be reproducible and repeatable, and their results and outcomes should reach a wider audience. In the context of this work, the outcomes and results are published in well-respected conferences, journals related to the research fields of learning analytics.

This chapter presents the research paradigm I followed as a basis for this work. In the following chapters, I describe the research process applied in this work and present the achieved results, based on these research guidelines.

4 FOUNDATIONS

This chapter provides the salient points of research areas, theories, relevant information, and terms which are closely related to the work in this dissertation. Each point is presented and observed from the perspective of scaling up learning analytics in higher education.

4.1 Blended Learning definitions

Blended learning is a fundamental redesign of the instructional model where the instruction is shifted from the lecturer towards the students, where students become active and interactive learners. There is no single comprehensive definition of blended learning, but different researchers provide summarized definitions that range from covering the content/learning delivery method, didactical goals and approaches, and technology-enhanced learning. I provide summaries from several blended learning definitions from different researchers, to point out the main characteristics of blended learning scenarios. The recurring theme in all definitions is the combination of digital and physical learning environments. According to Friesen (Friesen, 2012), blended learning designates the range of possibilities which arise from combining digital media and internet with well-established forms of face-to-face classroom learning which requires the presence of teacher(s) and students. Oliver and Trigwell (Oliver & Trigwell, 2005) summarize four concepts that characterize blended learning:

1. Blended Learning combines web-based technologies to accomplish an educational goal;
2. Blended Learning combines different pedagogical approaches to build an optimal learning outcome regardless of instructional technology;
3. Blended learning combines any form of instructional technology with face-to-face instruction
4. Blended Learning combines instructional technology with actual tasks and jobs within the learning process.

Garrison and Vaughan define blended learning as “the thoughtful fusion of face-to-face and online learning experiences” while emphasizing the need for reflection upon the traditional learning approaches to redesign the learning and teaching processes and approaches in this new domain of learning (Garrison & Vaughan, 2012).

Blended learning practice in itself is a mixture of different pedagogical approaches which combine the effective socialization of students and teacher within the classroom and different technological benefits of technology-enhanced learning (Poon, 2013). The idea behind blended learning is to get the strengths of two worlds, i.e. face-to-face oral communication and technology-enhanced learning, to combine them in an optimal way to provide a learning experience which is compatible to the learning context and educational goals. Blended learning scenarios in higher education context have different benefits and challenges for both the teaching staff and the students. Blended learning includes face-to-face learning with learning technologies. The scope of this thesis tries to improve the blended learning scenarios by analyzing and supporting the learning technologies part of the blended learning scenarios. The following two sections summarize the benefits and some of the challenges for both teachers and students.

4.1.1 Teachers in blended learning scenarios

When teachers and instructors are designing their courses within the frame of blended learning scenarios, they have to consider and decide upon their didactic design. These decisions affect how students perceive, construct, and process knowledge in these courses. The didactic design of blended learning courses should comprise of the five following topics: learning outcomes, the course learning resources and environment, the methods of interactions between students and the instructor, the interaction among the students themselves, and individual learning processes and opportunities (Paechter & Maier, 2010).

1) *Learning outcome*s are the competencies which students should have after successfully finishing the course. The students need to acquire conceptual and methodological knowledge and build social and personal competencies, such as working in teams, and successful self-regulation and monitoring of their own learning processes and progress (Paechter & Maier, 2010).
2) The teachers must prepare the *course learning resources and activities* while aligning them to the learning outcomes, and for the affordances and constraints of the course room on the learning platform. The quality of the resources and how well are they coupled with the learning platform have a substantial effect on the course and the students. These course learning activities should be matched with the face-to-face sessions of the course (Paechter & Maier, 2010).
3) The teacher/instructor should create *communication channels* with the students. The teachers have to provide the structure and important information regarding the learning resources. Additionally, they should support the students by providing timely feedback of their accomplishments, they need to motivate the students to apply themselves, and assist them to engage in different learning activities. These multiple exchanges of information about the learning resources and socio-emotional information may influence students' engagement, satisfaction, and motivation to learn better in a course. In practical terms, the teachers should develop a set of student interaction protocols (emails, discussions, announcements) which need to be followed in the course, and strategies to implement "learning support" for the students (Paechter & Maier, 2010).
4) The teachers/instructors should foster peer interactions among students by incorporating different methodologies within the course design that include group work; possibilities for information exchange about the learning contents and resources; and other activities that invoke cooperation among the students. This can invoke feelings of group cohesion and can affect their engagement in teamwork and satisfaction from the course. The teachers should act as catalysts and facilitators when it comes to online discussions, manage group work, also providing appropriate support and information when it comes to time management and study skills necessary for completing the course (Paechter & Maier, 2010).
5) The teachers/instructors should design the course, the learning resources, and the formative assessment in such a way that gives the students many opportunities to practice and apply what they are learning at their own pace. They should have the choice within the course outcomes, to learn by themselves and regulate their own learning processes (Paechter & Maier, 2010).

The faculty that decides to apply blended learning strategies and scenarios in their courses can expect an increase in interaction between them and the students and interaction among the students themselves. Blended learning scenarios and methodologies can help in increasing student learning engagement and provide flexibility in running the course within the learning environment. The learning platform with its distinct features for collaboration and communication between the participants within a course (Vaughan, 2007).

The teachers/instructors have to invest a lot of time and effort to design and create their courses with blended learning scenarios and elements. Developing and adapting the learning resources, developing, and mapping the communication channels, engaging in recording the lecture, developing the different assessment artifacts takes a lot of time and effort. Additionally, not all teachers/instructors are tech-savvy and they need technical and professional support for redesigning the course, creating the different learning resources and artifacts, and developing the skills necessary for implementing blended learning scenarios. As a final point, the teachers/instructors need information and feedback that their course design works, and that students do apply themselves within the duration of the course (Poon, 2013).

4.1.2 Students in blended learning scenarios

Students who take courses which include blended learning methodologies and scenarios can benefit from them in having time flexibility, pace their learning, develop their own learning strategies, and improve their learning outcomes (Poon, 2013).

Course design that is implemented with blended learning scenarios and methodologies provides possibilities for the students to work from home, and on-the-go. This affords flexibility for students because they can decide by themselves when and how to complete an assignment, watch lecture videos, or take part in a course discussion. Such flexibility provides a greater range of course schedule flexibility and can fit well with their other courses or daily activities (Poon, 2013).

Blended learning design combined with formative assessment can help students achieve better learning outcomes. The students have to develop their own learning strategies and their own pace of learning during the course to access and follow the versatile learning resources in order to work on the different tasks and assignments as part of the formative assessment of the course while doing that from the comfort of their home/residence. Blended learning reduces the effort to create and foster continuous learning (Poon, 2013).

However, blended learning with less face time does not equal less coursework. The students must invest effort and time to spend on the learning platform and do the expected coursework. The flexibility to organize their own learning can be dangerous, because time management is a mastered skill, and many students might have problems with it. Blended learning scenarios depend on the premise that students will do independent work (or continuous work) in the time span between the face-to-face meetings. In blended learning, students have the responsibility for learning (Poon, 2013). This can be difficult for students accustomed to being passive learners within a traditional lecture format. Another factor that hinders students in these learning scenarios is the technology itself. Students need to have enough knowledge of and are ready to use technology as part of their learning. They have to be able to navigate the information and communication technology used in blended learning (Poon, 2013).

4.2 Data privacy rules and regulations in blended learning

In the following sections, I summarize the concepts of privacy as both, students and teachers have the right to privacy and informational self-determination in the context of blended learning.

4.2.1 The concept of privacy

Privacy is a basic human right, and as such is an established element of the legal systems in countries within the European Union (Drachsler & Greller, 2016). The concept of privacy allows a right to a person to be let alone, and the right of informational self-determination. Meaning, it is up to the person himself to decide upon about showing his personal information in a given setting, with the intent to protect himself. It is important to mention that privacy is different from anonymity, and from data security. These two concepts are related to privacy in the sense of providing tools and techniques for the right of informational self-determination. However, they do not represent privacy. The right to privacy is collectively agreed upon and legally defined within the legal texts of the data privacy laws. I review the data privacy legal frameworks in power within the context of this dissertation, higher education institutions in Germany, which is a founding member of the European Union (Drachsler & Greller, 2016).

The European data privacy law structure is hierarchical. The European Parliament and the Council of the European Union pass fundamental directives (or from 2018 binding regulations) with a text that outlines the basic principles and guidelines about the directives. In this case Directive 95/46/EC (European Union, 2016; The European Parliament & The European Council, 2016), which is the Data Protection directive that regulates the protection of individuals with regard to the processing of personal data and on the free movement of such data. In compliance with these texts, the different EU member countries pass their own data privacy laws, which can implement, amend, and extend the European legislation. In Germany, this is the Bundesdatenschutzgesetz (BDSG) which implements the Directive 95/46/EC. Additionally, each federative state within Germany should pass its own data privacy law, that derives (inherits) from the federal law, in this case, the Datenschutzgesetz Nordrhein-Westfalen (DSG NRW). The two salient notions of the data privacy law in Germany (and in NRW) are the right of informational self-determination and the notion of data minimalism (Bundesdatenschutzgesetz, 1990).

4.2.2 Right of informational self-determination

The right of informational self-determination mandates that any citizen has the right over his personal data. This personal basic right is derived from the German Grundgesetz and confirmed by the German federal constitutional court in 1983 (See BVerfG, 1 BvR 209/83 vom 15.12.1983, Under C II 1. des Urteils; Rn 152). This means that any organization (public or private) which saves any data related to an individual has to disclose to this person what kind of data is being saved. If this data includes any additional data than the one publicly available for this person, the person has a right to request the deletion of this particular data. The request itself is legally binding and must be fulfilled (Bundesdatenschutzgesetz, 1990).

In the context of blended learning, the institution that provides the learning platform for the teaching staff and the students must have the capabilities to filter out and present all collected data associated with an individual user. Also, the institution (or the responsible body within the organization) must be able to remove this data, without compromising the quality of service. Furthermore, every person has the right to inquire about the sources of his personal data within the learning platform and forbid the transfer of his personal data to third parties. In other words, it must be possible that a person “can be forgotten” from the learning platform so that his data should not be reused or repurposed.

4.2.3. Data minimalism and data collection

The concept of data minimalism in the context of data privacy is that no personal data should be collected that is not necessarily needed to supply the offered service (The European Parliament & The European Council, 2016). This is a general statement which is addressed within the organization’s internal regulations which organize the provision of the service. In this case, the responsible body for providing e-learning services (including the learning platform) has to create the necessary regulations and infrastructure and define which personal data should be collected from the users. The EU directive (The European Parliament & The European Council, 2016) provides six general obligations about collecting personal data:

- The personal data must be processed fairly and lawfully;
- The personal data must be used for specified, explicit, and legitimate purposes;
- The personal data must be safeguarded from secondary use and further processing;
- The personal data must be adequate, relevant, and not excessive;
- The personal data must be accurate and up to date;
- The personal data must not be stored longer than necessary;

The technological advances and infrastructure provide a wide range of possibilities for collecting personal data from the learning platform. However, not all data needs to be collected, nor should be collected. The responsible body for providing e-learning services should collect personal data which is necessary for providing the different e-learning services, which in this case is the blended learning technical infrastructure and services. All other pieces of personal data should be not be collected without a purpose. If there is a necessity for collecting data for research purposes within the learning context, the data privacy directives and laws also regulate the collection of personal data for improving the learning processes, learning experiences, and the learning technologies that support learning within an institution. The European Commission stance on this is that the member states of the EU should ensure that legal frameworks allow higher education institutions to collect and analyze learning data. The full and informed consent of the students must be a requirement and the data should be used only for educational purposes (Vassiliou, McAleese, & Chair, 2014).

The practical implementation of this dissertation lies within the boundaries of the described concept of privacy and the data privacy protection laws. I provide the detailed implementation of the data privacy preserving strategies within the next chapters of this dissertation.

4.3 Analytics in education

Analytics in education shares a lot of features, methodologies, and techniques with the more general research field of (data) analytics. In the following sections, several research domains and fields related to learning analytics are introduced, which provide approaches and methodologies used by Learning Analytics.

4.3.1 Analytics in general

The IBM Tech Trends Report (IBM2011, 2011) have ranked analytics as one of the four major technology trends in the 2010s. Analytics in its essence is a data-centric approach and relies heavily on data collection, extraction, and analysis of (abundant) data. For example, the e-commerce and web communities are responsible for generating the need for Business Intelligence and Analytics. Corporations and e-commerce vendors have developed and transformed the e-commerce market by implementing and deploying electronic shopping platforms enhanced with recommender systems. Search engine services provide have also developed complex advertisement algorithms and ecosystems to further develop advertisements and long-tail marketing platforms to reach millions of users and their specific needs via tailored searches and personalized recommendations (Chen, Chiang, & Storey, 2012). In healthcare, analytics helps in detecting diseases at earlier stages which can be treated with ease and increased effectiveness. Additionally, the hospital treatments can be analyzed and optimized by using historical data about length of stay, success rate of different surgeries, identify at-risk patients for medical complications, disease progressions to improve the chances of curing the patients, while reducing time, cost, and effort from the medical staff (Raghupathi & Raghupathi, 2014). Web analytics is a widely used process which improves websites to leverage and enhance the customers' user experience and increase the web site's profitability. Web analytics uses user-generated data, analyzes it, and delivers insights and information about what is happening on the website in terms of usage and interactions, and using this insight and information to optimize the website for its users (Waisberg & Kaushik, 2009).

Learning Analytics uses concepts and methods from related research and application fields which include web analytics, statistics, action research, academic analytics, educational data mining, recommender systems, and social network analysis. The closely related fields (such as information visualization, educational data mining, and academic/institutional analytics) are presented in the following sections.

4.3.2 Information visualization

Information visualization of data and analytics results has the role of transferring the analysis results and data insights to the end users of learning analytics and help them in reading and understanding the analytics results and support them in their interpretations and sense-making. The sense-making and knowledge creation from visualizations is the main goal of the research area 'information visualization (IV)'.

Card, Mackinlay, and Shneiderman (Card, Mackinlay, & Shneiderman, 1999) describe visualization as "the use of computer-supported, interactive visual representations of data to amplify cognition (Fekete, Van Wijk, Stasko, & North, 2008)." In their definition, they assert that the ultimate purpose of information visualization is to amplify cognition and understanding with the help of external visualizations. The most suitable tasks for Information Visualization are exploratory tasks which involve browsing through a large information space. In many cases, the user or person using such system does not have a clearly specified goal, but by examining and exploring the data she can learn more about it, discover innovative ideas, and gain deeper insight about the represented data. This exploration often may influence the user's understanding and evoke questions and tasks based on his newly developed insights and knowledge. The authors also provide six ways of how visualizations enhance cognition:

- Increasing memory and processing resources available
- Reducing search for information
- Enhancing the recognition of patterns
- Enabling perceptual inference operations
- Using perceptual attention mechanisms for monitoring
- Encoding info in a manipulable medium

Learning Analytics relies mainly on Information visualization methodologies to convey the analytics results because visualizations have considerably bigger potential to be more revealing and more precise than conventional statistical computations and tabular representations of data (Klerkx, Verbert, & Duval, 2017). Moreover, it is vital that the visualizations of analysis data are interactive to allow the user to interact and explore the data. Images and static visualizations can serve answers to a limited set of questions, which in turn can hinder the user's understanding because the act of exploring the data can inspire new questions for which the static visualizations cannot offer answers. The temporal and the dynamic aspects of the data afford for human interactions with the data, which in turn needs the creation of meaningful visualization tools essential in exploratory data analysis (Few, 2009).

There are many approaches to Information Visualization about visualizing data and analytics results. In this dissertation, the focus of Information Visualization is covered through the use of learning analytics dashboards (Verbert, Govaerts, et al., 2013). According to Verbert, dashboards capture and visualize traces of learning activities to encourage awareness and reflection, inspire sense-making, and enable the learners to define their goals, and track these goals. Learning analytics dashboards provide a salient overview of relevant metrics and information to shape and improve the learning process and encourage teachers and students to reflect upon on their work and learning experience (Klerkx et al., 2017; Verbert, Govaerts, et al., 2013).

4.3.3 Educational data mining

Educational Data Mining (EDM) develops and applies methods and techniques from statistics, machine learning, and data mining to explore the data that comes from an educational context with the purpose to understand students and their setting, identify patterns within the data with the purpose to inform and improve the educational practice (R. S. J. D. R. Baker & Yacef, 2009; MacNeill, Campbell, & Hawksey, 2014). The domain of Learning Analytics and Educational Data Mining, their processes and goals are quite similar to each other. Both research fields are data-intensive approaches for improving education. Both work in the same educational context, work with educational data and aim to improve the learning processes and setting through analysis and insight obtained through analyzing the educational data (M. A. Chatti et al., 2012). However, these two research fields have differences in their approaches, applied methodologies, and tackle practical issues like data privacy, accessibility and reach in a different manner. Table 3 offers a succinct comparison of the two research fields based on their origins, discoveries within the analysis and data, the focus of the field, and the applied methodologies and techniques.

Table 3. Comparison of learning analytics and educational data mining (Siemens & Baker, 2012)

	Learning Analytics	Educational Data Mining
Discovery	Leveraging human judgment is key, and automated discovery is another tool to accomplish this goal	Automated discovery is key; leveraging human judgment is a tool to accomplish this goal
Reduction and Holism	Stronger emphasis on understanding systems as one unit in their full complexity	Stronger emphasis on reducing systems to components and analyzing these individual components and the relationships between them
Origins	Learning Analytics has stronger origins in the semantic web, “intelligent curriculum”, outcome prediction, and systemic interventions	Educational Data Mining has strong origins in educational software and student modeling with a significant community in predicting course outcomes
Adaptation and Personalization	Focus on informing and empowering teachers and learners	Focus on automated adaptation (e.g. by the computer without a human in the loop)
Techniques and Methods	Social Network Analysis, sentiment analysis, influence analytics, discourse analysis, learner success prediction, concept analysis, sense-making models, interactive information visualization	Classification, clustering, Bayesian modeling, relationship mining, discovery with models, complex, domain expert visualizations and results

Learning Analytics and Educational Data Mining research fields premise that data and analytics will have a transformative impact on education. Both research fields strive to provide substantial results which can support policymakers, administrators, curriculum developers, institutions, and government bodies in understanding and accepting the power of data and analytics to advocate and foster the deployment of data analytics in the upcoming era of data-driven education (Siemens & Baker, 2012).

4.3.4 Institutional analytics (Academic analytics)

Academic Analytics or Institutional Analytics is closest research field to Learning Analytics. If one takes a closer look at the potentials and definition provided by Campbell and Oblinger (Campbell et al., 2007): "*Academic analytics can help institutions address student success and accountability while better fulfilling their academic missions. Academic systems generate a wide array of data that can predict retention and graduation. Academic analytics marries that data with statistical techniques and predictive modeling to help faculty and advisors determine which students may face academic difficulty, allowing interventions to help them succeed.*" Basically, Academic Analytics has the potential to create actionable intelligence to improve teaching, learning, and student success. Academic Analytics uses data generated from course management systems, student response systems, and similar tools, analyzes is and tries to relate this generated data to student effort and success. Most academic analytics initiatives seek to predict students who are in academic difficulty, thus allowing faculty and advisors to customize learning paths or provide instruction tailored to their specific learning needs (Campbell et al., 2007).

The primary stakeholders of academic analytics are educational institutions and the management department. They receive analytics results via indicators and reports about student retention rates, graduation rates, and other performance data about how well the higher education institution and its units (faculties, colleges, schools, departments, or fields of study) are performing in fulfilling their goals and missions. The main goal of analytics is to leverage the institutional decision making with actionable intelligence derived from educational data (van Barneveld, Arnold, & Campbell, 2012). These decisions supported with quantifiable and empirical data can help the educational institution to have a better and enhanced reputation and improve its accountability while increasing the possibility of effective use of resources on different institutional levels. Similarly, learning analytics is also concerned about student success, but its' context is on a lower level of teaching and learning. Learning Analytics is concerned with supporting teachers and learning in actual courses in e-learning scenarios (blended learning and MOOCs) by providing analytics to support them and improve their learning/teaching experience.

4.4 Learning Analytics

This section provides definitions that have emerged in an attempt to define learning analytics. The subsections also discuss the setting and context of learning analytics, different models which try to describe the learning analytics process, and the methods of learning analytics delivery through dashboards and indicators.

4.4.1 Learning analytics definitions

The definition of Learning Analytics closely follows the Cooper's (Cooper, 2012) premise that analytics "*help us to evaluate past actions and to estimate the potential of future actions, so to make better decisions and adopt more effective strategies as organizations or individuals*". Analytics is the process of developing actionable insights through problem definition and the application of statistical models and analysis against existing and/or simulated future data (Cooper, 2012). The first definition is from the 1st Conference on Learning Analytics and Knowledge which was set out in the call for papers (Ferguson, 2012b): "*Learning Analytics is the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.*" The second definition comes from one of the grounding researchers George Siemens (Siemens, 2010), who defined Learning Analytics as "*the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning.*" In a later publication, Siemens (Long & Siemens, 2011) slightly amended the definition as "*Learning Analytics is the collection and analysis of traces that learners leave behind, can help to understand and optimize (human) learning and the environments in which it occurs.*" Elias (Elias, 2011) defined Learning analytics "*as an emerging field in which sophisticated analytic tools are used to improve learning and education. It draws from, and is closely tied to, a series of other fields of study including business intelligence, web analytics, academic analytics, educational data mining, and action analytics*". According to Brown (Brown, 2012), "*the process of systematically collecting and analyzing large data sets from online sources for the purpose of improving learning processes is called learning analytics. The purpose of LA is to tailor educational opportunities to the individual learner's need and ability through actions such as intervening with students at risk or providing feedback and instructional content. Learning analytics focuses on the application of known methods and models to address issues affecting student learning and the organizational learning system*" (Avella, Kebritchi, Nunn, & Kanai, 2016). The last definition comes from the Horizon Report from 2011 and 2016 (Johnson & Adams, 2011; NMC Horizon Report, 2016). The definition from 2011 Horizon Report states that "*Learning analytics refers to the interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues. Data are collected from explicit student actions, such as completing assignments and taking exams, and from tacit actions, including online social interactions, extracurricular activities, posts on discussion forums, and other activities that are not directly assessed as part of the student's educational progress. Analysis models that process and display the data assist faculty members and school personnel in interpretation. The goal of learning analytics is to enable teachers and schools to tailor educational opportunities to each student's level of need and ability.*" In the Horizon Report for 2016, the definition is slightly amended and states that "*Learning analytics is an educational application of web analytics aimed at learner profiling, a process of gathering and analyzing details of individual student interactions in online learning activities. The goal is to build better pedagogies, empower active learning, target at-risk student populations, and assess factors affecting completion and student success* (NMC Horizon Report, 2016).

Learning Analytics and its processes are geared towards education, which means the definitions have the educational context as an underlying frame. The definitions are different in some details, but they share the fundamental resemblance on creating useful intelligence and actions that foster learning from educational data.

4.4.3 Learning analytics processes

In this section, I present different learning analytics processes which describe and model how learning analytics works. Chatti et al. (2012) presented an iterative three-phase learning analytics process (Figure 1) which consists of three major steps: data collection and pre-processing, analytics and action, and post-processing.

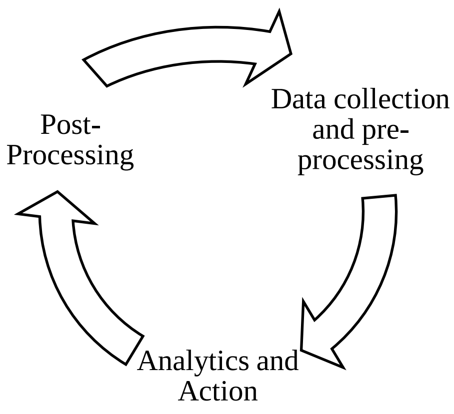

Figure 1. Three-phase learning analytics process (M. A. Chatti et al., 2012)

Data collection and pre-processing: The first step is to collect educational data. The data from the educational environments comes in different forms, formats, amounts, and sizes. Therefore, it is highly prudent to clean up, aggregate, and transform the data into a common and interoperable format, so that different learning analytics method can utilize it (M. A. Chatti et al., 2012).

Analytics and Action: After the data is pre-processed and transformed, the next step is to use different analytics methods to analyze the data. The choice and range of these analytics' methods are co-dependent on the requirements and goals of the users. The different analytics methods can discover patterns and extract actionable information from the educational data. Information visualization techniques are also applied in this step to better represent and visualize the pieces of actionable information. This step also includes the stakeholders' activities based on the presented information. These activities can include (but are not limited to): monitoring, analysis, assessment, adaptation, intervention, and reflection (M. A. Chatti et al., 2012).

Post-processing: The post-processing phase of the process stands for the continuous improvement of the analytics application while learning from previous cycles, iterations, and usages. These improvements can be on the data collection and pre-processing, such as new/adapted data sources, new ways of data collection, or aggregation methodologies. Additionally, there can be improvements on the analytics methodologies, the algorithms, refinement on the visualizations, introduction, and discovery of new questions and visualizations, or improvements of the user interface and streamlining the user experience (M. A. Chatti et al., 2012).

Verbert et al. have developed a process model for learning analytics which distinguishes four stages (Figure 2). The *awareness* stage is concerned with all kinds of data that can be visualized in different ways including activity streams and tabular overviews and summaries. The *reflection* stage focuses on the users' asked questions and assessment of the utility and relevance of the questions, thus identifying possible connections between the data and the appropriate questions. The *sensemaking* stage tries to answer the identified questions in the reflection process to create new insights. In the *impact* stage the goal is to create sustainable impact and induce new understanding and, if necessary, change in behavior (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013).

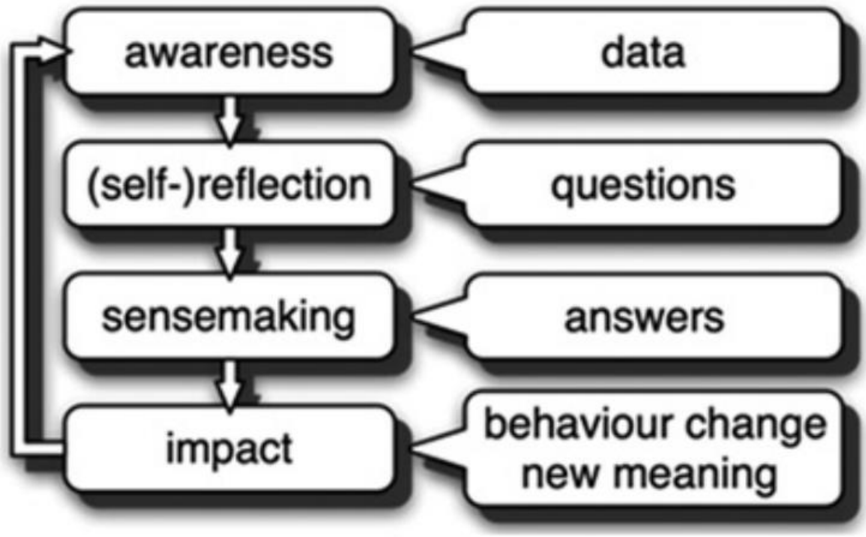

Figure 2. Four stages process for learning analytics (K. Verbert et al., 2013)

Elias (2011) presents another approach for modeling the learning analytics process, which the author based on an actionable knowledge conceptual framework called "knowledge continuum" from Baker (B. M. Baker, 2007). This "knowledge continuum" defines four components, *Data, Information, Knowledge, Wisdom.* These components are linearly structured, starting with Information at the bottom, which is in the end transformed into wisdom. Therefore, data (facts, information) which by themselves have no meaning, are transformed into usable knowledge.

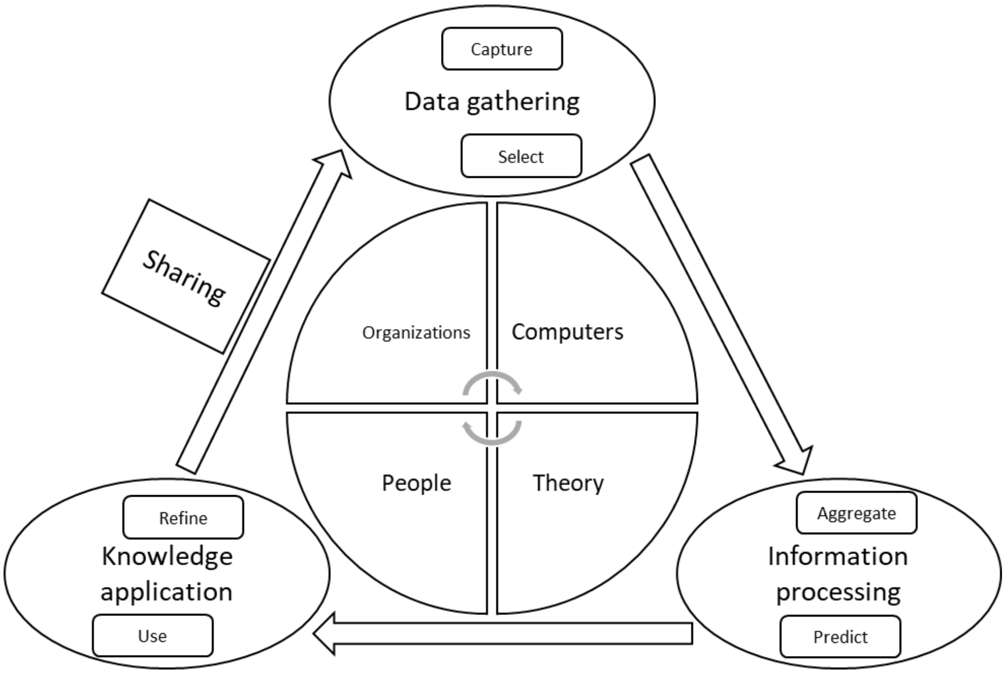

Figure 3. Learning analytics continuous improvement cycle (Elias, 2011)

This process model also consists of a cycle (see Figure 3) and has three main components: *data gathering*, *information processing*, and *knowledge application*. This process combines tools, theories, stakeholders, and organizations as the driving forces, and each component is crucial to learning analytics implementation. Put together, they complete an ongoing pattern of three-phase cycles aimed at the continual improvement of learning and teaching (Elias, 2011).

Clow developed a Learning Analytics Cycle which he argues, closes the feedback loop through appropriate intervention (Clow, 2013; Clow & Hall, 2012). Clow based his work on Kolb's experimental learning cycle, Schön's theorization on learning and reflection, and Laurillard's Conversational framework. The cycle (Figure 4) itself starts with the learners who generate data from their activities. This learner-generated data is processed and transformed into metrics and analytics to provide insight into the learning process. The last step is the intervention, meaning providing the metrics and analytics back to the learners via dashboards. The metrics and analytics are used to instigate one or more interventions that have some effect on the learners (Clow, 2013; Clow & Hall, 2012).

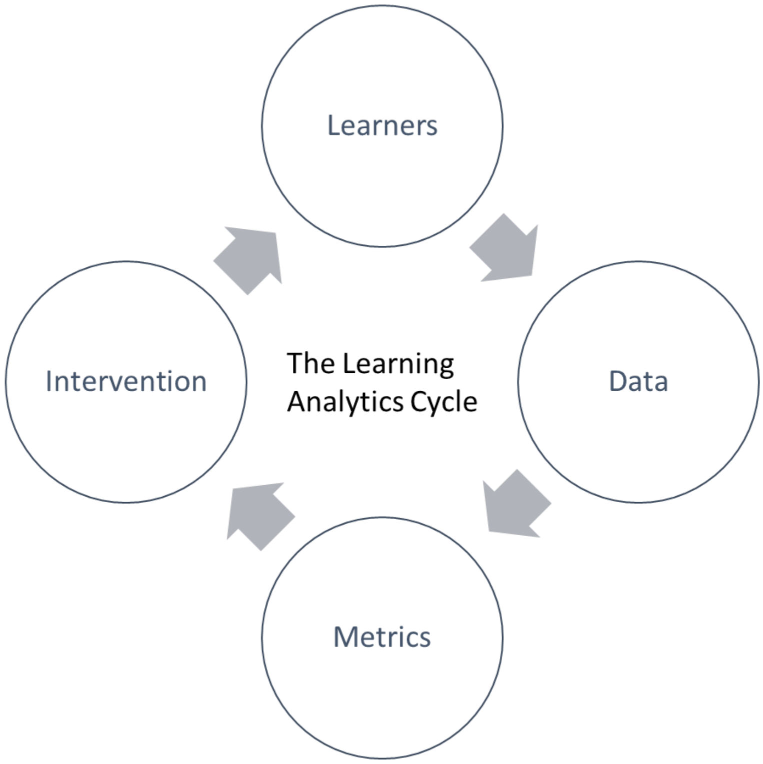

Figure 4. The learning analytics cycle according to Clow

Siemens proposed a Learning Analytics Model (LAM) that introduces a systemwide approach to learning analytics. According to him, the systemic approach would ensure enough resources and support from the management and administration to develop analytics and provide interventions to students on large scale (Siemens, 2013). The model incorporates seven components: *collection*, *storage*, *data cleaning*, *integration*, *analysis*, *representation and visualization*, and *action*. The details of the components and their connection are shown in Figure 5. Since this model stands for a comprehensive strategy of deploying and using learning analytics, it is highly unlikely that one individual possesses a combination of skills and knowledge to accomplish the implementation of the model. Hence, the institutional support and systemic approach can ensure that the right people with the right skills can be put together as a team to work on the different components of the learning analytics model (Siemens, 2013).

Within the work of this thesis, the presented models were used as a theoretical basis for designing the practical processes and methods necessary for scaling up learning analytics in blended learning scenarios. The approach of this thesis is presented in chapter 5, based on the different learning analytics process modes. Chapter 6 contains the technical implementation of the approach, whose evaluation is presented in Chapter 7. Chapter 8 discusses the findings in relation to the implementation and evaluation and validates and verifies the requirements.

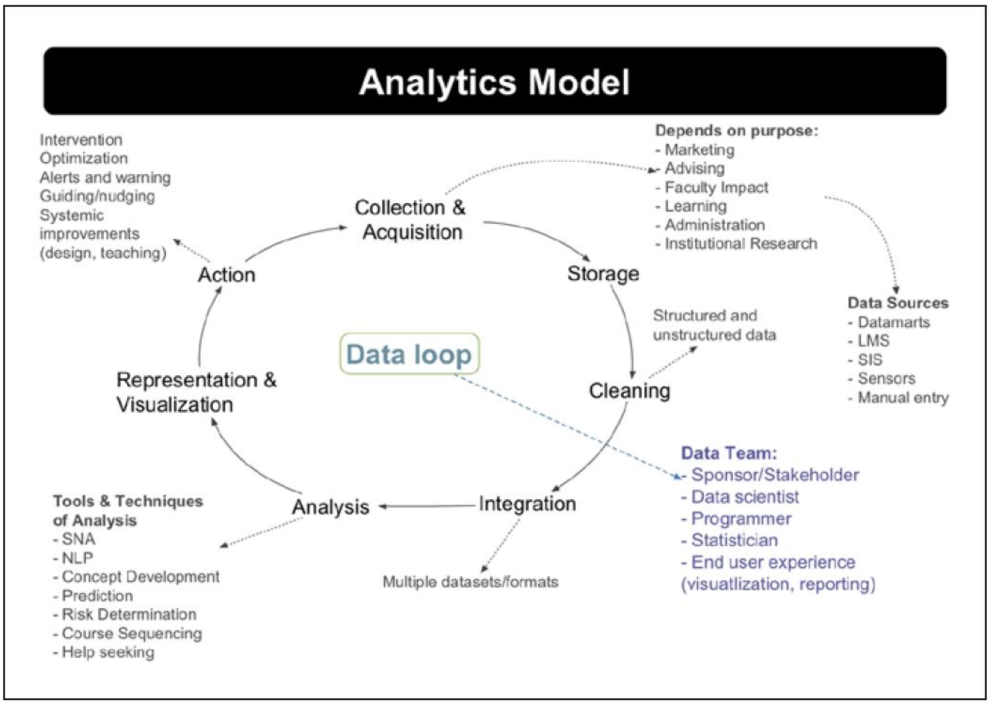

Figure 5. Learning analytics model for system-wide approach (Siemens, 2013)

4.4.4 Setting (Context): HEI vs. MOOCs (Blended learning vs. Distance learning)

The learning context of blended learning and distance learning (in this case Massive Open Online Courses, or MOOCs) differs in the way how students or learners interact with the learning materials. Blended learning has a substantial portion of face-to-face learning and direct contact with the teaching staff, their peers inside and outside the physical course room, and learning resources which are augmented and enhanced with technology (Dziuban, Moskal, & Hartman, 2005; Mattingly, Rice, & Berge, 2012). The students take advantage of the benefits of both learning contexts. They still have the traditional bonding and sense of community when interacting, collaborating, and cooperating with their peers, while enjoying the freedom and flexibility that e-learning provides. Therefore, when applying learning analytics, only the online, or e-learning part of the implemented blended learning scenarios can be analyzed and visualized with analytics indicators. The students have the opportunities and freedom to study, collaborate and cooperate outside of the online learning environments. These learning activities cannot be logged or quantified, and as such can go unnoticed from the learning analytics processes. This constraint has to be accounted for, and clearly conveyed through the analytics visualizations and indicators to avoid misinterpretations and more importantly, avoid teaching interventions based on these misinterpretations.

On the other hand, the pedagogical approaches of distance learning, in this case MOOCs are based on interactions and combinations of learning activities of watching learning videos, downloading course materials, working on formative assessment activities such as assignments and quizzes, and social activities on the discussion forums where the learners can get in touch among each other, and with the course instructors, or teachers. According to Anderson and Dron (Anderson & Dron, 2011), the distance learning pedagogical approaches applied to MOOCs are connectivism, cognitive-behaviorist, and social-constructivists. The most proliferated MOOCs approach is the so-called e**X**tended Massive Open Online Course, or xMOOC where the course is pre-defined and pre-designed in a guided fashion and follows a cognitive-behaviorist pedagogical approach (Daniel, 2012). The xMOOC has a fixed structure, following a 'traditional' university course with a top-down approach and structure in which the learning objectives and goals are defined by the instructors, who are also responsible for developing the curriculum and establishing the formative and summative (if any) assessments. The xMOOCs put a strong emphasis on individual learning rather than learning from peers. The social contacts, collaboration, and knowledge sharing have lesser emphasis through providing social features like discussion forums, while students/learners are encouraged to passively receive input from experts (Conole, 2016; Jadin & Gaisch, 2014). This distance-learning approach which strongly relies on students' motivation and self-regulation to advance in an xMOOC, while keeping retention rates high has many challenges, among which the most significant one is the lack of knowledge about the ways the students/learners interact and use the learning resources. Furthermore, the teaching staff/instructors do not receive the same kind of explicit and implicit feedback from the students/learners which is present and given to them in a traditional face-to-face classroom (Mattingly et al., 2012). To help MOOCs alleviate and solve the challenges of keeping retention rates high, and the learners are motivated and using the provided learning resources, xMOOCs platforms have incorporated extensive tracking functionality which results in enormous amounts of log data that describes the learners' online actions. The platforms are using this data to make decisions about the learners' past activities, their progress within the MOOC and try to predict about future performance and try to recognize possible issues (Mattingly et al., 2012).

The difference of these two pedagogical contexts is that for MOOCs the complete quantification and analysis of the complete dataset generated by the learners is necessary to get insights, information, and feedback of what is really happening in a given MOOC. On the other hand, for courses that use blended learning scenarios, the teaching staff collects valuable knowledge and feedback directly from the students in their face-to-face meetings within the physical classroom. In blended learning scenarios learning analytics can only provide insights, information and feedback only for the learning that happens online, and the teaching staff can direct and regulate their intervention towards the online parts of their course.

4.4.5 Learning dashboards and indicators

The learning analytics implementation and method of delivering analytics results to the stakeholders in this dissertation are by applying the concept of learning analytics dashboards. The most inclusive definition of what a learning dashboard is, was coined by Schwendimann et. al (Schwendimann et al., 2017) as "*a single display that aggregates different indicators about learner(s), learning process(es) and/or learning context(s) into one or multiple visualizations*". The concept and definition of what a learning dashboard is encompass aspects and concepts from fields as Human-Computer Interaction, Information Visualization, and the concept of indicators which use information, learner data and analytics results as an underlying foundation. The purpose of achieving one or more goals is a distinctive trait of a learning dashboard (or analytics dashboard). According to Teasley (2017), and Klerkx (2017), an analytics dashboard visualizes the important information needed to achieve one or more goals, and this visualization of information is arranged and consolidated to a single screen with the advantage to be monitored at a glance.

The learning dashboards in their form incorporate multiple indicators and visualizations to provide personalized feedback back to the users. The information presented through the indicators is highly visual and as such provides an abstract representation of usage and performance that requires the user to understand it and connect the represented data to specific actions (Teasley, 2017). Additionally, it is expected that the users should have the necessary experience and background knowledge to understand and interpret the data. However, in real-world scenarios, the data literacy, and data-driven decision making have not been investigated in detail with respect to understanding the indicators represented within the dashboard, and mapping this understanding with the appropriate and necessary actions (Teasley, 2017). The most desirable effect of the learning dashboard is to promote awareness, reflection and sense-making out of the visualized traces of the learning activities (Corrin & de Barba, 2015; Klerkx et al., 2017). In the context of this dissertation, one of the implemented learning dashboard support the teaching staff in conducting their teaching activities in blended learning scenarios, thus enabling them to adapt their teaching activities and learning resources, be aware of what is going on in their courses, thus providing analytics to discover and provide potential for improvements.

The dashboard design and what kind of indicators and visualizations vary depending on the learning design, the e-learning scenarios, and the available data. Another factor that has a large influence of what is entailed in a learning dashboard are the data privacy and legal constraints that can limit not only what kind of data is collected, but also how this collected data is analyzed and presented back to the appropriate users. From a technical perspective, a learning dashboard consists of visual elements, or indicators, which can be charts, graphs, textual data representation, and alert mechanisms (Charleer, Santos, Klerkx, & Duval, 2014; Kim, Jo, & Park, 2016). The analytics dashboard can include indicators based on analysis on the collected data about the artifacts produced by the learners, their social interactions, their use of the provided learning resources, the amount of time they spent on the learning platform, and the formative and summative assessment results (Klerkx et al., 2017).

The basic idea behind an indicator is based on presenting analytics results in small pieces of information, by using information visualization concepts for easier interpretation (Dimitracopoulou, 2008; May, George, & Prévôt, 2011). The inherent modularity of a dashboard incorporating different indicators enables personalization and enhances the user experience by providing a suitable set of indicators. Dyckhoff et al. define indicators as specific calculators with corresponding visualizations tied to a specific didactical question (Anna Lea Dyckhoff et al., 2012). The main goal of an indicator is to 'indicate' a certain situation that has been discovered with the data analysis back to the user. They provide the information and guide the user to a conclusion. However, one indicator can lead different users to different (and sometimes conflicting) conclusions. As mentioned before, the user's background knowledge, data literacy, and understanding can have a profound effect on the user's understanding and on the following user's actions and interventions (Anna Lea Dyckhoff, 2014). The most common visualization types used in indicators in a learning dashboard are bar charts, line charts, tables, network graphs, scatterplot, pie and donut graph, radar chart, word cloud, stop light. There are also more complex and composite visualizations used as well, such as treemaps, box and whisker plots, explanatory decision trees, parallel coordinate's graphs, and plant and tree metaphors (Bodily & Verbert, 2017).

4.5 Learning analytics evaluation

Learning analytics dashboards (and solutions) are interactive systems and products that support people in their everyday workplace activities. They provide user experiences that extend the way the stakeholders work, communicate and interact with them (Preece, Sharp, & Rogers, 2015). However, these interactive systems need to be assessed to ensure that they are behaving as expected, meet the requirements from the stakeholders, and assess whether they are fulfilling their aims and purpose(s). Ideally, evaluation should be intertwined throughout the entire design life cycle and development, and constantly providing feedback and modification to the design. According to Dix (2004), evaluation in the context of interactive systems has three main goals: to assess the extent and accessibility of the system's functionality, to assess the user experience of the interaction between the user(s) and the system, and to identify any specific problems with the system. Evaluation or evaluation methods can be divided into two main categories: Evaluation through expert analysis (evaluation without users), and evaluation through user participation (evaluation with users) (Dix, Finlay, Abowd, & Russel, 2004).

4.5.1 Evaluation without users

Evaluation can start in the initial stages of the design of an interactive system (ideally, before any implementation work), as an evaluation through expert analysis. For these types of evaluation methods, there is no need for user participation. The advantages of this early stage evaluation are to discover major errors and flaws within the design and find areas that are likely to cause problems because they violate known cognitive principles. These evaluation methods are relatively cheap to implement and can be used in different stages of the development process (storyboarding, design, prototyping) (Dix et al., 2004). In the context of this dissertation, I consider two approaches for expert analysis: *heuristic evaluation* and *cognitive walkthrough*.

Heuristic evaluation is a method that involves usability specialists and experts independently judging or critiquing a system to produce potential usability problems. The name of this evaluation method is derived from the term heuristics, which in this context is used as a general principle, or a guideline that can guide a design decision (Dix et al., 2004). Jakob Nielsen and Rolf Molich coined the name of the method. They developed a set of simple and general heuristics to structure the evaluation of early designs, storyboards, and prototypes. This is a simple, cheap, and very flexible method which can yield satisfactory results within a short time period. The most important rules of this evaluation method are that several evaluators involved and that each evaluation session is independent (Dix et al., 2004). The evaluators use sets of heuristics as an aid to help them discover usability problems, and each evaluator assesses the system and notes violations of the provided heuristics that would indicate a possible usability problem (Dix et al., 2004; Nielsen, 1994). The evaluator also supplies an opinion about the severity of the identified problem by using a single rating scale (Nielsen, 1993):

0= this is not a usability problem
1= cosmetic problem, can be fixed if there is extra time for the project
2= minor usability problem, fixing this problem has a low priority
3= major usability problem, important be fixed and should have high priority
4= usability catastrophe, must be fixed before release

Nielsen’s ten heuristics (Jakob Nielsen, 1995; Nielsen, 1995) that can be used for evaluating an interactive system (taken word-for-word):

1. **Visibility of system status.** The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
2. **Match between system and the real world.** The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
3. **User control and freedom.** Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
4. **Consistency and standards.** Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5. **Error prevention.** Make it difficult to make errors. Even better than good error messages is a careful design that prevents a problem from occurring in the first place.
6. **Recognition rather than recall.** Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one

part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

7. **Flexibility and efficiency of use.** Accelerators — unseen by the novice user — may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
8. **Aesthetic and minimalist design.** Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
9. **Help users recognize, diagnose, and recover from errors.** Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
10. **Help and documentation.** Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

After all of the evaluators have completed their separate evaluations, all of the identified issues, problems, and pitfalls are collected, and their mean severity calculated. Afterward, according to the severity and complexity, they have to be tackled by the development team (or person) (Dix et al., 2004).

The **Cognitive walkthrough** is a usability inspection method that evaluates a design, or a system to establish how easy a system is to learn. The focus of this evaluation method is on learning through exploration. This is derived from the observation that many users prefer to learn how to use software through exploration with a hands-on approach of learning by doing/exploring, rather than using a user's manual, or by training (Wharton, Rieman, Lewis, & Polson, 1994). The process behind the cognitive walkthrough is similar to other walkthrough processes (for example, code walkthrough), and it is a review process in which the author of one aspect of the system or the design presents it to a group of experts, or peers. The number of experts can be one to three. The evaluators or peer(s) then evaluate the presented solution by using suitable criteria (can also include heuristics) which address the design issues. The evaluators go through each step in the task of the proposed design/feature and provide reasoning why this particular step is good, or not good for a new user (Dix et al., 2004; Wharton et al., 1994). The experts must have experience with general interface design, and a good understanding of who the users are, and what kind of tasks the users are expected to perform with the interface (Lazar, Feng, & Hochheiser, 2017a). There are four artifacts that must be prepared for conducting a cognitive walkthrough (Wharton et al., 1994):

1. Specification, or a prototype of the system.
2. Description of the task the user is going to perform on the system.
3. Complete written list of the needed actions to complete the task with the system
4. Indication of who the users are, and what kind of previous knowledge and experience the evaluators can assume about them.

Having this provided, the evaluators go through each action provided in the third point, following the predefined sequence, and for each step, they try to answer the following four questions:

1. Will the user try to achieve the right effect?
2. Will the user notice that the correct action is available?
3. Will the user associate the correct action with the effect they are trying to achieve?
4. If the correct action is performed, will the user see that progress is being made toward the solution of their task?

This is a time consuming and tedious evaluation method. Therefore, it is crucial to document the entire process in a structured way for each action. An effective way to do this is to develop a standard form for each action that logs the answers of the four questions, and each answer that is negative or identifies a possible problem to be logged separately on a problem report sheet. The evaluators can also provide an opinion about the severity of the problem which can later be reused for prioritization and scheduling the fixes and modifications of the system (or the design) (Dix et al., 2004; Wharton et al., 1994).

4.5.2 Evaluating with users

Evaluation with users usually happens in the later stages of development, when there is at least a working prototype of the system in place. This working prototype can be a simulation of the system and its interactive capabilities, or a basic functionality working prototype, or a full-fledged and implemented interactive system (Dix et al., 2004). There are two distinctive styles of evaluation with users: laboratory studies and field studies. In laboratory evaluation studies the users are taken out from their environment and taken to a lab (or an ordinary quiet separate room) to take part in the controlled test. On the other hand in a field study evaluation, the evaluator goes to the environment of the user to observe the system in action, where the evaluation is conducted in a working environment which provides the context with all constraints and the real setting and situation in which the evaluated system would be used. In the context of this dissertation, I consider three evaluation methods with users: *think-aloud observation method*, *questionnaire/survey method*, and the *case study method* (Dix et al., 2004; Lazar, Feng, & Hochheiser, 2017b).

Think-aloud observation is a form of evaluation method where the user is asked to complete a predefined set of tasks on the evaluated system. The evaluator observes and records the user's action and asks the user to explicitly elaborate and verbalize their actions, and their thoughts by 'thinking aloud' (Dix et al., 2004; Lazar, Feng, & Hochheiser, 2010). The user can describe what she believes is happening, the reasons behind a specific activity, and what is she trying to achieve with her activities and interactions within the system. The entire evaluation session can be logged with different tools which provide different levels of data granularity thus influencing recall and analysis of the conducted evaluations. The methods used to record the session can be pen and paper, audio recording, video and screen recording, and computer logging. For the best results, a combination of all provides the most comprehensive set of raw data for evaluation, which can be reviewed at a later stage by several experts (including the evaluator). However, whenever there is audio/video recording, the effort and time to transcribe, analyze, and annotate everything are very expensive in terms of time and resources. To make the process less strenuous and more enjoyable, the evaluator can encourage the user different questions about her behavior, or will himself answer users' questions, and clarify problems and ideas. This approach gives leeway to the user to be more open in criticizing the system, the user can feel more confident in conducting the predefined tasks and can feel supported whenever a confusion arises. Such an approach can maximize the effectiveness of identifying problem areas (Dix et al., 2004; Lazar, Feng, et al., 2017b).

Questionnaire/Survey is a research method that is a well-written and well-defined set of questions to which an individual is asked to respond (Dix et al., 2004; Lazar, Heidi Feng, & Hochheiser, 2017). The users self-administer them (individual) without a present researcher. The lack of the researcher's presence affects the depth of the collected data and feedback. However, their asynchronous nature helps in reaching to a wider participant group, it takes less time to administer and can be rigorously analyzed. They have to be well designed, with good structure, robust, and as such can result in data with a high level of validity. They allow the researchers to capture the "big picture" of how many individuals are interacting with the evaluated interactive system, what problems they are facing, and what kind of interactions and actions they are taking

(Dix et al., 2004; Lazar, Heidi Feng, et al., 2017). Surveys are most suitable for measuring attitudes, awareness, intent, and collecting feedback about experiences, the users' characteristics, and over-time comparisons. As an evaluation method, they are less suitable for precise measurements or identifying usability problems in an interface. The most common question styles which can be used in a survey/questionnaire in the context of evaluation are the following:

- Broad questions that help to establish the user's background within the user population.
- Open-ended questions ask the user to provide his own opinion about the system, or the interface to collect general subjective information and feedback.
- Scalar questions ask the user to judge a specific statement on a numerical scale (Likert scale).
- Multiple choice questions are useful for gathering information on a user's earlier experiences.
- Ranked questions ask the user to order different items in a list to discover her preferences.

To ensure the validity and reliability of the survey, the questions must be pilot tested, to ensure that they are clear, unambiguous, and unbiased. The survey design should make it easy for the participants to understand the questions and use the tool to provide their answers (Dix et al., 2004; Lazar, Heidi Feng, et al., 2017).

A ***case study*** is a research method that conducts an in-depth study of a specific instance (or a small number of instances) within a specific real-life context (Lazar, Feng, & Hochheiser, 2017c). Close examination of such cases can be used to build understanding, generate theories and hypotheses, present evidence of an existing behavior, or to provide insight that would otherwise be difficult to gather. A case study is a detailed examination of one or more specific situations, which is happening in the context of the user (in the field). The detailed examination is documented and logged with data coming from multiple data sources, and the emphasis is put on qualitative data and analysis. The number of examined situation can in some special cases be one, but it is highly recommended that the researchers consider a small set (if available) of situations or users. Case studies focus on observing phenomena that are out of their control (unlike controlled experiments in a lab). This helps researchers/evaluators to develop detailed understandings of the interaction techniques and coping strategies of the users. The inability to control the context and the environment demands data from multiple sources to act as corroborating evidence. The multiple data sources that provide validating evidence can increase the confidence and reliability of the observations and deal with the data quality if only one data source was used (Lazar, Feng, et al., 2017c).

The goals of a case study can be fall in these categories:

- *Exploration*, which deals with understanding novel problems or situations with the hopes of informing new designs and interactions possibilities. In this case study, researchers can observe how potential users accomplish tasks, interact with available tools and systems.
- *Explanation*, which deals with developing models which can be used to understand the context in which a specific technology can be used. Technologies are often used in ways that were not considered in the initial design, and case studies can provide an understanding of these unexpected uses and outcomes.
- *Description*, which documents an interactive system with its context, or the process that led to a proposed design. In some case studies, the design process behind the system can be the focus of a case study for the researchers (Lazar, Feng, et al., 2017c).

4.5.3 Evaluating information visualization

Learning analytics borrows methods from information visualization for presenting the data analytics, abstractions, and results to the end users. Therefore, as part of the evaluation processes, there is a notable prerequisite to evaluate the interactions and visualizations of the learning analytics tools. Visualization techniques and interfaces are generative artifacts, and they only bring value to the user in a specific context. The value that visualizations generate is dependent on the context in which the user is, the data itself, and the purpose for which this data is visualized. Furthermore, the developed or evaluated visualization, or interface is a specific implementation of some more generic visualization techniques. Hence, the only evaluation that can be successfully conducted, is the evaluation of the particular instance (Ellis & Dix, 2006). According to Ellis (2006), empirical evaluation of generative artifacts is unsound, and the evaluation in itself, cannot tell whether the given visualization works or does not work. However, if an evaluation cannot prove whether a visualization is good or correct, there are ways in which visualization can be validated (Carpendale, 2008).

The combination of justification and evaluation can provide sound validity of whether a given visualization methodology works well within the given context and purpose. The justification for using a particular visualization technique can include already existing published results of analysis and experiments, empirical data from experiments and studies, and expert evaluation and opinion (Ellis & Dix, 2006). The justifications can be enhanced with carefully designed empirical studies with users, and comparison with earlier works. The most suitable evaluation of information visualization in the research context can be an explorative evaluation, to provide new things and insights about the implemented ideas and concepts and provide what is useful to the evaluators/researchers. The main difference from formative evaluations which aim to improve a design, or summative evaluations which aim to provide a seal of approval, explorative evaluations aim to provide knowledge (Ellis & Dix, 2006).

4.6 Conclusion

In this chapter, I have presented theories and research work as the basis of the proposed solution in this dissertation. Blended learning is the context in which learning analytics prototypes will be deployed. The developers of such tools must know the context in which they need to develop and deploy learning analytics solutions and dashboards. The base of practical implementation of learning analytics for a wider audience is shaped by the existing data privacy laws and regulations, the available data and shaped by the concept of data minimalism. The information visualization methods are the core for developing learning analytics dashboards which implement the outlined learning analytics processes. The last part of the chapter covered the evaluation of learning analytics presenting the theoretical background of the research methods used to evaluate the proposed solution in this dissertation. The next chapter uses the existing knowledge to derive a conceptual approach to scale up learning analytics in blended learning scenarios applied in a higher education institution.

5 SCALING UP LEARNING ANALYTICS

This chapter provides the salient points of the conducted preparation for scaling up learning analytics in blended learning scenarios. The representation of the process for scaling up learning analytics is separated into three main components: collecting learning analytics requirements, institutional preparation, and technical implementation. Collecting learning analytics requirements cover the identification of stakeholder groups and by using different techniques borrowed from software engineering and business market analysis. The institutional preparation covers the legal and practical preparations of the higher education institution for learning analytics, and the identification and implementation of processes to foster the infrastructural and technical preparation. The technical implementation outlines the preparation for development of the data management and warehouse techniques, the outlining, application and development of the analytics engine and algorithms. The results of these preparations and process are used as the basis for the actual implementation presented in chapter 6.

5.1 Essential software requirements

Learning analytics tools and services are interactive software systems whose designs are driven by different choices and requests from the users for whom these interactive systems are developed, and the activities and actions these systems need to perform for the users. One of the most challenging tasks of building software systems is deciding precisely what to build. This decision is a difficult conceptual work that includes establishing detailed technical requirements which consist of interfaces to and for the users, to and for the machines, and to other systems. If this conceptual work is done wrong, it is very difficult and expensive to fix at later stages and will definitely cripple the resulting system (Wiegers & Beatty, 2013). The resulting artifacts of this conceptual work are the software requirements for building a software system. According to Sommerville and Sawyer (1997), *“Requirements are a specification of what should be implemented. They are descriptions of how the system should behave, or of a system property or attribute. They may be a constraint on the development process of the system.”* The requirements include the behavior of the system under specific conditions (environmental, technical, and legal), the user’s view of the system, the human-computer interactions, and also the developer’s perspective of the internal characteristics of the system (Wiegers & Beatty, 2013). Overall, software requirements have three distinct levels: user requirements, business requirements, and functional requirements. In addition to these three levels, there is a set of non-functional requirements which affect the system design and implementation. These four levels of requirements also include subsets and types of requirements, it is always a good practice to define a set of consistent adjectives which change the term “requirement” which help in avoiding misinterpretations and misunderstandings. Table 4 presents the terms which are used in this chapter in the context of creating learning analytics requirements, along with their meaning.

Table 4. Terms used in the requirements engineering process (Wiegers & Beatty, 2013)

Term	Definition
Business requirement	This is a high-level objective of the organization that builds a product.
Business rule	This can be a policy, legal obligations, laws, standards, or regulations that define or constrains some aspects of the context and the situation. This is not a software requirement, but this rule (or rules) serve as the origin of several types of requirements which have an influence on the result.
Constraint	This is a restriction which is imposed on the available choices to the developer for the design and construction of a product.
External interface requirement	This is a description of a connection which the system has with another software system, or between the user and the system, or a connection with a hardware device.
Feature	Features are one or more logically related system capabilities that provide value to the users and are described by a set of functional requirements.
Functional requirement	This is a description of a behavior that the system will exhibit under specific conditions.
Nonfunctional requirement	This is a description of a property or characteristic that the system must exhibit or a constraint that it must respect.
Quality attribute	This is a special kind of a nonfunctional requirement that describes services or performance characteristic of the system.
User requirement	This is a goal or task that specific user from different user groups must be able to perform with the system. In some cases, this can be a desired attribute of the system.

The entire process of creating the requirements for a given system, or the *requirements engineering* can be logically divided into requirements development and requirements management. The requirements development process consists of four steps: elicitation, analysis, specification, and validation (Wiegers & Beatty, 2013).

The *elicitation* covers all crucial activities which discover the user requirements, such as surveys, interviews, business innovation processes, literature reviews and analysis, interface analysis and prototyping and other activities. The actions from these activities identify the users and the stakeholders of the system, their tasks and activities, expectations, and goals which are fulfilled or reached with these tasks and activities. Additionally, these activities also explore and define the environment and context in which the new system will be used and working with the users/stakeholders helps in understanding their needs, quality expectations(Wiegers & Beatty, 2013).

The *analysis* activities provide a more detailed and precise understanding of the elicited requirements and represent them in sets of requirements in multiple ways. In this phase, the task goals are identified and separated from the functional requirements, their quality expectations, the business rules and regulations, the solution suggestions, and other information that was elicited in the with the elicitation activities (Wiegers & Beatty, 2013). In this phase, the high-level requirements are decomposed and described into details and are classified accordingly to their content, context, and importance. This decomposition delivers functional requirements, the importance of the quality attributes, define the first coupling between some requirements and the software components of the system's architecture, and their relevance in relation to the system and the goals. While analyzing the elicited requirements, gaps and missing requirements are also discovered with regards to the scope and the system. The analysis activities can also discover unnecessary requirements which are simply discarded (Wiegers & Beatty, 2013).

The requirements *specification* is the representation and documentation of the requirements in a persistent and well-organized fashion. The requirements have to be transformed from the collected user needs into written requirements and suitable diagrams which ease the comprehension, review, and use by the development team (Wiegers & Beatty, 2013).

The requirements *validation* confirms that the collected set of requirements will enable the development team to build the right solution which satisfies the objectives of the project. This is achieved through a continuous review of the requirements to correct problems and errors before they are implemented by the development team and developing acceptance tests and criteria to confirm that the requirements meet the needs and achieve the objectives (Wiegers & Beatty, 2013).

The management of the requirements follows in parallel of the described four phases. The goal is to define their baseline, their development, and ensure that all the functional and nonfunctional requirements are finished and polished so that can be transferred to the development team. Additionally, all the changes and revisions have to be inspected and their impact on the existing set of requirements have to be evaluated, and all of the changes are introduced in a controlled fashion that does not break the existing requirements. The connections, relationships, and dependencies that exist within the requirements are managed and defined so that individual requirements correspond to their designs, source code and tests, their changes, and updates throughout the project. The last activity in managing the requirements is to keep the project plans and the evolving requirements synchronized and up-to-date (Wiegers & Beatty, 2013).

5.1.1 Requirements engineering practices

I applied different requirement engineering practices in developing the three levels of requirements for learning analytics in blended learning scenarios. For the business requirements, I used the results from market research and innovation technique called Outcome Driven Innovation to define the scope and to define the user-defined metrics and the e-learning processes segments for improvements. I have conducted brainstorming sessions, literature reviews, and have defined the stakeholders in details and developed user personas to capture and describe their goals and requirements. I used surveys, questionnaires, interviews, and literature review and document analysis for eliciting the learning analytics requirements. Based on the user personas and the elicited requirements, I created suitable use cases for understanding the users' requirements concerning implementing learning analytics. The validation of the business and user requirements was done with applying exploratory data analysis to discover the validity and correctness of the collected requirements. In parallel, I conducted an exploratory analysis of the existing laws, rules, and regulations existing within the higher education institution, on state, federal, and EU level. The results from this analysis provided the legal framework for scaling up learning analytics. The following paragraphs outline definitions and a short summary of each of the applied techniques for developing the learning analytics requirements.

Outcome driven innovation (ODI) is a holistic innovation approach for business and market analysis which focuses on a job-to-be-done theory (Christensen, 2010). The users employ products or technological solutions to receive a certain outcome or goal, or to get a job done. In this context, a job is defined as a fundamental goal or a problem that users (or customers) are trying to solve in a given situation (Ulwick, 2014). The ODI approach focuses on uncovering the metrics customers apply to evaluate solutions and aims to convert them into measurable items, because to evaluate a solution or a product, a set of metrics is applied to measure how effectively a product, service, or a solution can contribute to completing a job or contribute the degree of job achievement (Ulwick & Bettencourt, 2008). Following this approach, one can gain a deep understanding of the core jobs their products have and how customers measure successful or failed job execution (Christensen, Hall, Dillon, & Duncan, 2016). The ODI method is mainly used for development of new products and services but can be suitable for developing a strategy for digitalizing higher education because the results of this method are a transformation of fuzzy needs into measurable outcomes (Ulwick, 2011). The ODI approach uses traditional customer-oriented research techniques such as user interviews, and validation surveys (Ulwick, 2005). In the first step, in-depth interviews are conducted to inquire about the jobs to be done (in this case, learning and teaching) to explore the different learning and teaching processes, with the purpose to uncover their latent needs. The focus of the interviews was on their needs, with a special emphasis is given to analyzing the opportunities to identify the important and unsatisfied needs (Ulwick, 2005). The end result of these interviews and qualitative study should be the needs and metrics used to evaluate the learning and teaching processes as a job with underlying needs. These identified needs and metrics are transformed into quantifiable outcome statements, which are validated in a quantitative study. The validation step is a large-scale survey which collects and analyzes quantitative data by using ODI heuristics. The validated quantifiable outcome statements are then inspected and analyzed by applying exploratory data analysis, which increases the knowledge base about their validity and value (Ulwick, 2005).

Exploratory Data Analysis (EDA) is a well-established practice in the statistical literature and in the data sciences which provides knowledge and learning from data (Behrens, 1997). EDA techniques and aims include checks on data quality, the calculation of summary statistics, plotting and creation of proper graphs and data visualizations, and application of a variety of more complicated data-analytics techniques. Most of these techniques are based on visual and graphical data representations, which eases data exploration and analysis in providing the structural data secrets or new, often unexpected insight and knowledge of the data. The objectives of EDA are (1) to describe the data at hand, and (2) to find patterns in unsorted and unorganized data sets that allow researchers to build rich mental models of the different phenomena being examined (Chatfield, 1985). Chatfield provides a sensible summary of useful features to be considered when conducting exploratory data analytics (Chatfield, 1986). First, an essential preliminary is the formulation of the problem and the aims of the investigation. Afterward, the analysts should continue with assessing the structure of the data, its description, and most importantly, the quality in terms of completeness and accuracy. If the analyst was not responsible for the data collection, then it is important to know how the data was collected. This means that the data should be checked regarding errors, outliers, missing data, and missing observations (Chatfield, 1985; Chatfield, 1986). After the data is screened and scrutinized, the analysts should continue with conducting descriptive statistics on the data set. Summary statistics should be calculated on the data set, and for important sub-groups of the data. These statistics wherever possible include the mean, the median, and the standard deviation for each variable, as a descriptive measure for comparing the variation of samples. After this step, Chatfield (1985) also recommends plotting the data in different ways that seem appropriate to get different insights within the data. Another point is that through this analysis, data transformation and modification should be always considered because it can help in deciphering and understanding the data at hand. The last part of exploratory data analytics is the informal use of inferential methods and formulation of different models. This stage is particularly useful when the objectives of the investigation and analytics are not merely confined into describing the data. In this case, researchers wish to generate and test hypotheses, build models and analyze the data by using an appropriate inferential procedure which can be linked to other data sources (C. Chatfield, 1985; Chris Chatfield, 1986). The general principles of model formulation include collaboration with appropriate experts, to incorporate a lot of background theory to check whether a model formulated on empirical and/or theoretical grounds is capable of reproducing the main assumptions of the data. Exploratory data analytics should give the analyst a feel for the data, for picking out the key features of the data, and in generating hypotheses based on the data. The concluding remark of exploratory data analysis is that it works very well for exploratory examination of a relevant data set. Thus, it should be seen as an integrated stage of general statistical inference which is effective in supporting of model and hypotheses creation processes (C. Chatfield, 1985; Chris Chatfield, 1986).

The exploratory data analysis on the anonymous usage logs (EDA) from the e-learning platform at RWTH Aachen University aims to gain more insights about the users' engagement and needs. I used the results of the exploratory data analysis to examine the identified student needs from the ODI approach to validate their relevance and accuracy based on usage data from the learning platform in use at RWTH Aachen University. Finally, the validated needs from the ODI are used as a groundwork to transform the identified needs into data-driven actionable intelligence by developing need-based learning analytics (Ferguson, 2012a). In this way, I transform the user needs through the lens of analytics and restructure them as tangible goals and requirements to develop services and infrastructure and improve the learning and teaching processes based on the underlying e-learning platform.

During ***brainstorming sessions***, different ideas are collected within a pre-defined and agreed-upon timeframe among a group of three to ten people. There is also one person who is the moderator and is responsible for documenting the generated ideas. The basic rule of a brainstorming session is that ideas are not discussed, judged, or commented. They are simply spoken out and documented. The participants in a brainstorming session can use the ideas of the other participants, to change them, or build on them, or develop new and original ideas. After the brainstorming session, the ideas are collected, analyzed, and classified with a thorough analysis. With this elicitation technique, a large number of ideas can be collected in a relatively short time period while encouraging collaborative work. It is imperative to remain unbiased so that new and effective requirements or solutions can be developed (Dix et al., 2004; Pohl & Rupp, 2015a, 2015b).

Literature reviews or document analysis involves research, reading, and reviewing existing publications, scientific literature, technical reports, software documentation which contain potential software requirements (Pohl & Rupp, 2015a; Wiegers & Beatty, 2013). This is a very tedious and time-consuming process in the initial phase of collecting requirements, but if done properly, ensures that there is no "re-inventing the wheel" situation with the software project, or preventing breaking of existing laws, rules, and regulations. Ideally, one can find requirements specifications from similar applications, business processes, lessons-learned reports, research projects and prototypes descriptions, user manuals, institutional regulations, and legislation. In these documents can contain usable information about necessary features, user interface decisions, user interactions, regulations, compliance towards industry and data standards (Pohl & Rupp, 2015a; Wiegers & Beatty, 2013). When using peer-reviewed research reports and publications, there is always the possibility to do comparative reviews and identify shortcomings in other products and understand in which directions the product development and research are moving over time. Doing literature review and analyzing documents reveals information because it gathers knowledge from many sources, although it can be time-consuming and very tedious to apply (Pohl & Rupp, 2015a; Wiegers & Beatty, 2013).

Surveys/Questionnaires are a series of questions design to elicit information from the users about their needs and goals concerning a specific context and topic (Pohl & Rupp, 2015a; Preece et al., 2015; Wiegers & Beatty, 2013). They are an inexpensive method that aims to collect precise and unbiased statements from the stakeholders regarding their requirements. This technique assumes that the participants are capable of explicitly expressing their knowledge and they are committing their time and effort for this type of requirements elicitation. Questionnaires can collect a lot of information in a short amount of time from large user populations and can be administered across geographical or physical boundaries. The biggest challenge is forming well-written questions, thus making the preparation of a proper survey/questionnaire requires a thorough knowledge of the domain in question, and knowledge about the psychological guidelines for creating effective questionnaires. Hence, using results from document analysis, or a literature review of the domain or related products (projects) before creating a questionnaire is a highly recommended topic (Pohl & Rupp, 2015a; Preece et al., 2015; Wiegers & Beatty, 2013).

Interviews involve asking the stakeholders a set of questions about a product, or project in a specific context topic (Pohl & Rupp, 2015a; Preece et al., 2015; Wiegers & Beatty, 2013). They can be classified as structured, unstructured, or semi-structured, depending on how meticulously the interviewer follows the set of prepared questions. Agile projects rely on interviews as a mechanism to promote user-centered or user-involved design per the participatory design and development paradigm. Whenever conducting an interview to collect requirements, the interviewer or the developer have to have established the goals of the interview, the context, and scope of the interview. It is advisable to remain focused on the objective, and not waste the participant(s) time. Again, as in the case of surveys/questionnaires, the preparation period involves reviewing the literature and drafting questions, materials, and rough system models beforehand. It is sometimes advisable during these interviews to suggest ideas and alternatives during the interviews via the process of the active listening topic (Pohl & Rupp, 2015a; Preece et al., 2015; Wiegers & Beatty, 2013). An experienced interviewer always has control over the situation, asks about specific aspects or answers and ensures that the answers are complete and unambiguous.

Personas are a fictitious, specific and concrete representation of the target users (Adlin & Pruitt, 2010). They are descriptions of a hypothetical and generic person who serves as an exemplary user to get descriptive and detailed specifications about the potential users who have similar characteristics and needs (Wiegers & Beatty, 2013). Personas can help in understanding the requirements and help in designing the user experience to meet the needs of specific stakeholder groups. They define end-user groups by archetypes to raise empathy for end-users within development teams. The basic idea of a persona is to model the user groups in detail like age, gender, interests, character traits, needs, constraints, and interests within the context of the system. By providing a humane element within the development process, it is harder to disregard their needs, especially when extra effort is necessary to develop or address those needs. Personas should be truly a representative of their stakeholder class, based on the context and their demographics. Each persona has a given name and personal details, and these details are very useful whenever a design decision or compromise needs to be made (Wiegers & Beatty, 2013).

Use Cases creation places between the business requirements that set the project objectives, and the functional requirements that describe what the development team needs to implement (Pohl & Rupp, 2015a; Wiegers & Beatty, 2013). Requirements elicitation by use cases shifts the perspective from the product-centric requirements towards user-centric requirements elicitation to understand what the users need to accomplish with the system. This elicitation approach describes users' tasks that will be performed with the system, or user-system interactions which will have a beneficial outcome for some users, or stakeholders. A *use case* defines a sequence of interactions between an actor and a system. The interaction between the actor and the system should result in a valuable outcome for the actor. The structure of a use case is always fixed and is composed of a verb and an object. The verb always comes first, and then followed by the object (Pohl & Rupp, 2015a; Wiegers & Beatty, 2013). The use case has to be short, concise and convey a clear and unambiguous interaction/action. The essential elements of a use case are the following:

– The unique identifier and short name that states the user goal

- Brief text description that describes its purpose
- Trigger condition that initiates the execution of the use case
- Zero or more preconditions that have to be satisfied before the start of the use case
- One or more post conditions that describe the system's state after the use case completion
- List of sequential steps that depict the interactions between the actor and the system

5.2 Learning analytics requirements elicitation

In the context of this dissertation, the stakeholders for which learning requirements will be elicited are the **students**, the **teaching staff**, the **university's administration**, and the **IT staff** responsible for running and maintaining the e-learning tools and services.

The innovation strategy Outcome Driven Innovation was used for finding the e-learning objectives (or the business objectives) of the students and the teaching staff. The implementation and the results of the outcome-driven innovation are presented, and the results are used as business objectives for the user groups of teaching staff and students. The results are then validated with exploratory data analysis to inspect their validity and relevance in the context of the applied blended learning scenarios. For finding the needs, goals, and perspectives of the decision-making processes and needs of the administration, and the IT staff, brainstorming sessions and semi-structured interviews were conducted. The analytics indicators for all stakeholder groups as system features will be designed and developed by basing them on earlier analytics work and research from the field of learning analytics. The literature review of the research project will concentrate on descriptions about developed indicators from other projects to provide a comprehensive set of indicators, their target users, what kind of data they need, and their suitability regarding the identified business objectives. The institutional preparation defines the business rules and constraints that must be compiled. This includes the development of rules and regulations on an institutional level which are in accordance with the state and federal data privacy laws, and in accordance with the new GDPR directive of the European Union. The interface requirements will be collected and streamlined with the process of rapid application prototyping and formative evaluation with and without users. Finally, the quality attributes will be developed and deployed in accordance with the regulations of the higher education institution concerning deploying software and services in a productive environment.

5.2.1 Outcome driven innovation data collection and analysis

The study was conducted at the RWTH Aachen university located in Germany within the research project "AIX Future Teaching and Learning" funded by the "Donors' association for the promotion of humanities and sciences in Germany". The ODI study was conducted by the Department for Technology, Innovation, Marketing, and Entrepreneurship from the Faculty of Economics. The RWTH Aachen University was chosen as study object because it is one of the leading technical German research universities with more than 43,000 enrolled students, and this method aims to identify and discover the needs and problems of the students and the teaching staff within the applied learning processes. By exploring the needs and problems the students and the teaching staff encounter, RWTH Aachen University aims to develop a holistic digitalization strategy, on which different faculties can build upon for implementation of different blended learning and e-learning scenarios.

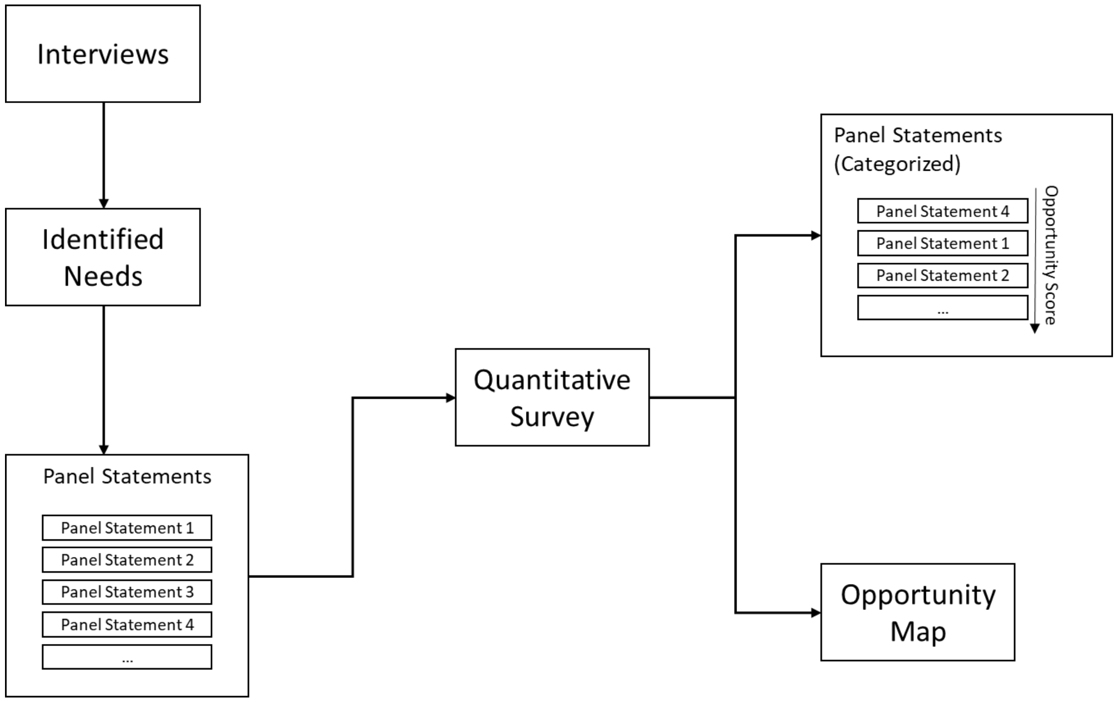

Figure 6. Outcome Driven Innovation process workflow

Figure 6 represents the workflow of the longitudinal outcome innovation case study conducted by Piller et al. from 2015 to 2017 at RWTH Aachen University (Piller, Brenk, & Nacken, 2017). Overall there are two phases, qualitative and quantitative. In the qualitative phase, semi-structured interviews are conducted, from which selected panel statements are identified, which are later used in the quantitative survey for validation. The quantitative validation of the panel statements is done through categorization according to their opportunity score in descending order, or graphically via an opportunity map.

In the *interview* phase, 45 students were interviewed with the interviews being recorded and transcribed for later analysis (Piller et al., 2017). For the teaching staff study, 34 lecturers from different faculties were interviewed to discover their teaching needs, problems, and teaching activities. The interviews followed the semi-structured strategy with prepared and predefined questions with the emphasis of extracting the needs of the interviewees, and with the possibility to allow the interviewees to provide their own ideas and input regarding their learning and teaching experiences (Piller et al., 2017). The students had a different educational background and had different fields of study, and the teaching staff was randomly contacted and chosen from the different faculties and departments at RWTH Aachen University. After the transcription of the semi-structured interviews, they are analyzed with the purpose of extracting customer/user needs out of the interview by clustering the transcripts according to needs and problem categories (Piller et al., 2017). After analyzing and describing specific and desired outcomes out of the interviews, full outcome statements with a certain structure is built. The structure of an outcome statement is presented in Figure 7.

[Direction of Improvement] … [Unit of measure] … [Object of control] … [Contextual clarifier] … [Example of object control]

Figure 7. Panel statement structure (Ulwick, 2014)

In the final steps of the analysis, all the generated outcome statements are evaluated and combined to remove duplicates, redundancies, and complexities (Piller et al., 2017). Piller et al. used frequency of the relevant outcome statements and their distinctiveness as criteria in the selection process of outcome statements for the quantitative evaluation of the findings. The outcome statements are finalized and this finalization and transformation result in a set of panel statements (Piller et al., 2017). Overall, the student set contained 36 panel statements, while the teaching staff set contained 43 panel statements. The two sets of panel statements served as the basis for the quantitative surveys in which the panel statements are evaluated on two factors: importance and satisfaction. The quantitative surveys consist of questions which are structured as 5-point Likert scale questions from which the importance and satisfaction factors are calculated. For example, if 75% of the respondents give a 4 or 5 score for the importance of a panel statement, and 53% of the respondents give a 4 or 5 score for the satisfaction of a panel statement, then the score is calculated 7.5 for importance, and 5.3 for satisfaction (Piller et al., 2017).

Both surveys were distributed online and via e-mail and overall 4589 students responded for the student ODI survey, and 268 persons from the teaching staff responded to the teaching staff ODI survey. The analysis of the results of both ODI surveys categorizes the panel statements by calculating their opportunity score (Piller et al., 2017). The opportunity score of each can range from 0 to 20, and it is calculated in the following fashion [Opportunity score] = Importance + max[Importance - Satisfaction]. If one takes the previous example, the opportunity score is equal to (7.5 + (7.5 – 5.3)) = 9.7 of this panel statement. The opportunity score can range from 0 to 20, and a score which is larger than 10 defines an opportunity for innovation, larger than 12 means high innovation potential, while 15 extraordinary innovation potentials. All scores below 10 can be disregarded because those identified needs are either fulfilled and there are enough solutions and services that cover them. The result of both surveys is a hierarchical set of the panel statements ordered by the descending order of the opportunity score. The panel statements with the opportunity for innovation are presented in the next section, while the complete list of panel statement can be found in the appendix and in the work of Piller et al. (Piller et al., 2017).

5.2.2 Explorative analysis of platform log data

I performed an explorative data analysis to confirm the results and findings of the ODI studies. I collected and analyzed statistical data from the learning platform implemented as a central e-learning service at RWTH Aachen University. The underlying platform on which the statistical analysis of the outcome statements is based is the L²P Learning platform at RWTH Aachen University. Each semester around 3.000 courses are created and managed on the learning platform. On daily basis, there are from 16000–22000 unique clients (users), and 1.5 – 2.5 million requests, while on weekends the numbers drop to 800–10000 unique clients, and 0.5 – 1.2 million requests. Around 25 percent of the use comes from mobile devices. The exploratory data analysis was conducted on the anonymous logs of the learning platform at RWTH Aachen University. For each course taught at RWTH Aachen University, the teaching staff can create a virtual course room on the learning platform. The learning platform consists of course rooms divided into different semesters, and each course room itself has six module groups: *Course organization, Information distribution, Learning resources, Assessment modules, Collaboration modules, and Course settings.* Each of these module groups contains various modules (functionalities) to support the learning and teaching processes with the learning platform. The learning platform log data was collected for three semesters (summer semester 2016, winter semester 2016/17, and summer semester 2017) with the help of the Systems and Usage Team from the IT Center at RWTH Aachen University. They used the firewall logging possibility to deliver anonymized usage logs related to the learning platform and exposed them for access via protected database service access point. The logging functionality collected and saved the HTTP requests from the users towards the learning platform. I chose to use the requests' logs because the HTTP protocol is an application-level protocol which has been in use by the World-Wide Web global information initiative since 1990 (Fielding et al., 1999). Once I established access to the raw log data, I applied Myatt's four steps practical approach to making sense of data (Myatt, 2007). According to Myatt (2007), each practical exploratory data analysis should include four steps: (1) Problem definition, (2) Data Preparation, (3) Implementation of the analysis and (4) Deployment of results.

Problem Definition. The goal of the exploratory log data analytics is to understand the usage and user engagement on the learning platform and get deeper understanding and insight into the different learning pattern and scenarios implemented within the learning platform for the digitalization of learning initiative. The primary users of the learning platform are students and teaching staff (professors, lecturers, teaching assistants, and tutors). The users can access the learning platform via desktop devices (personal computers, laptops) via web browsers, and by installing applications that can download (synchronize) learning content from courses to permanent storage on their personal computers. On mobile devices (tablets, and phones) they can access the platform via mobile web browsers, and an RWTH Campus App that also downloads (synchronizes) learning content from courses to their mobile devices.

Data Preparation. The next step of the exploratory data analysis is to prepare the log data that was provided to me. In this stage, I familiarized with the data, cleaned it from unnecessary information, and partitioned it to be analyzed (Myatt, 2007). The raw data arrived in the form of seven different parameters identifying a single HTTP request made to the learning platform. These seven parameters come from the HTTP Protocol definition by the World-Wide Web Consortium (Fielding et al., 1999). For each individual user action, one or more HTTP requests were generated from the user and logged in the raw data collection. The data was provided in one big stream, and as such, it was difficult to clean, and then analyze. I had to transfer it to an additional analytics server and partition it into smaller meaningful chunks, in this case into days. I set up a raw data transfer method, that got the data for a single day, created a raw data table for one day, and transfer it on the analytics server. After the data partition and transfer, I cleaned up the unnecessary data. For the analysis, the data clean up resulted in data logs about user read/view and create/edit activities (Myatt, 2007).

Implementation of the analysis. The data analytics on the partitioned and cleaned log data was conducted with a custom analytics tool that was specifically developed for this purpose. The implementation of the data analysis forms the creation of a customized data analytics tool. This tool summaries data based on the problem definition and is designed to implement methods that can provide descriptive and inferential statistics based on the data summaries. The creation of data summaries can be divided into two major groups. The first group of summaries has derived data tables about different read/view activities on the learning platform. The second group of summaries has derived data tables about different create/edit activities on the learning platform. The first step of the implementation of the custom analytics tool was to design and create the tables that will hold the derived and analyzed data. Since, learning and teaching activities are temporal activities which evolve over time, each daily summary of different events relates to a date. This enables in the deployment stage to create visualizations about data trends, time correlations, detect relationships, and get a feel for the analyzed data about users' engagement over time. The second step of the analysis was to develop the core methods that take different chunks of the raw data, analyze its content, transform, and identify patterns within the sequences of characters and strings. Then I have aggregated the identified patterns as findings and saved these in the database. In this step, I employed methodologies including regular expressions, string pattern analysis, and combined them with well-designed SQL queries to optimize the process of analyzing the raw data (there were around one billion entries in the raw data tables). The last step of the analysis was to save the corresponding analysis results and save in the configuration that the different raw data analysis methods for the selected day completed their analysis work. The analysis process itself was developed with a rapid application prototyping model, where many iterations were designed, developed, and tested within a brief period of time. The application was designed to automatically check for new raw data from the learning platform, imported it to the analytics server raw data, cleaned it, analyzed it, and saved the results (Myatt, 2007).

Deployment of Results. The last step of the exploratory analysis was to provide visual representations of the exploratory data analysis. For this purpose, I built a Web API architecture with services (Booth et al. 2004), and three different data representation and visualization approach for the user. The Web API application consists of different methods, which load and group together the data from the analysis tables and deliver the results in a JSON format to the visualization mechanism. This, in turn, represents the data in the form of different graphs and charts. The application itself is modular and it can be easily extended if different forms of data representation and visualization are necessary. For the visualizations, I developed a graphical interface that provides the analysis of the user engagement and behavior on the entire learning platform. Since user engagement on a platform level entails a broad range of activities and interactions, I applied the university faculty and temporal semester structure. This means that I can see the users' engagement and activities on the platform separated by weeks, months, and semester, and more importantly, by faculty. The grouping and layout present the analysis of user activities within the courses on the learning platform of a given faculty during the entire time of the available log data or during a single semester. The graphical interface also represents analysis that captures the technical aspects, system events, and devices that access the learning platform. It provides information and detailed distribution about the operating system of the devices, with which web browsers the users had accessed the platform, which devices have the biggest failure rate, authentication and compatibility problems with the platform's technical infrastructure. I used these data representations to understand how different users from different faculties engage in the learning and teaching processes from different learning scenarios at RWTH Aachen University.

5.2.3 Validated outcome driven innovation results – students

The results from 45 in-depth interviews show that students are looking for ways to enhance their learning process with new digital learning tools and they appreciate the increased flexibility. However, students also value traditional teaching concepts and face-to-face learning with a lecturer, which is perceived as more motivating and have a higher perceived degree of credibility. Therefore, e-learning solutions may be suitable to ease organization and execution of learning activities but cannot fully substitute analog learning interactions with lecturers. From the quantitative survey with 36 evaluated outcome statements, Piller et al. (2017) identified 14 under-served needs in the learning process of students, which pose high opportunities for innovation (see Table 5). They have used the opportunity equation to rank each outcome statement according to the opportunity score and only consider outcome statement with an opportunity score of 10 or more. Based on the characteristics of the selected opportunities they derived four need-based clusters as central "jobs-to-be-done" within the learning process, which are presented in Table 5 (Piller et al., 2017).

The first opportunity cluster as job-to-be-done characterizes the availability of learning resources and information as important and unsatisfied outcome statements. This involves increasing access to literature and information and reduced efforts to find relevant learning resources. Furthermore, it is important that learning materials are consistent and steadily provided to reduce the possibility of missing materials. The students expect from the teaching staff to provide them with a lot more relevant and important information, scientific literature and other published work inside their courses (Piller et al., 2017). However, the exploratory data analysis reveals that teaching staffs use literature sparsely in the course room, although there is a separate module on the platform designed for such learning scenarios. Only in around **3%** of the courses on the learning platform, the teaching staff explicitly uses the literature module to provide copy-righted and scientific work to students in their courses. The university library offers a service via the literature module on the e-learning platform to check whether the teaching staff can use a copyrighted published work within their course. Another service offered by the library is to digitalize (scan and provide in electronic form) book chapters and other paper-based published work, which is not available as an online resource. I performed an analysis on the files being uploaded in the learning materials sections, and I discovered that teaching staffs rarely use learning resources containing copyrighted original published work. From the analysis, I was able to determine that there had been only **1367** copyright check requests, and only **172** requests by the teaching staff to digitalize scientific resources. From the results, I have identified that challenges often concern the provision, delivery, and internalization of digital learning resources and content. In this regard, the university has to advance its core competencies from analog to digital provision of knowledge. Hence, the teaching staff needs to ensure that students have appropriate access to literature, learning resources and information while reducing the effort to find relevant resources by using the e-learning platform.

Table 5. Student outcome statements with high opportunity for innovation and improvement

Opportunity Cluster	ID	Outcome Statement	Opp Score
Availability of learning materials and information	3	Maximize the likelihood that literature is free, flexible and online available.	13.8
	20	Minimize the amount of missing learning material.	13.2
	13	Reduce the effort to obtain additional information about a lecture (e.g. additional examples, further explanations, application tasks etc.).	11.7
	27	Reduce the effort to find certain information and content using (e.g. on e-learning platforms etc.).	10.8
Pragmatism of learning	7	Increase the flexibility to rework lectures individually (e.g. from home).	13.3
	10	Increase the graphicness and clearness of course content to improve its understandability.	13.1
	18	Maximize the number of exercises with a practical orientation.	11.7
	28	Increase the fun factor of learning (e.g. in lectures).	10.7
Certainty within the learning process	34	Reduce the risk of not being able to solve an exercise on my own.	12.5
	21	Reduce the insecurity of getting incorrect information.	10.8
	8	Maximize the number of opportunities to get direct feedback on exercises (e.g. based on e-learning).	10.6
Encouragement & support	26	Minimize the likelihood that I misestimate my personal time management throughout the semester.	12.4
	9	Maximize the number of achievements as success experiences within the learning processes.	12.2
	14	Reduce the effort to plan the semester.	11.5

In the second opportunity cluster, students are looking for more pragmatism of academic learning in terms of higher flexibility, fun, graphicness/descriptiveness, and increased practical orientation. This cluster of pragmatism in learning should increase the motivation by building on more practice orientated approaches, which prepare students for their future profession or enhance the ability to re-work a course on their own (e.g. by providing class videos) (Piller et al., 2017). However, the exploratory data analysis showed that only in around **10%** of the courses, the teaching staffs provided video-based learning resources for re-working the course content. Additionally, the analysis showed that students used their phones/tablets on regular basis outside the university network. Around **20%** of the usage of the learning platform was with applications for downloading the learning resources via the API of the learning platform or using a phone app provided by the university that has integrated features to load learning resources. At the beginning of each semester, around **30-40%** of the devices on the learning platform are mobile devices (see Figure 8). This means that students are actively engaging the platform on the go. During the semester time, the number of devices falls to **20-25%**, which means that one-fourth of the entire platform activity happens on mobile devices. Thus, there is a strong need to access the learning platform on mobile devices. In this case, the user interfaces and the learning platforms should be optimized for such devices, so that the effort to access these materials is minimized and enable flexible learning "on-the-go" which is independent of a desktop computer. From this analysis, one can see that most of the uploaded learning resources and materials are not compatible (fit on a variety of smaller screens, good readability, and visibility etc.) with such mobile learning scenarios, which cause problems and low satisfaction when learning on mobile devices.

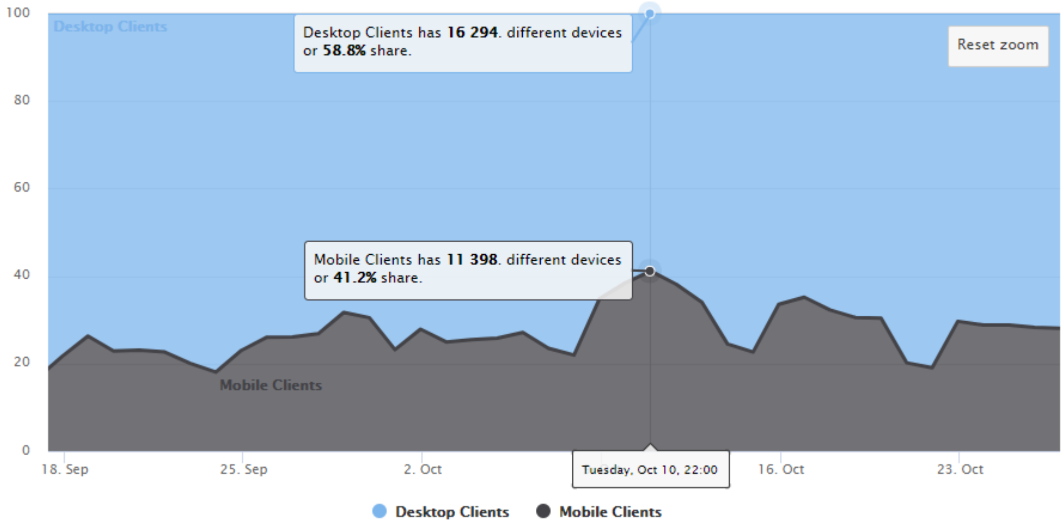

Figure 8 . Usage share (%) of mobile versus desktop devices

The third cluster of outcome statements emphasizes students' need of certainty within the learning process. Teaching staffs need to assure the provision of relevant information to students and minimize the distribution of misleading or incorrect information. Additional, students have the need to receive direct feedback about exercises on the e-learning platform and reduce the risk of not being able to solve these exercises on their own (Piller et al., 2017). Based on the data analysis, I have analyzed students' behavior and engagement in the learning materials module over a time period of a semester from different faculties. I discovered different patterns of behavior among different faculties on the e-learning platform. For example, the students from the largest Faculty 4 of Mechanical Engineering have high peaks of usage at the beginning of the semester, and at the end right before the exams (see Figure 9, left side).

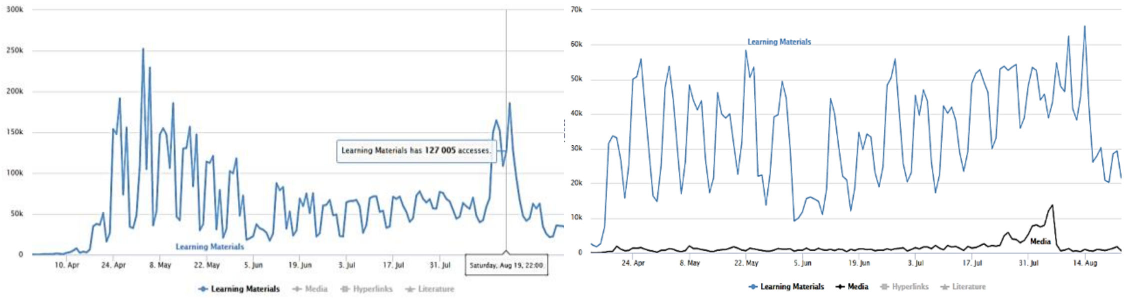

Figure 9. Students' behavior and engagement in the learning materials module of Faculty 4 of Mechanical Engineering (left) and Faculty 1 of Mathematics, Computer Science, and Natural Sciences (right)

Students at the second largest Faculty 1 of Mathematics, Computer Science, and Natural Sciences show a different behavior compared to students from the Faculty 4 of Mechanical Engineering. Student engagement is evenly distributed over the semester with less volatility (see Figure 9, right side). To investigate these two patterns, I segmented the ODI data according to faculty affiliation and it turns out that outcome statement 8 (see Table 5) to maximize direct feedback is not an opportunity for Faculty 1 (opportunity score below 10). The result of the ODI segmentation explains the different pattern of learning material usage on the platform because students in Faculty 1 are strongly engaged with exercises, lab experiments, programming session etc. throughout the whole semester. Therefore, they already engage on a regular basis with their teaching staffs and the provided learning material on the platform, while getting more direct feedback and information. Through the continuous communication and engagement with the learning material on the platform (e.g. through assignments) students from Faculty 1 get up-to-date information and steady feedback regarding course organization and requirements. The continuous interaction on the platform reduces uncertainty because students are getting a clearer understanding of the expectation of successfully completing the course. In contrast, students from Faculty 4 stressed in their interviews that often the course size is too large to engage in direct communication for feedback to teaching staffs or professors (Piller et al., 2017). In these cases, frontal teaching approaches with only a final examination at the end of the semester are often applied to manage the large numbers of students by formalizing the teaching processes. From a student perspective, this approach provokes a high degree of information gathering at the beginning and end of the semester to plan efficiently the course completion and reducing uncertainty about the learning process.

The last cluster of outcome statement emphasizes the need for encouragement and support during the learning process that reduces uncertainty and organization efforts within the learning process. Students need more support and encouragement when organizing the semester and engaging in formative assessment activities as part of their learning process (Piller et al., 2017). In this regard, students need more support and tools for managing their time and engagement during the semester. The university needs to provide more transparent and efficient tools for time management. Additionally, they need to inform students throughout the entire semester that they need to adapt their behavior and engagement in learning and assignments for successfully complete a course (Piller et al., 2017). The exploratory data analysis reveals that on average, only **5-8%** of the courses per semester use the assignments module at the e-learning platform as part of their teaching scenarios. Furthermore, this module is automatically connected with the gradebook module, where all the scores from individual assignments, electronic tests, and other coursework and assessment activities can be entered and distributed to students. This module for providing assignment and exam results is data privacy conformant, and the teaching staff is strongly encouraged to use it. The gradebook module provides features that let students compare their scores with the rest of the class through graphics and visualization of grade and score distribution. However, only around **20 %** of the teaching staff uses this module. According to the data analysis, it is one of the most used features on the platform by the students when teaching staffs apply it. This is a strong indicator that students need to be informed about their progress and situation within the course. Nevertheless, using both modules for assignments and grading require larger personal investments of time and effort from the teaching staff. Therefore, only **2%** of the courses per semester use these two modules in conjunction to increase students' engagement and encouragement throughout the semester. For assignment and grading, teaching staffs rather choose the learning materials module of the platform, as a place for uploading assignments and exercises. In this case, the assignments and the following organization of score, grades, and feedback are carried out offline and not online on the platform itself. In **30-40%** of the courses on the platform for every semester, this scenario is chosen. However, using the e-learning platform primarily as an upload tool of documents does not enable the provision of a detailed overview of progress and success/failures throughout the semester. In this case, teaching staffs are not able to provide advice and encouragement by analyzing student progress on demand, especially not for large courses. According to the results, key action points for the development of user-centered e-learning solutions are interactive encouragement and support within the learning process, which is based on incentivization, motivation, and self-management of students.

In conclusion, students are looking pragmatically at their learning activities by requesting e-learning tools to support them to reduce the effort to study and thus maximize their learning outcomes. Furthermore, they would like to have more real-world examples and success stories that can better prepare them for their professional careers after their university education. The students understand the benefits and affordances of the new technology they use on daily basis and would like to have the opportunity and flexibility to use the learning resources and materials on their own. The students would also like to have a wider range of better freely available learning resources and scientific literature. In the context of learning analytics, this can be interpreted that the students would like to have analytics tools that will help them in achieving better results by effectively using the proper and suitable learning resources (which have high-quality content, real-world examples, and applicable knowledge and information). The developed learning analytics tools need to access multiple sets of data, and different analytics algorithms and interfaces are needed to provide student support in their learning experiences.

5.2.4 Validated outcome driven innovation results – teaching staff

Brenk et al interviewed 34 lecturers from different faculties about their teaching needs, problems, and learning activities (Piller et al., 2017). Overall, the interviews had shown that faculties differ in their evaluation of e-learning services. Whereas some faculties already experiment with e-learning, other faculties are reluctant to deviate from traditional face-to-face teaching approaches. In total, they identified 43 needs in the qualitative research and used these for the quantitative evaluation, out of which there are eight under-served needs and they categorized them into four opportunity clusters (Piller et al., 2017).

Table 6. Teaching staff outcome statements with high opportunity for innovation and improvement

Opportunity Cluster	Nr.	Panel Statement	OPP Score
Encouragement for continuous learning processes	23	Increase the likelihood that students prepare and rework course content independently.	16,9
	15	Increase the likelihood that students engage in continuous learning processes.	14,6
	39	Reduce the likelihood that students are not intrinsically motivated to study for a lecture and rather learn to achieve a certain grade result (e.g. learn only to pass an exam etc.).	14,1
Scientific writing	40	Increase the possibility for students to learn academic/scientific writing.	12,7
Sureness while using e-learning in teaching	26	Reduce the likelihood of infringements of copyrights when using e-learning formats.	12,4
	33	Minimize the opportunity to cheat in electronic exams.	10,6
Organizational effort	12	Reduce the time effort related to bureaucratic processes when setting up a new course.	14,7
	18	Reduce the time effort to become familiar with new e-learning technologies and their application areas.	11,3

First, lecturers want students to engage in continuous learning processes and encourage students to be prepared for their lectures, to interact in class and discuss the learning content. This means that teachers need to be aware and reflect upon the student behavior in the course room regarding the course content and learning resources they provide for the students. Teachers invest a lot of time and effort in designing and implementing their learning scenarios, creating and providing these learning resources to the students. Therefore they need to intervene when they identify that the learning resources they provide, are not being reworked on regular basis by the students. This is a clear indication that they need tools to observe the student engagement with the learning resources on the learning platform. This can be achieved by providing descriptive statistics and analytics, and student engagement distribution over time on the learning resources in a course on the learning platform. The teaching staff would like to have intrinsically motivated people who would like to learn new skills, acquire knowledge that could be reused, and not just learn to pass the exam. The teaching staff should have tools and mechanisms that show the student motivation and commitment to the course materials, and goals. Again, this argumentation goes into the direction of students applying themselves to learn new skills and acquire relevant knowledge.

Secondly, the panel statements refer to the teaching of academic skills to students. University lecturers have a great focus on scientific methods and want to pass on these skills to students. They want to pass on academic skills for writing scientific works to the students, and at the same time avoid legal risks as copyright violations. The teaching staff should have features on the learning platform that enables them to maximize the usage of original published work and scientific literature which is protected by copyrights in the learning process. Also, teachers need to be aware that there are such functionalities in their course room/learning platform, and that they should use them as part of their learning scenarios. As mentioned in section 5.24, only in **3%** of the courses on the learning platform, the teaching staff explicitly use the literature module to provide copy-righted and scientific work to the students in their courses, although in reality a lot more scientific literature is distributed via the learning materials module on the learning platform

Finally, innovation potentials also refer to the administration and organization of lectures. Most teaching staffs are challenged with the bureaucracy and the time consuming administrative work when setting up or operate a lecture (Piller et al., 2017). The lecturers want to reduce the effort when setting up a course on the learning platform. Many teachers reuse materials and resources from previous semesters because over time they have iterated and optimized their pedagogical approaches, and learning resources to fit the blended learning scenarios. This way they can save time, and they can invest efforts in more productive work for their teaching and research activities (Piller et al., 2017). For this concrete need, the learning platform offers a feature for the teachers to import learning resources and content from a previous semester in order to reduce the effort for setting up a new course. The import feature is used in **55%** of the courses (Figure 10), by the teaching staff to re-import learning content from old courses on the learning platform.

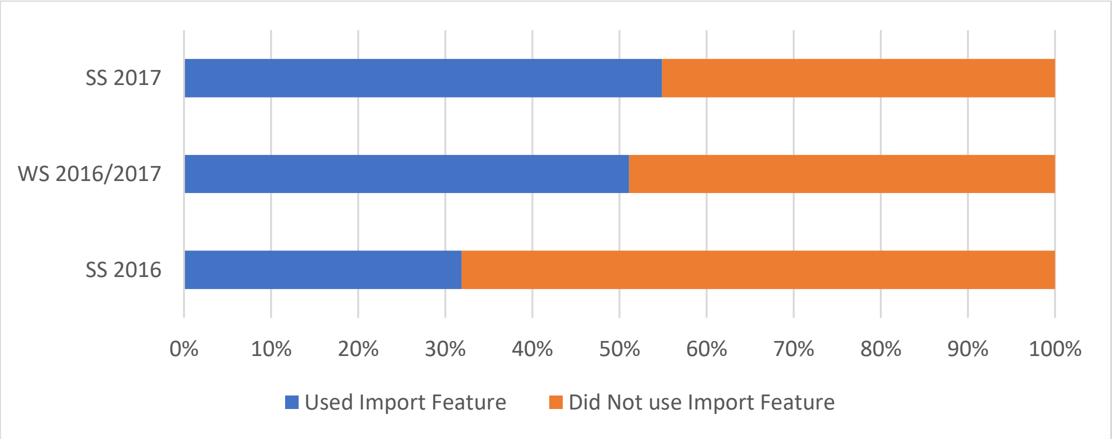

Figure 10. Importing content from old to new courses

I also analyzed the way how the teaching staff interacted with the learning materials module. The module itself affords two different types of interactions for uploading learning resources: drag-and-drop, and through uploading pop-up window. When comparing uploads with drag-and-drop and creating learning resources explicitly via the uploading pop-up, the teachers prefer the second action, although it took more steps and more time (Figure 11 shows the distribution of the use of both features). One reason might be that they are not aware of the drag-and-drop functionality of the learning platform.

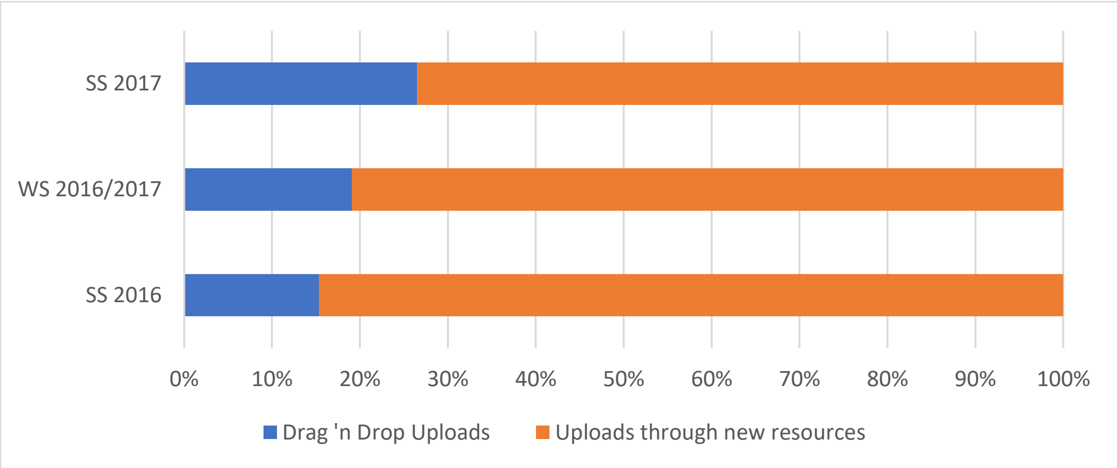

Figure 11. Different ways of uploading learning materials on the learning platform

The results indicate that lecturers lack competencies to work with e-learning services. They question the pedagogic value of pure e-learning concepts and state that tradition face-to-face concepts are not substitutable in some fields. However, e-learning may pose some solutions to currently under-served needs: It can be suitable to reduce the bureaucratic and organizational efforts for lectures and help to foster continuous learning of students. This need also supports the assumption that lecturers want to interact with students and to discuss the learning content in a dynamic, but reliable teaching environment.

In conclusion, the teaching staff wants an overview of what is happening inside their courses over the semester, expect student engagement, continuous learning, and expect that students have intrinsic motivation and are able to work by themselves through the entire semester. On the other hand, the teaching staff would like to reduce the time and effort for creating and managing a course on the learning platform, and reduce the time and effort when using e-learning tools and services. In the context of learning analytics, the teaching staff needs tools and analytics that can provide them overview about their course, the student engagement and use of the provided learning services over the duration of the semester. However, their need for a decrease of invested time and effort for using e-learning tools and services can limit the development and deployment of new and complex e-learning tools and interfaces. Hence, the learning analytics tools should be easy to use and understand, provide insight and feedback of what is happening inside a given course, while reducing the time to use and decreasing the cognitive load on the users (the teaching staff).

5.2.5 Administration and IT staff semi-structured interviews results

In the following section, I outline the salient points of the goals and perspectives of the different people within the administration and IT staff (Lukarov & Schroeder, 2017b). The stakeholder group includes the blended learning coordinators[2] (one for each faculty), the committee for e - learning (consisting of all BL coordinators, plus the management leadership from the institutes responsible for e-learning processes and implementation), the rectorate and dean(s) responsible for teaching, academic affairs, planning, development, control and institutional research, and the center for teaching and learning (including the developers' team). Each of these users (or user groups) has different responsibilities, perspectives, and goals concerning their e-learning process and scenarios. The goals and perspectives from the different user groups were collected in two phases (Lukarov & Schroeder, 2017b).

In the first phase, I conducted a literature review and several brainstorming sessions to identify and sketch out the different perspectives of the different user groups. The second phase was a series of semi-structured interviews with different people from the identified users' groups to gain an insight about how they collect and gather pieces of information, and how they use this information in their decision-making process. The intermediate results from the first phase were used as a basis for the guidance of the semi-structured interviews in the second phase (Lukarov & Schroeder, 2017b). The different involved groups and the salient points of their respective goals and perspectives are summarized in Table 7.

Table 7. Administration goals and perspectives for learning analytics (Lukarov & Schroeder, 2017b)

User group(s)	Goals and Perspectives
Developers Team IT Staff Platform Support and Qualification Team	• Learning Platform usage patterns • Concurrent users and requests (traffic) • Different modules use frequency • OS/Devices distribution among users • Errors and bugs identification
(Blended Learning) for e-learning Committee Blended Learning Coordinators	• Resources planning • Implementation of different for e-learning approaches • Feedback and guidance for faculty and students o Continuous Learning vs. Cramming • Success rate/Evaluation of Blended Learning initiative • Scientific literature/library resources use
Rectorate Department for planning, development, and control	• Accountability and university image • Resource allocation • Effective use of learning resources and infrastructure among different faculties

The analysis of the results of the interviews showed that that different groups responsible for different aspects of the e-learning processes have similar or even common goals, but still have different perspectives. For example, the developers' team is interested to learn how different faculties use the various functionalities of the learning platform in their course rooms and how (when, how much, which devices) the users interact with the different modules in these courses. This way they can evaluate their modules, find patterns of usage, potential problems, system load, and other parameters that could help improve the platform from a developer's/technical point of view. At the same time, the blended learning coordinators, who are responsible for implementation of the blended learning initiative would like to know the same thing, but from a different perspective (Lukarov & Schroeder, 2017b). They also need this information, how different users from different faculties use the learning platform; on which modules, they use to

[2] Blended Learning Coordinators are responsible for implementing the blended learning initiative at RWTH Aachen University. They are part of a rectorate appointed committee whose goal is to oversee the university-wide implementation of blended learning concepts.

implement their teaching and learning approach; does their approach rely on lecture scripts, or also includes videos and media; whether collaboration and discussion is useful; is formative assessment integral part of their courses; and many questions that follow under implementation of different for e-learning approaches and strategies.

Another example is that the members of the department for planning, development, and control are interested in how different faculties use the underlying learning and technical infrastructure (Lukarov & Schroeder, 2017b). They need to have an overview and compare how different faculties implement their e-learning initiative(s), which also coincides with the perspective of the coordinators, who need to oversee and analyze that on individual faculties to come to different conclusions and decide upon different actions and measures. Another aspect that was revealed during the interviews was that to gather and aggregate information about how to analyze and decide over is the lack of tools and lack of readily available and understandable data, which can support them during their work. They do have different data available from them through different channels (mostly qualitative) such as number of (new) students, students evaluating teaching, log data, interviews, and surveys, but nothing conclusive nor systematic on a regular basis. Furthermore, they must invest a lot of time and resources just to understand the data, because it comes in a variety of forms, factors, and amount, and thus needs a versatile skill set which goes beyond the scope of their work. This makes it particularly challenging to analyze and assess the e-learning scenarios and situations just to get an overview of the situation, and nearly impossible to make informed decisions and undertake measurements to achieve their respective goals regarding the e-learning initiatives and activities at RWTH Aachen University.

5.2.6 Literature reviews and document analysis results

The literature review and document analysis were conducted on publications from the relevant conferences, journals, and books whose area of research is technology enhanced learning and learning analytics. Furthermore, technical and summary reports from government bodies and educational organizations were also collected and analyzed for existing research, initiatives, and technical implementations of learning analytics. The set of analyzed documents included the complete proceedings from the conference on Learning Analytics and Knowledge (LAK), and the European conference on technology enhanced learning (EC-TEL), the German e-learning conference, different journals about learning analytics, e-learning and technology-enhanced learning, and reports from the European Commission and educational organizations. The literature review and document analysis were geared towards broad collection and analysis of research questions concerning technology enhanced learning with the context of the design, development, and application of learning analytics tools. The relevant publications (case studies) were analyzed for the following goals:

- Collect research and experiences about didactical approaches to didactical goals, questions, e-learning scenarios, and learning theories with relation to technology-enhanced learning in relation to the application of innovative technologies and media in learning scenarios.
- Collect research, practical applications and technical details about implemented learning analytics tools.
- Collect technical and implementation details concerning learning analytics indicators, their definition, and description, intended users, what kind of educational data they are using, applied analytics algorithms and applied data privacy mechanisms.

Table 8. Classification of questions asked by teaching staff in blended learning (A. L. Dyckhoff et al., 2013)

Qualitative evaluation of learning	Quantitative evaluation of learning	Student/learner differences
How difficult/easy is it to use the learning offer? Why do students appreciate the learning offer? How do students like/rate/value specific learning offerings?	How often do students use a learning environment (per week)? When and how long are the students accessing specific learning offerings (during a day)? Are there specific learning offerings that are NOT used at all?	By which properties can students be grouped? Do native speakers have fewer problems with learning offerings than non-native speakers? How is the acceptance of specific learning offerings differing according to user properties (e.g. previous knowledge)?
Learning resources versatility Are students using specific learning materials (e.g. lecture recordings) in addition or alternatively to attendance? Will the access of specific learning offerings increase if lectures and exercises on the same topic are scheduled during the same week?	**Correlation between learners and resources** Which didactical activities facilitate continuous learning? How do learning offerings have to be provided and combined to with support to increase usage? How many (percent of the) learning modules are student viewing?	**Effects on performance** Is the performance in e-tests somehow related to exam grades? How do those low achieving students profit by continuous learning with e-test compared to those who have not yet used the e-tests?

The collected questions and didactical goals were separated into six questions categories based on Dyckhoff's (2011) meta-analysis classification: (1) qualitative evaluation of learning, (2) quantitative evaluation of learning, (3) student/learner differences, (4) learning resources versatility, (5) correlation of learners and learning resources, and (6) effects on performance. The question groups with the collected example questions are presented in table 8, based on Dyckhoff's and Lukarov's approach (A. L. Dyckhoff et al., 2013). The complete list of questions can be seen in Dyckhoff's publication, and in the Appendix.

The first group of questions are exploratory in nature and can be answered via indicators that collect qualitative data from surveys, interviews, or automated feedback collection within the learning platform. The group of questions about the quantitative evaluation of learning can be answered via quantitative analysis of the (anonymous) log data from the learning platform. The same is true for questions concerning the learning resources versatility and its correlation with their usage scenarios from the learners. The most difficult questions to answer are the ones that involve the collection and correlation of historical, personal and performance data. The literature review and document analysis showed that the teaching staff already have many questions concerning the resources they provide, the students' behavior and the correlation between them.

The second objective of the literature review resulted in analyzing 74 learning analytics tools and research projects. The underlying work was based on previous work by Dyckhoff et al. (2013; 2014) which developed a categorization scheme for the identified tools and learning analytics indicators. I took this existing categorization scheme and extended it with new tools, newly identified (and developed) indicators, and dimensions within the categorization scheme. For each identified learning analytics tool, I conducted additional research for literature, publications, and documentation for discovering its intended users, type of data, analysis methods, and what kind of indicators the tool contained. After reading and analyzing the comprehensive set of publications and technical documentation, 272 learning analytics indicators were collected. The indicators' names and descriptions were collected in relation to each learning analytics tool extracted from the scientific publications and technical reports at hand. Afterward, I applied the categorization scheme that mapped the indicators to their respective tools; mapped the indicators to their intended users (stakeholders); mapped the indicators to what kind of data they need for their analysis; and mapped the indicators according to the identified teachers' questions from the previous literature review objective. The goal behind these categorizations was to identify to which extent the development of the learning analytics tools was driven by the available data, and whether the development of these learning analytics tools was aligned with the stakeholder needs' and the didactical aspects of the applied learning scenarios. The complete list of indicators and their categorization schemes are available in the Appendix.

Concerning the categorization according to stakeholders (Table 9), 78 indicators are intended for the students, 267 are intended for the teaching staff, 12 indicators for the university's administration, and 32 to for the IT staff responsible for providing the technical infrastructure and e-learning services to the teaching staff and students. The research projects concentrate on providing indicators mostly for the teaching staff and students. The administration and the IT staff are overlooked in comparison with the number of indicators available for the teaching staff and the students.

Table 9. Distribution of LA indicators concerning target users

	Students	Teaching staff	Administration	IT Staff
Indicators list	**78/272**	**267/272**	**12/272**	**32/272**

Concerning the categorization according to indicators and what kind of data they need (Table 10), 105 indicators can provide results with anonymous data, 134 indicators can provide results with pseudonymized data, 118 need personal/private data to provide results, while six indicators need course organization/campus related data. The combined results show that for 200 indicators it is enough to have anonymous or pseudonymized data to deliver analytics results.

Table 10. Distribution of LA indicators with regard to data sources

	Anonymous Data	Pseudomized Data	Personal/Private Data	Course/Campus Related Data
Indicators list	**105/272**	**134/272**	**118/272**	**6/272**

The last categorization according to indicators and the research questions, where the indicators were mapped to the identified corresponding questions. The purpose of the mapping was to identify which learning analytics indicators can partially, or fully support finding the answer to the questions at hand. Furthermore, this mapping of indicators and questions also raises reasonable concerns about the data-literacy of the users, since researchers and domain experts can understand the connection between them but there is no guarantee that the average user would also make the same connection between her research questions and the provided indicators. Most of the questions belonging to the category of quantitative evaluation of learning can be answered with the available indicators because they use readily available log data as an abstract indicator for access of resources, and time spent on the learning platform. The questions concerning the qualitative evaluation of learning are not answered with the existing indicators because of an identified trend of developing indicators with the premise of what can be analyzed and calculated with the existing data generated from the learners and the learning platform, and this data does not include data from surveys and ratings (Although there are developed indicators that try to somewhat calculate this by using different algorithms and calculations of the generated log-data and branding them as engagement indicators). Many of the questions about the correlation of learning activities and teaching activities are not answered at all by the indicators, although the collected data would suffice to answer them within a data privacy conformant analysis. The questions from the students/learner differences and effects on performance categories are the least ones addressed by the collected indicators of the tools. The biggest problem with providing indicators for them is related to data privacy because such indicators would have to access a lot of personal data and information with the supposition that these indicators will deliver invaluable insight and empirical improvement to their stakeholders. From a researcher's perspective, the most valuable questions remain unanswered, because developing such complex indicators would require usage and correlation of qualitative data, official academic data, profile data, personal records and grades, teacher data.

As a final remark, it is important to state that both questions and goals and the learning analytics tools and indicators stem from research projects in the corresponding field of learning analytics. This is a notable factor because it influences the way these tools are developed, and their intended purpose. The number of implemented and deployed analytics tools in a productive setting (on large scale) is very small (Ferguson et al., 2016). This notion can be supported when looking at the goals of learning analytics concerning tools, educators, and students from Dyckhoff and Lukarov (2013) in table 11.

If one compares these identified goals from the research with the ones from the validated outcome driven innovation results, the difference between them is striking. The learning analytics goals from the research community are far more reaching and ambitious, while the teaching staff and students are concerned with much more pragmatic objectives and goals whose context is the day-to-day teaching and learning activities within the implemented e-learning scenarios. For a large-scale introduction of learning analytics and its adoption from the stakeholders within the learning process, it is crucial to consider the actual goals and needs of the stakeholders. For example, according to the ODI results, for the teachers, it is crucial for them to see whether the students use the learning resources and engage in continuous learning. On the other hand, the students want to optimize their learning in such a way so that with minimal effort to achieve the best results.

Table 11. Goals of learning analytics concerning tools, teaching staff and students(A. L. Dyckhoff et al., 2013)

Learning analytics is supposed to	Educators are supposed to	Students are supposed to
• track user activities • capture the interaction of students with resources / the interactions among students • gather data of different systems • provide educators/students with feedback/information on students' activities • provide an overview • highlight important aspects of data • provide different perspectives • offer possibilities for (peer) comparison • draw the users' attention to interesting correlations • pinpoint problematic issues • establish an early warning system	• monitor learning process/way of learning/students' effort • explore student data / get to know students' strategies • identify difficulties • discover patterns • find early indicators for success / poor marks/drop-out • draw conclusions about the usefulness of certain learning materials and success factors • become aware / reflect / self-reflect • better understand the effectiveness of learning environments • intervene / supervise / advice / assist • improve teaching / resources / environment	• monitor own activities / interactions / learning process • compare own behavior with the whole group / high performing students • become aware • reflect / self-reflect • improve discussion participation / learning behavior / performance • become better learners • learn

5.2.7 Developing Personas

The results from the conducted interviews, the validated results from the two conducted ODI studies, and the literature review and document analysis when aggregated and combined provide a comprehensive collection of requirements for learning analytics. This comprehensive collection was also used as a knowledge base for the brainstorming sessions concerning the creation of personas of the intended stakeholders. The personas for each stakeholder group were created in two brainstorming sessions with the help of two young researchers who did the work as part of their master theses (Mentiu, 2018). The developed personas were structured with a poster taken from the Creative Companion which covers all important questions and information about them. Each developed persona contains details that describe a real person like name, gender, age, occupation and main character features, and specific details concerning their personality and the context in which they are going to use the developed software (Mentiu, 2018). These details were developed by answering questions like: Which expectations do they have from the tool? What is important for them in terms of user interface? How would the tool fit in their daily schedule and work? What would stop them from using the tool? Furthermore, personal questions were also collected to answer and to understand how active/or social are the personas? Do they have hobbies?

The answers to these questions were collected, revised and then inserted into the persona's poster in Figure 12 (Mentiu, 2018). The personas poster held the following sections:

- **Name** – a realistic name of the persona. The name can be augmented with a wordplay to provide additional information to the name and to emphasize its importance.
- **The descriptor** is a short-selected set of keywords describing the persona as an individual
- **Quote** contains several sentences that a persona would have said about herself.
- **Who is it?** contains descriptive facts about the persona such as age, occupation, marital status, country of origin, and residence.
- **What goals?** contains the points of interest for this persona. For example, why would this persona use the analytics tool?
- **What attitude?** outlines the approaches that this persona would most likely take when using the system. Such information covers the personal preferences towards the user interface, the types of tools and elements she would like to have to extend her workspace.
- **Which behavior?** covers the daily scenarios in which she would use the system, and which are the most common problems and tasks she would like to solve with the system. The behavior contains useful descriptions of how the user would perceive the system and what should the development team watch out for while developing features and interfaces for this persona.

The first step after developing the skeletons of the different personas was to map them to the different stakeholder groups (Mentiu, 2018). For the teaching staff, there are two personas, a university professor, and a teaching assistant. The university professor persona also covers the administration stakeholder group because most of the university bodies and administration committees consist of faculty members and instructors. The teaching assistant persona is very important, because they are the main group of users that organize the lecture, the exercises, and are responsible for the successful implementation of the didactical approaches and the learning scenarios within a given course (Mentiu, 2018).

Figure 12. Persona core poster example by Creative Companion

The student's stakeholder group received two personas because in the brainstorming session and analysis distinct traits of two student groups were: those who aim to have excelled academic performance and students who struggle/want to get by. The first want to achieve outstanding levels of performance and knowledge, while the second aims to identify the minimum satisfactory level that can be met to finish their studies. The IT staff (or the development team) received a separate persona (Mentiu, 2018). The short descriptions for the personas follow here, while the detailed persona posters can be seen in the Appendix. Special care was put into developing their names, which outline their main concerns and goals, their gender, social background, their field of study and academic aspirations (Mentiu, 2018).

Teaching Staff and Administration Persona: Elisabeth van der Rate, 47 years old, Dutch, married. She is a professor of Economics and Finances at Maastricht University. She is the chair of the Committee for Resources Allocation for Improvement and Innovation at Maastricht University. She is also a parent of two students (Mentiu, 2018).

Teaching and research assistant: John Datov, 33 years old, male, German, single. He is a research and teaching assistant at the Faculty of Mechanical Engineering, and responsible for organizing the exercises connected to the lecture: Introduction to Mechanical Engineering I (Mentiu, 2018).

Students: Johannes Chillermann, 19 years old, male, German, single. He is a freshman studying Biology and trying to earn a bachelor's degree (Mentiu, 2018).

Lili Tuto, 26 years old, female, Chinese, single. She studies in a master's program in computer science and is working as a student assistant and is tutoring assignments for bachelor courses in mathematics and programming (Mentiu, 2018).

IT Staff: Ravi Infra Stracturi, 28 years old, male, Indian, married. He works at the IT Center who provides the technical infrastructure and development effort for the various e-learning services and the learning platform at the university. He is a software developer who is responsible for maintenance and development of existing and new features on the university's learning platform (Mentiu, 2018).

5.2.8 Personas, use cases, and indicators

The next steps were to create use cases for the developed personas. Two composite use cases were created for each persona and they were created to have a story-like manner when describing them and contained three main components: problem description, solution, and a result. After the use cases were created, the list of indicators and the categorizations from section 5.2.6. The use cases will be used when developing the learning analytics indicators, their visualizations, and their groupings and representations within the interface for each intended user. The development team can decide what kind of input is needed from the user, and what kind of data representation and visualization strategies should be applied and how would the tool be presented to each user. Here, I supply three examples of use cases, while the rest of them can be found in the Appendix.

John Datov (teaching assistant) Use Case 1 (Mentiu, 2018)

Problem: John uses e-tests as part of the implemented learning scenarios in one of his courses. He has uploaded an additional paper that explains how to solve the more difficult questions in the upcoming quiz. After grading the quiz, he is puzzled to see that only 30% of his students succeeded in solving the question. He wants to know what happened because he provided a solution to the challenging problem.

Solution: John opens the analytics tool in his course room and can easily see how many people have accessed and seen the additional paper that provides solutions to the challenging problems. He can see that only around 30% of the students have accessed the paper.

Result: John is aware of the situation is caused by the fact that students have not accessed nor seen the additional paper as a valuable learning material.

Elisabeth van der Rate Use Case 1 (Mentiu, 2018)

Problem: As a member and chair of the committee for allocating resources and improving the e-learning strategies and implementations she needs to decide which department needs new and better video equipment for producing lecture videos and media as more engaging learning resources. On the next budget meeting, she needs to give a presentation about video-based learning at her university and identify whether and how different departments and faculties use media and videos as part of their e-learning scenarios.

Solution: Elisabeth opens her the analytics tool, from where she can easily select the different faculties and different departments and groups from each faculty. Afterward, she finds the media usage indicators and visualizations, where she can select different video types and can generate a comprehensive comparison visualization that shows the distribution of used media and videos per courses, semesters, departments, and faculties.

Result: During the committee session she uses the generated visualizations to argue which departments or faculties need more financial support for video production equipment and allocates the necessary funding.

Johannes Chillermann Use Case 1 (Mentiu, 2018)

Problem: Johannes is in the middle of the semester and he would like to know how he is performing by now and what is required for him to qualify for the final exam.

Solution: Johannes opens the learning analytics tool and looks for the indicators that show the following information:

- Number of completed assignments, number of points and percentage with the relation of the assignments left in the course
- His current state in the course, compared to his peers, and to the minimum requirement to qualify for the exam to pass the course
- List of learning materials, and the approximate time needed to read/process them

Result: Johannes is now aware of how much more work he needs to invest in the course and his current learning progress and situation. During this activity, he discovers new learning resources which can help him.

The solutions provided with the use cases were mapped to existing indicators to identify which indicators, sets of indicators, combinations, and aggregations would provide insights, actionable intelligence, and results that help the persona in fulfilling the goals and achieving a positive result (of the use case) (Gospodinova, 2018; Mentiu, 2018). Two sessions were conducted, and the collected indicators were analyzed, discussed with the context of the personas, and matched the indicators with the personas and the use cases. The resulting sets of many indicators were too large (in some cases more than 50 indicators), therefore a prioritization and classification of them was necessary. The main reason is that providing sets of tens of indicators on a single interface can quickly overwhelm and overload the user's understanding and cognition (Gospodinova, 2018; Mentiu, 2018). The choice was made to have around seven to ten indicators per persona, by following the findings from research about the human capacity for processing information given a limited amount of time (G A Miller & Miller GA, 1994; George A. Miller, 1956). Furthermore, not all scenarios and identified goals had corresponding indicators from the collected indicator set. This meant that new potential indicators had to be defined and tailored towards the goals and needs of the stakeholders. Many of the identified indicators were combined to form one and more insightful indicator, some of the indicators were removed or marked as not important considering the limited design-space and time for indicators and initial finalized lists of indicators per persona were identified. These indicators are not complete, nor a comprehensive list that covers all possible indicators or learning scenarios available but covers the most suitable indicators per given persona (Gospodinova, 2018; Mentiu, 2018).

Elisabeth van der Rate's indicators list includes (Gospodinova, 2018; Mentiu, 2018):

- A number of students, sessions, and requests over time in each course, department, semester, or faculty.
- Trends in student activity based on the time spent online.
- Student reactions and interactions based on teachers' activities within a course.
- Quantification of the use of e-learning offerings over time: learning resources, electronic tests, and assignments, types of learning resources, collaboration activities, and

engagement in mobile learning. These should be available per course, department, semester, or faculty.
- Correlation between use and performance in e-tests and assignments over time per course.
- Timely adoption of learning resources per course.

John Datov’s indicators list includes (Gospodinova, 2018; Mentiu, 2018):

- Combination of time spent on the course, which resources are used, when during the day the students are studying or using the learning offerings. This should be available over a time period for a course.
- Learning path analysis, which shows how students access learning resources over time, assignments and exercises and their correlation.
- Correlation between use and performance in e-tests and assignments over time per course.
- Trends in student activities over time during the semester.
- A number of unique users per resource over time, and resources that have not been used over time.
- Time-dependent distribution of students in discussions, and their connection to lectures and assignments.
- The grade distribution for students in lectures and assignments.
- Most popular resources, and time spent on them.

Johannes Chillermann’s indicators list includes (Gospodinova, 2018; Mentiu, 2018):

- My time spent learning per week, month, and/or semester
- Compare my results of the assignments, and e-tests to the rest of my peers
- Most used and useful learning resources over time by my peers
- Compare my learning path/trajectory with my peers
- The grade distribution per course of the previous semester/year/course iteration
- My progress, workload and time spent learning over time compared to my peers
- Reminder about deadlines.

Lili Tuto’s indicators list includes (Gospodinova, 2018; Mentiu, 2018):

- All indicators that are relevant for Johannes Chillermann
- Time distribution of submissions of assignments
- Online discussions after my tutoring session: patterns, new topics
- Online activities within the course after my tutoring session
- Grade and points distribution of my tutoring group in comparison with other groups, and the rest of the course
- Mistakes that often come together while solving assignments/e-tests
- Most difficult questions/assignments/e-tests

Ravi Infra Stracturi’s indicators list includes (Gospodinova, 2018; Mentiu, 2018):

- Resources use over time (amount of transferred data, number of users, locations, platforms, devices, operating systems, browsers, modules)
- Resources creation over time in different courses/departments/faculties
- Learning platform’s use with correlation to unique users over time
- Desktop/Mobile devices and usage distribution over time
- Distribution of bad requests on the platform (number/percentage of users who receive bad requests, location, devices, operating system, browsers, modules)
- Time distribution of students’ activities on the learning platform by modules

- Comparisons of technical usage between departments, faculties, over different semesters/years

The collection of indicators for development for each persona concludes the collection of the functional requirements and features for each stakeholder group. The next step was to identify the non-functional requirements and characteristics that the developed system must respect.

5.2.9 Non-functional requirements

The non-functional requirements were also collected through the literature review from the comprehensive set of publications and learning analytics tools available. The analysis results were later confirmed by the distinct types of evaluations conducted on the developed prototypes. The collected non-functional requirements are:

- *Usability*: The user interface needs to be understandable, suitable methods for data representation and visualization must be chosen, and the user should be guided through the analytics process
- *User Experience:* the user experience needs to be pleasant, and the analytics interfaces should be streamlined and easy to use. The overall experience should convince the user to use the analytics tool on a regular basis
- *Usefulness:* the learning analytics system should provide relevant and meaningful indicators that help the users to gain insight about what is happening on the learning platform, and support them in fulfilling their goals
- *Extensibility:* the system needs to be designed in a way that easily allows development of new features, indicators and incorporating them in the existing system
- *Sustainability:* the learning analytics system needs to be built with solid and proven technologies which can be safely deployed and hosted for an extended period. Furthermore, the system must be safe and work despite hardware and software updates regarding the underlying technical infrastructure.
- *Performance:* the system must have enough processing power and handle multiple requests from multiple courses and return results/visualizations to sustain streamlined user experience
- *Data privacy:* the system needs to preserve confidential user information and protect the user's data and identity.

5.3 Institutional Preparation for scaling up learning analytics

The institutional preparation for scaling up learning analytics covers the development of rules and regulations concerning the different e-learning services on an institutional level. Learning analytics implementations are (or will be) just one tier of the different available e-learning services in a given higher education institution. The developed rules and regulations provide the legal foundation and the legal framework and appoint the responsible bodies for the actual implementation and provision of the different e-learning services. The management and the use of these e-learning services and systems in higher education institutions are controlled by internal organizational instructions (Organisationsanweisung), the official documentation of the procedures and operations for handling personal data (Verfahrensverzeichnis) for a given system or application, and the description of the official documentation of the procedures and operations for handling personal data (Verfahrensbeschreibung) for a given system or application.

The data privacy in learning analytics is a relevant issue and is an integral part of the non-functional requirements because it covers the aspect of collecting, storing, and analyzing personal and sensitive data that impacts the privacy of the users of these systems. However, the users' privacy must be taken into consideration and protected as part of all technical solutions and technology that are employed in supporting the learning and teaching processes in a higher education institution. Therefore, the legal solution and framework must encompass all systems and services that store private and sensitive data, and the development and deployment learning analytics tools and services is just one part of those services and must be covered by it. The university with its internal governing bodies (the university's government, rectorate, and its Senate) can create and develop official regulations that govern and regulate the use of technology and e-learning services within the learning processes. This led to the development of the so-called "eLearning Ordnung zum Schutz personenbezogener Daten bei multimedialer Nutzung von E-Learning-Verfahren an der Rheinisch-Westfälischen Technischen Hochschule Aachen", or translated Regulations for the protection of personal data in multimedia applications and use of e-learning methods at the RWTH Aachen University. This official document outlines the rules which apply to all services and e-learning solutions and e-learning processes that use and are processing personal and sensitive data within the university for the purpose of scientific training. The process for developing these regulations was coordinated by Prof. Dr.-Ing Ulrik Schroeder, who at the time was the chief executive officer at the Center for Innovative Learning, and responsible for the design and development of the learning platform at RWTH Aachen University. The legal texts and content were written by the RWTH Aachen University's legal department with technical consultations with me, and with the Data Privacy Officer at RWTH Aachen University to make sure that the regulations were in accordance with the state and federal data privacy laws in Germany, and would be in accordance to the (back then still newly, and not yet in power) General Data Protection Regulation (GDPR) of the European Union. The entire process lasted for more than a year, and during the official process, the rectorate met on multiple occasions to update and improve the texts of the different regulations. The regulations were officially proclaimed as incumbent at the beginning of 2016.

The regulations consist of 14 paragraphs which outline the scope of the regulations, defines the affected persons, the basic principles and rules, and outlines the responsible body and its duties towards the affected persons. Within the regulations, distinctions among the different types of personal data are outlined and divided into inventory data, usage data, content data, and media recordings of the lectures. There is a dedicated paragraph that handles research and use of personal data within this research, the consent of the affected persons, and how long such data can be legally and safely stored. In the following section, I provide the salient points of the official regulations.

The ***scope*** of the regulations covers all the personal data and individual information that identifies an individual person. These persons can be students, teaching staff, guest lecturers and any person who is involved in any of the teachings and learning at the university. This also includes the development team who handles the provision of e-learning services and solutions. In the scope of these regulations belong all the e-learning processes, software and tools which are used by the identifiable persons for learning, teaching, and examination at RWTH Aachen University which collect, store, use, process and change, transfer, block, and delete personal data of the involved persons.

The ***basic principles and rules*** state that the persons responsible for the provision of technical solutions and e-learning services should always stick to the concept of data-minimalism which means that they should collect only the necessary personal data and strive towards collecting as little personal data as possible. They are also responsible for developing clear and concise concepts for protecting the personal data they need for the type, scope, and purposes of processing personal data. These data privacy concepts have to be developed and provided before the provision of an e-learning service, and they have to be available to the users at all times. Additionally, if it is technically possible or reasonable, they should provide mechanisms for the users to stay anonymous on the system or use it with a pseudonym. These persons are allowed to process personal data and may share it or make it accessible and available to persons who are part of the university, or persons who participate in a course or other learning scenarios if this data is needed for achieving or completing a specific e-learning scenario, or a course.

The three ***types of personal data*** which can be collected are inventory data, usage data, and content data. The persons responsible for provisioning the e-learning services may only process personal data such as name, address, matriculation number, or email address if it is only required for registration, or for use of the e-learning services at RWTH Aachen University. Usage data about how the person(s) behave online within the learning environment(s), can be collected, processed, and analyzed only by the responsible persons for provisioning e-learning services, and only if it is necessary and is in the context of improving the provisioning of e-learning services. Additionally, they can monitor, aggregate usage data from the users who are using different e-learning services for the purposes of improving the learning offerings, the learning platform, and other e-learning services. The same rules apply for data that is generated by the persons on the learning platform, on the different communication and collaboration services within the e-learning services.

These three types of personal data can be collected and processed only by the responsible persons for the purpose of optimizing the teaching and learning offerings, and the persons involved within these processes have a clear benefit, and their rights and privacy are not violated by such usage of this data. Furthermore, the processing of personal data concerning the gender, nationality, educational background, or the course of study can be processed and used only by explicit expression of consent by the affected persons. All other data from these three types can be processed and presented in the pseudonymized form. If this data is shared or transmitted to other bodies for research purposes, it must be completely anonymized. Another privacy concern arises when lecture recordings are created and contain the faces of the live audience. These lecture recordings are permissible if they are required or chartered by an educational mandate of the university. However, the participants must be informed about the type and form of recording and transmission of these recordings (or live streams) before the recording starts. Any outside use of these recording requires the consent of the lecturer. The implementation of automated electronic assessment for important milestones (such as exams) also collects personal data from the students, and as such can be proofed by an examiner or a professor at the request of a student concerned to the responsible examination board. All electronic submissions of electronic tests or assignments must have a digital timestamp immediately after the submission.

The ***users' consent*** is binding if it is based on their free decision. The users must be aware of the intended purpose of the processing of their personal data, and where applicable and necessary, also about the consequences of the refusal of consent. Consent must be given in writing unless another form is appropriate due to special circumstances. The users' consent can also be electronic if the persons responsible for provisioning the e-learning services can ensure that the users have given their consent consciously and clearly, and this consent is safely recorded and saved. Additionally, the users must be able to recall the consent at any time and revoke it with effects in the future. If they have revoked the consent, then their personal data must be deleted, or at least anonymized, unless there are other regulations that require their storage. If the deletion or anonymization of their personal data poses a threat to their achieved points and grades, they need to be informed before the deletion or anonymization. Participation in a course may not be constrained or depend on the consent of the users to the use of their data for other purposes.

There is ***data expiration*** defined for the three types of personal data. The inventory data shall be stored until a person deregisters or ex-matriculates from the university. Request of the person can delete this data at an earlier point in time. The usage data can be stored if this is necessary for the implementation and improvement of the e-learning services. However, this timeframe should not be longer than five years. The same is valid for the content data. It should be deleted after five years.

The responsible persons for provisioning the e-learning services are also responsible for the integrity of the collected personal data. They need to take the necessary technical and organizational measures to protect the collected data, as long as it is necessary for implementing the e-learning services and it is within reasonable costs. The developed data privacy concepts for the different e-learning services have to warrant that the (1)purpose for the collected personal data is rational and reasonable; (2) only authorized persons to have access to the data, and these authorized persons will not read, copy, modify, or delete personal data without authorization; (3)it can be subsequently verified and established whether and by whom personal data was modified, or deleted, or where this personal data was passed on; and (4) the personal data is protected against accidental destruction or data-loss.

The originally published regulations are used as a basis to protect the right to privacy of the users of the different e-learning services and solutions at RWTH Aachen University. However, they are general and encompass the topic of data privacy in a broader sense. Whenever an e-learning application is developed and deployed for productive use, there is a separate process that needs to be conducted as a requirement of the German state and federal data privacy laws. This process is the creation of the official documentation of the procedures and operations for handling personal data (Verfahrensverzeichnis) with a description of the official documentation of the procedures and operations for handling personal data (Verfahrensbeschreibung) for the developed and deployed e-learning application. This process is conducted between the technical lead of the development team, and the appointed data protection officer of the higher education institution. These two official documents are describing and documenting the collection, storage, and processing of personal data which are later provided to the data protection officer. He is required by law to publicly provide these documents (or parts of these documents) and make them available to everyone in an appropriate manner upon request. The structure of this document follows a structure predefined by the state data privacy law:

- Name and address of the institution/unit that processes personal data
- The complete and detailed purpose and the legal basis for the processing of personal data
- A complete list of the types of stored and processed personal data
- A complete list of the persons affected by the stored and processed personal data
- A complete list of the authorized persons, or groups of persons that are allowed access to the personal data
- A complete list of personal data that is transferred or forwarded, and a complete list of recipients of this forwarded personal data
- A complete list of the sources from where the data comes from/originates
- The predefined expiration date of the collected personal data
- Data transfer to third countries (outside Germany and outside the European Union)
- A complete description of technical and organizational measures regarding
    - Data confidentiality
    - Data integrity
    - Data availability
    - Data authenticity
    - Possibility to revise and audit data
    - Data transparency
- A complete description of the technical infrastructure that hosts and runs the application
    - Type of application (intranet, web)
    - Required operating systems
    - Required proprietary software

- Installed firewalls and network security

The necessary institutional preparations concerning the legal aspects of the development and deployment of all e-learning services within a higher education institution were completed. The introduction of learning analytics tools as a service for the teachers means that before the application is deployed, the software system must be discussed with the data privacy officer, thoroughly documented and the official documentation of the procedures and operations for handling personal data must be submitted and reviewed to receive an approval to publish it for general use.

5.4 Technical implementation outline

The last part of the preparation for scaling up learning analytics in blended learning scenarios is the outline of the technical implementation. In this step, the collected requirements presented in the previous section will be organized in a well-known requirements specification protocol. The next step of the technical implementation is to identify the data sources and a proposal for designing a data warehouse which complies with the created regulations regarding processing and protecting personal data. The last part will outline a sustainable architecture for implementing a learning analytics solution which uses the data warehouse, analyzes the data with different analytics algorithms, and makes the results available for delivery and visualization.

5.4.1 Learning platform data

I provide the salient points of how users can access the platform, how is the personal courses' overview for each user, and the building blocks and modules of an individual course on the platform to better describe the setting and how the usage data required for learning analytics is generated and collected. First, the primary users of the learning platform are students and teaching staff (professors, lecturers, teaching assistants, and tutors). Their user activities and interactions are intercepted by the firewall, and copies of the HTTP requests are saved in a raw format into the database for further analysis, inspection, support purposes, and the identification of errors or problems with the learning platform infrastructure and modules. The users can access the learning platform with desktop devices (personal computers, laptops) via web browsers, and by installing applications that can download (synchronize) learning content from courses to permanent storage on their personal computers. On mobile devices (tablets, and phones) the users can access the platform via mobile web browsers, and an RWTH Campus App that also downloads (synchronizes) learning content from courses to their mobile devices. After the successful log-in procedure, the users are taken to their personal dashboard of the learning platform, where they can find their current courses, their courses from previous semesters, an aggregated calendar that shows all the user's events, and other personalized widgets they might have added for a quick overview of their courses. From this location within the system, the most used interaction of the users is to enter in a specific (learning) course room on the learning platform. The course room (Figure 13) itself has six module groups: *Course organization*, *Information distribution*, *Learning resources*, *Assessment modules*, *Collaboration modules*, *and Course settings.*

In the *course organization* belongs the course calendar, where all the important course dates are shown, and any additional dates and events created by the teaching staff. To this group belongs also the customizable course dashboard and the course information page. The teaching staff can personalize the course dashboard with different apps and widgets that show summaries and information for different modules and activities from within the course. The course info page is acts as a syllabus that shows the course summary, description, and course goals; the course lecturers and teaching staff; the course events and duration of the semester; and the course room lifecycle. It summarizes the course organization in terms of teaching and time management.

The *information distribution* group of modules provides methods for information distribution from the teaching staff to the students within a course. For this purpose, there are two modules: Announcements, and Emails. In the **Announcements** module, the teaching staff can post a course announcement with an expiration date for the students within this course. With the **Emails** module, the teaching staff can contact all students, or individual groups within a course, by sending an email directly to them. The students can review both announcements and emails at any given point in the respective modules during the semester.

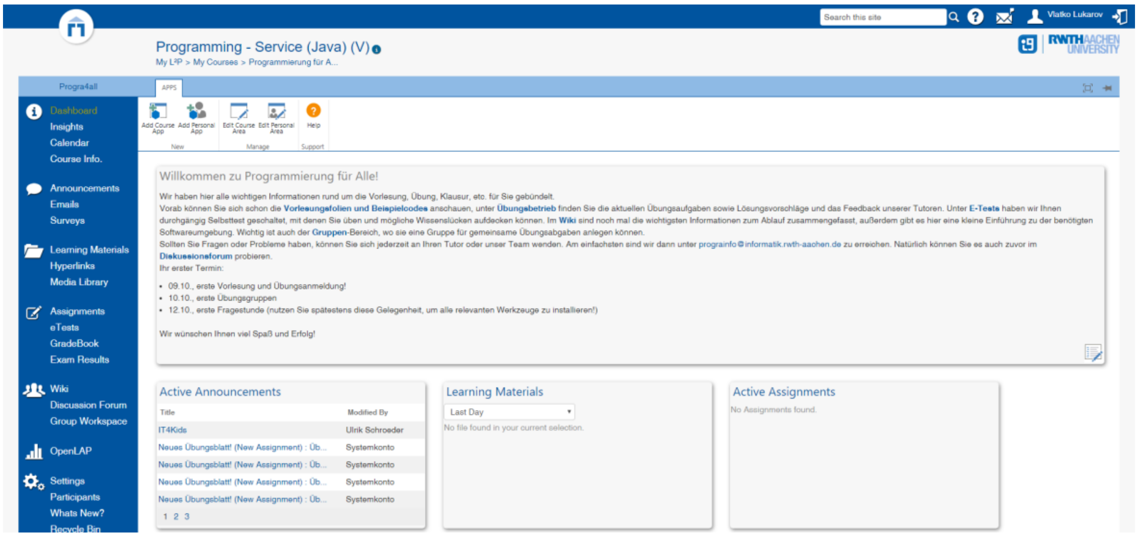

Figure 13. Snapshot of a dashboard from a course on the L^{2}P learning platform

The *learning resources* contain four modules: Learning Materials, Media Library, Hyperlinks module, and Literature. The teaching staff can upload and distribute learning materials and files following a desktop metaphor and behavior. In the **Media library**, they can upload, link, and embed media resources (images, audio files, and videos) created by themselves, or copied and linked from streaming servers, or video sharing platforms. In the **Hyperlinks** module, the teaching staff can link and provide any kind of electronic resource/hyperlink from the internet. In the **Literature** module is used for providing scientific literature and copyrighted publications. Additionally, the teaching staff has the possibility to send requests for digitalization of books, magazines, and other literature resources which are not available in an electronic form, to be provided back to the students. Essentially, the four modules support a wide range and different types of learning resources to support the needs of the users.

The *assessment modules* provide support for formative assessment in the learning scenarios. There is an **Assignments** module which supports online submissions and corrections of the student submissions. The results of these assignments are directly transferred to the **Gradebook**. The teaching staff can use the gradebook for managing the results of offline assignments (on paper). There is also an electronic tests module that provides quizzes and electronic tests for students, whose results can be transferred directly in the gradebook as well. Moreover, there is an **Exam Results** module, where the students can see their summative exam results, before being transferred to the registrar's office for an official grade.

The *collaboration modules* are where the students can collaborate with each other. They can share with each other different learning resources and materials within the **Shared Documents** module. They can discuss different course topics in the **Discussion Forum** of the course and compose and maintain different **Wiki pages** within the course. The last module is the **Group Workspace** where students can form groups, and privately share files with each other, submit assignments together, and organize their own private working group with their teammates.

The last three modules of the platform consist the course room management and configuration (*course room settings modules*). These modules are the **Course Participants**, the **Settings** page, and the **Recycle bin**. In the participants' module, the teaching staff can see all the registered students, invite and remove other people with different roles within the course room. From the settings page they can de/activate specific modules within the course, import learning resources and old content from any other course on the platform, limit the number of group members within a group, and other small tweaks for the course.

The raw data from the usage of the learning platform arrived in the form of seven different parameters identifying a single HTTP request made to the platform (Figure 14). These seven parameters come from the HTTP Protocol definition by the World -Wide Web consortium (Fielding et al., 1999).

Log Time	Client IP Address	Client Agent	Processing Time	Operation	URI	Result Code

Figure 14. Raw data structure from the data collectors

The first parameter is the exact date and a timestamp during the day when was a specific HTTP request generated from the user. The second and third parameters identify the client in the HTTP request. The client IP is the anonymized IP address of the user's device from which this HTTP request originated, while the client agent identifies from which device the HTTP request was made to the learning platform. The fourth parameter is the processing time for each request to the learning platform, or how much time it took to process the action from the learning platform. The fifth parameter is the HTTP operation method, or whether the activity was a simple read/view activity (GET), or it was a create/edit activity in any of the modules of a course on the learning platform. The sixth parameter was the URI or the unified resource identifier of every item/resource/page on the platform. The URI identifies the resource upon which the request is to be applied. In this case, the URI is built in such a way to identify the semester, the course, the module, and the item which was requested or created by the user activity. The last parameter is the HTTP status code, which conveys information how each request was completed (Fielding et al., 1999).

logtime	clientip	clientagent	processingtime	operation	uri	resultcode
2017-11-24 00:08:55.343000	89e291e3-ffff-0000-0000-00000000...	Mozilla/5.0 (Windows NT 6.1; Win64; x64) AppleWe...	15	POST	http://www3.elearning.rwth-aachen.de/ws17/17ws-5397...	200
2017-11-24 00:08:55.280000	89e206a1-ffff-0000-0000-00000000...	Mozilla/5.0 (Windows NT 6.1; WOW64)	265	GET	http://www3.elearning.rwth-aachen.de/_vti_bin/l2pservi...	200
2017-11-24 00:04:05.540000	d4c94f54-ffff-0000-0000-000000000...	Mozilla/5.0 (Windows NT 6.1; Win64; x64) AppleWe...	125	POST	http://www3.elearning.rwth-aachen.de/ws17/17ws-5013...	200
2017-11-24 00:04:05.540000	d4c94f54-ffff-0000-0000-000000000...	Mozilla/5.0 (Windows NT 6.1; Win64; x64) AppleWe...	172	POST	http://www3.elearning.rwth-aachen.de/ws17/17ws-5013...	403

Figure 15. Raw data excerpt showing actual data from persistent storage

For each individual user action, one or more HTTP requests were generated from the user and logged in the raw data collection. Figure 15 shows a data excerpt from the raw data collection. I provide an example to clarify the log data collection procedure. For instance, a student goes to one of his courses on the learning platform and starts watching a lecture video. The system saves the specific point in time as log-time when the student started watching the video; the IP address and the internet browser of the tablet as the client agent; the time in milliseconds it took to process the request to start playing the video; the HTTP operation method GET because the student is simply viewing content; the URI of the video itself because the video is a resource on the platform; and the HTTP status code, that the request itself was successfully completed by the system. These seven parameters are saved as an entry into the raw data.

5.4.2 Software requirements specification

The software requirements specification holds together all of the functionalities and capabilities that the developed system must provide, its main characteristics, and the business (legal) rules and constraints it must respect. The specifications must be detailed and complete concerning the system's behavior under various conditions and outline the expected qualities in terms of performance, usability, stability, and scalability. In the industry, there are already pre-defined and created standards for Software Requirements Specifications (SRS) for software projects (Wiegers & Beatty, 2013). In the context of this thesis, an existing SRS documentation structure is used to document the collected and analyzed learning analytics requirements. Table 12 provides a detailed overview of the SRS structure in the software engineering aspects of this dissertation concerning the requirements. The requirements specifications are used as a basis for the design and implementation of the technical aspects of the interactive system.

Table 12. Software requirements specification structure (Wiegers & Beatty, 2013)

Introduction • Purpose • Project Scope	The introduction identifies the developed product whose requirements are specified in the document including the revision/release number. Here it is also important to provide a short description of the software being developed, how its purpose is connected to the objectives and strategies, and how well does the solution fit within the current context.
Overall description • Product perspective • User classes • Operating environment • Design and Implementation constraints	In the overall description, there should be a high-level overview of the software system, the environment in which it will be used, the users of the system, and the known constraints, rules and regulations concerning the design and implementation of the system.
System features • Indicators and descriptions	The system features outline the functional requirements of the software system. In the context of this thesis, this will be the identification and specification of the different learning analytics indicators that have to be developed within the system.
Data Requirements • Data models • Data management	This section covers the identification, description, and manipulation of the various aspects of the data that the system will consume, process, and create as an output. Therefore, a sustainable data model and data management processes are outlined and described in this section.
Interface Requirements • User Interfaces • Software Interfaces • Communication Interfaces	This section provides description and information about how the software system will communicate with the users, with other systems, and possibly other hardware devices. Also, the specifications for the communication interfaces between the software and hardware systems are described.
Quality Attributes • Usability and usefulness • User Experience • Performance • Scalability	The quality attributes for this project cover the usability and usefulness specification and quality, the streamlined user experience, the performance expectations of the software system and its scalability.

5.4.2 Implementation outline

The implementation of the interactive system that provides learning analytics as a service can be divided into the development of four main components which are independent and interconnected with APIs. Figure 16 outlines the technical infrastructure of the four main components of the system with the basic workflow of the proposed system.

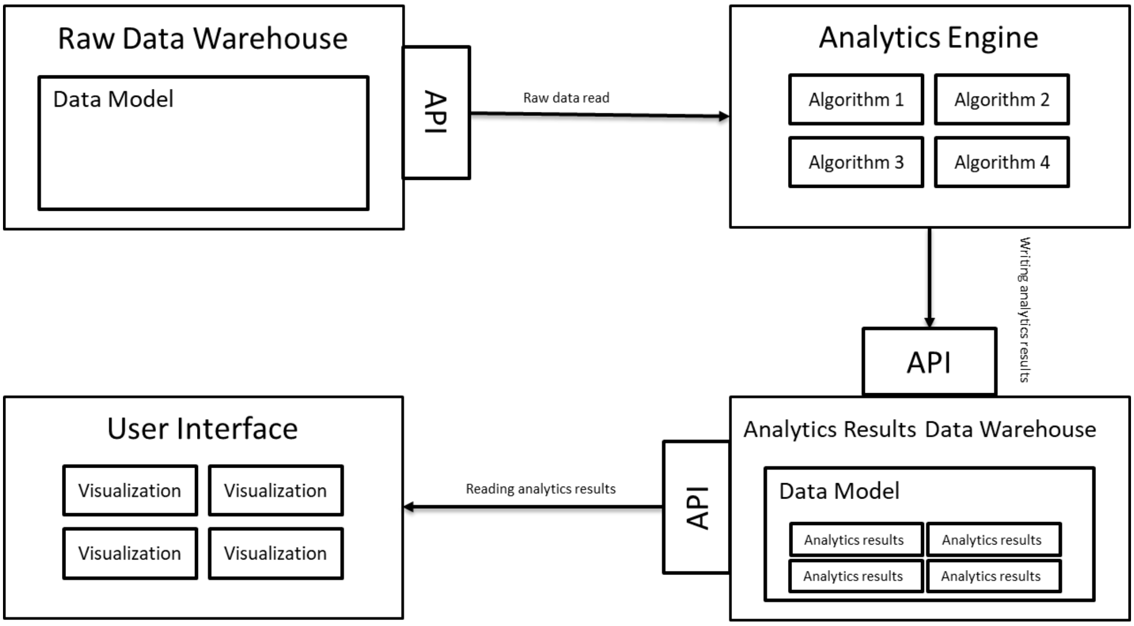

Figure 16. Sustainable learning analytics infrastructure

The raw data from the learning platform is imported daily in the raw data warehouse. Afterward, the analytics engine is triggered, and accesses the raw data from the warehouse, transforms it, processes it with the different analytics algorithms, and then stores it in the analytics results data warehouse. The user interface or the analytics indicators read the analytics results and deliver the data in the form of various visualizations to the users. Physically, the raw data warehouse and the results data warehouse can exist on single hardware or database server, but they represent two different logical units. The four components can be designed and implemented in parallel with underlying technologies approved from the IT department that has to deploy, run, and maintain the system on the scale over a long period of time. Therefore, before starting the implementation of the components, the development team must consult the system administrators about the available technical server infrastructure, the available database systems, the types of the web servers (including operating systems) and the necessary licenses. The use of new, expensive, untested and unproven technologies should be avoided in this usage scenario.

The analytics engine and the visualization techniques need to be modular and extensive. The design of the analytics engine architecture has to be able to incorporate new analytics algorithms, statistical methods, and other extensions, improvements, new features, and updates. The architecture of the different user interfaces also has to be easily extensible so that new indicators can be developed and deployed. It is crucial that the analytics engine and the user interface are completely de-coupled so that changes in one component do not affect the other component.

5.5 Summary

This chapter illustrated all the processes and the undertaken work which is fundamental for scaling up learning analytics in blended learning scenarios in higher education. The elicitation of learning analytics requirements was conducted by using well-established techniques from software engineering. For the two main stakeholder groups (teaching staff, and students) an innovation strategy from the business field of market research was applied to discover the potentials for innovation and development in the context of e-learning solutions. The identified needs and potentials were validated with exploratory data analysis on the log data from the learning platform to confirm their relevance and legitimacy. The conducted semi-structured interviews provided the requirements for the administration and IT staff stakeholder groups. Additionally, a comprehensive literature review though existing research on learning analytics systems, tools and implementations provided the goals and perspectives of the different stakeholder groups according to the researchers and their publications. An extensive list of existing and implemented indicators for the different stakeholder groups was also extracted with the aim to correlate their goals and the currently available indicators. The next step of the elicitation process resulted in the development of personas with representative use cases. As a finalization step, the elicited requirements from all methods were combined and correlated with the personas and a finalized set of learning analytics requirements was compiled and matching indicators were defined for each stakeholder group. The higher education institution created a legal solution and framework that encompassed all systems and services that store private and sensitive data, and the development and deployment learning analytics tools and services is just one part of those services and is covered by it. The university with its internal governing bodies (the university's government, rectorate, and its senate created official regulations (the eLearning Ordnung) for the protection of personal data in multimedia applications and use of e-learning methods at the RWTH Aachen University. The last part of the preparation for scaling up learning analytics as an integral e-learning service was the creation of a sustainable learning analytics architecture that warrants a robust technical implementation.

6 IMPLEMENTATIONS OF ANALYTICS SERVICES

This chapter presents the practical approaches and implementations of the projected sustainable infrastructure (chapter 5, section 5.4.2) and services for learning analytics in blended learning scenarios. Design-based methodologies for the research aspects and rapid application development for the software implementation aspects were chosen as underlying strategies for tackling the research and technological problems at hand. The design and implementation process was intertwined with the formative evaluation of the implementations, the results of these evaluations are presented together with the implementation details and results. The user-centered design process was conducted in several stages and focused on the four identified components of the analytics infrastructure: the design and implementation of the two data warehouses (for raw data, and analytics results), the analytics engine, and the user interface consisting of analytics indicators.

6.1 Rapid application prototyping and development

The learning analytics infrastructure with the learning analytics prototypes was iteratively developed in several development cycles (in parallel), following a rapid application prototyping and development approach. This approach benefited from the collected learning analytics requirements from multiple sources and with its intrinsic approach of minimal planning in favor of rapidly implementing working and functional prototypes of the different components of the interactive software system. The implementation followed an iterative and incremental model which provided the progressive growth of the functionalities and features of the different components of the analytics infrastructure. In between the iterations, different formative evaluation techniques were applied to refine and improve the collected requirements and to refine and improve the developed components of the sustainable analytics infrastructure. The following sections provide a summary of the implementation, the applied practical design decisions, and new findings concerning the technical implementation of the analytics infrastructure.

6.2 Analytics Data Management

The analytics data management covers the management of raw data that comes from the firewall servers, and management of the resulting data that comes out of the analytics engine. Microsoft SQL Server (MSSQL Server) was used as an underlying DBMS technology for both data warehouses. MSSQL Server as the technology was chosen as a technology based on discussions with the "Databases Administration" team from the IT Center at RWTH Aachen University as a long-term sustainable technology infrastructure which can be supported and maintained by their team. One of the biggest concerns when designing the solutions for managing the data was the sheer amount of data generated by the users because the data collection methods worked on the platform level. Hence the collection of traces of user actions and the identifiable different users ranged from hundreds of thousands to millions of events per day. The number of different users on the platform ranged from 12.000 to more than 30.000 users per day. Figure 17 shows historical data for one year about the number of requests/events generated and the number of different users on the learning platform. From October 1st, 2017 to October 1st, 2018, 322.211.812 events were generated with the daily user average of 25.916 different users. This poses a particular challenge because it calls for maintaining another large system infrastructure in parallel with the learning platform, for the sole purpose of storing and analyzing data, coupled with other hardware systems for the reason of providing analytics as a service within the learning processes. This non-functional requirement about a sustainable and "low-cost" infrastructural solution influenced the design and modeling of the data warehouse. I looked over different data models from the research fields of learning analytics and educational data mining to identify a possible model which could, in turn, be used to design both the raw data warehouse and (if necessary), the analytics results in the data warehouse.

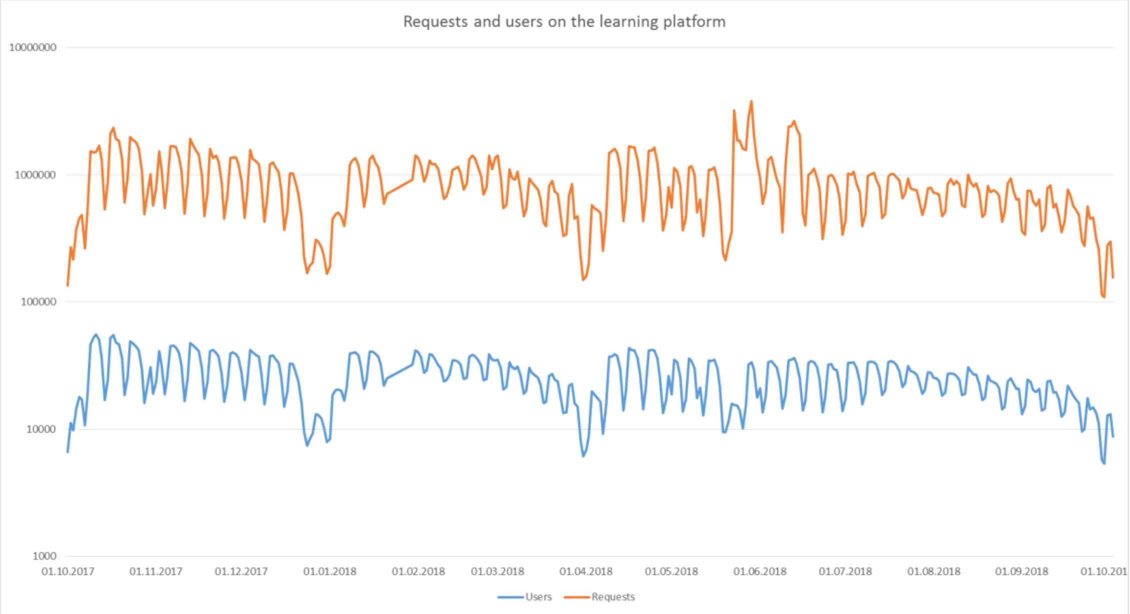

Figure 17. Users and requests on the learning platform in 2018

The nature of the collected data and the organizational structure of the platform with its limited data concerning concretely and consistently identifying an individual over time meant that data models which were centered around a user, or an actor were unsuitable for this purpose. This includes the eXperienceAPI or xAPI, the learning context data model LCDM, Activity Streams, and Learning Registry Paradata data models (Lukarov et al., 2014; Niemann, Scheffel, & Wolpers, 2012; Suthers & Rosen, 2011). I also looked over the Contextualized Attention Metadata (CAM) data model and the IMS Caliper which are event-centric data models that could fit into the scenario at hand. The IMS Caliper is free to use as a solution, but extensive certification processes have to be conducted, and the university as an institution needs to adapt its services towards the IMS Global standards. This requires a profound change and revision of all of the legal and technical solutions concerning e-learning at RWTH Aachen University. Another condition of this is the use, transfer, and collection of personal data and building profiles for the users which is unacceptable by the current university, government, and EU regulations. The CAM data model was the most suitable match for storing the raw data, but if the raw data was transformed in such a way which would lose the connection with the current course room structure. One can extend the data model to fit into this scenario, but that also means that a custom solution must be developed. Additionally, if one considers the rate of event generation by the users would quickly grow in the tens-of-millions events and intelligent partitioning of the raw data would be necessary to analyze it, without the introduction of new and more powerful hardware (without the introduction of brute force and calculating power). In other words, to use CAM data one had to create and generate more overhead data and complexity, so that it can be used as a suitable data model. Hence, after careful deliberation, I chose to use a custom solution that simply partitioned the incoming data into temporal chunks (days) and built a data privacy conformant service that took care of the management of the raw data. Additionally, this service was modular and can include methods that export the data in a CAM format (or any other format).

6.2.1 Raw data management

The raw data was collected from the Forefront Threat Management Gateway firewall managed and used for all IT services provided by the SuB department from the IT Center at RWTH Aachen University. The collected log data was transformed and transferred to a MySQL database hosted on the IT Center servers. The data itself was provided in one big stream, and as such, it was difficult to clean, and then analyze. Hence, I developed a data-privacy conformant strategy to transfer the data to a separate data warehouse and partition it into smaller meaningful chunks. I developed and deployed a scalable database and an automated raw-data import service which checked daily for new data. The raw data import process was straightforward. The service checked if there was new raw data available from the log data database, and if there was, the service split the available raw data into days (one chunk raw data = one day worth of log files), created a raw data table for each available day, and transferred it to the raw data warehouse for further processing and analysis. After the service partitioned and transferred the data, I set up an automated cleanup of the unnecessary data. Since, the available data sets are logs from a web application, they contain many service calls, files, and resources which define the style, color, page layout, client-side code and other web resources that are necessary for having streamlined web experience and when accessed, they also generate HTTP requests, which are saved as entries in the raw data. These resources include stylesheets, JavaScript files, service calls to other web resources, static images, and pages from the learning platform, logos, system files. I removed them from the raw data, and the data clean up resulted in data logs about user read/view and create/edit activities. Once the cleanup service removed all the service calls, it was ready to be processed and analyzed by the analytics engine.

To provide the raw data to the analytics service I created a small RESTful API that provided the raw data in a fast, standardized and uniformed way to the analytics engine. The design decision behind this approach was that if in the future the structure of the raw data changes, it would be straightforward to change/update the structure in one place. This decoupling of the raw data API and the analytics engine provided modularity and decreased the risks and problems in the raw data warehouse influencing the analytics engine, and the user interface.

6.2.2 Results data warehouse

I reviewed several data formats and models as potential candidates in which the analytics results could be stored and accessed. The data format should incorporate information about the learning and teaching activities on the entire learning platform, of each individual faculty, the different departments within each faculty, the teaching and learning activities within individual courses, individual modules within a course, and potentially, individual user actions. In addition, a small configuration dataset should provide permanent storage for configurations to keep track of the analyzed raw-data; to manage the execution of the data privacy methods which model the data expiration protocols; store telemetry data for diagnostics and successful execution of the different analytic methods. The data model had to be fast, scalable, and extensible while supporting many concurrent read operations during the day. Additionally, the update of data from the analytics results should not have any side-effects on the existing data and results. Figure 18 outlines the data flow process which starts from user-generated activities and ends with results in the results data warehouse.

The resulting data warehouse that stored the results was read optimized column-oriented database structure to support the large-scale and data-intensive applications that manipulated and displayed the data in the different indicators. I modeled the learning environment as a virtual learning space with two unique identifiers: the course room id (Lehrveranstaltungsnummer) and the date. These identifiers were used both as keys for accessing relevant data information out of the columns in which the analyzed data was saved and as a discovered/realized data from the raw data. This duality (serving both as data, and identifiers) saves a lot of redundant read-operations, decreases loading unnecessary data in the main memory by loading only the relevant data based solely on the given unique identifiers while doing many concurrent read operations from different datasets/tables.

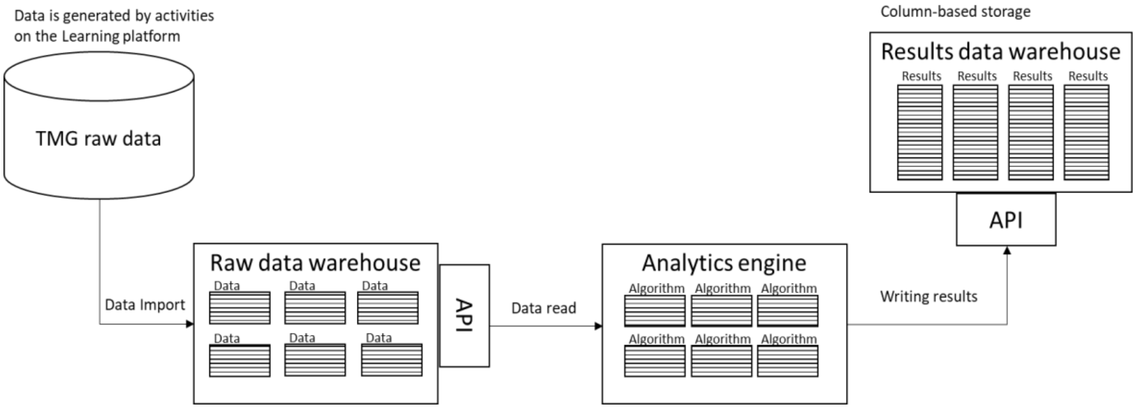

Figure 18. Data management components from the infrastructure

There is no central element within the results data, and instead of building a new relationship diagram, the structure of the course organization was simply borrowed from the campus management system. The borrowed course structure can be seen in Figure 19.

Faculty	Institute	Semester	LV Number	Course Title	Course Type

Figure 19. Campus management system course structure and information

Each course has a unique identifier, title, and type. One course can belong to one or more institutes, and one or more faculties. The business logic that retrieves this meta-data and the actual analytics data was incorporated in the CRUD RESTful API that managed all of the Create, Read, Update, and Delete activities within the data warehouse. The column-oriented database had three main abstractions: platform activities, course-based read activities, and course-based create activities. All of the data tables that stored the analyzed data were created specifically with built-in column store indexes with an updateable non-clustered index which performed well on large tables and scaled well as data tables that store analytics results grow over time. Practically, the analytic engine and the warehouse was designed in such a way, so that summaries and aggregation of all activities within each individual course on the platform were stored and analyzed. This made the activation and provision of the analytics results in every single course a straightforward task because the analytics results were already available in the warehouse. These three abstraction levels were designed to be extensible and one could easily update the warehouse with new structures, and new data without affecting the existing data and results.

Platform activities is an abstraction level that stored analytics results concerned with summarized learning activities on the learning platform. The unique identifier for this abstraction level was the date on which all of these events happened on the learning platform. These activities summarize statistics about how many users were on the platform on a given day; whether the users have used a desktop or mobile device including distribution among different devices, and aggregation of activities within the different learning modules on the learning platform on a given day.

Course-based read activities is an abstraction level that stored analytics results concerning all of the read activities in every single course on the learning platform. Read activity, in this case, was when a user has opened an individual course (either on her desktop computer, her mobile phone, or has used an app which uses the API of the learning platform) on the learning platform and has accessed, read, or downloaded learning content from this individual course. The unique identifiers for this abstraction level were the course id and the date. The analytics engine identifies different users (for this particular course), their activities on a given day, the devices from which these activities were initiated, and an aggregation of the read activities within each module within each course are saved within the data warehouse. Additionally, the different identifiable learning resources are also identified, and the interactions with them are also analyzed and stored in relation to their course and their module of origin.

Course-based create activities is an abstraction level that stored analytics results concerning all of the create activities in every single course on the learning platform. Create activity, in this case, was when a user (member of the teaching staff, or a student) opened a course room and created content (whether she created an announcement, or contacted the students via email, opened a new discussion, or edited a wiki page within the course). The learning platform has pre-defined roles and permissions structure, which allowed to explicitly identify what was a teaching action (teaching activity), and what was a collaboration activity (or in some cases student activity). The "roles and permissions structure" was also incorporated in this abstraction level (and also in the RESTful API) which enabled the development of visualizations that can correlate teaching activities with student activities. Again, all of the create content activities were identified by different users, their activities on a given day, the devices from which these activities were initiated, and an aggregation of the activities within each module for each course were saved within the data warehouse.

The downside of this data management strategy was that it was "strongly coupled" with this particular learning platform, and this particular campus management system. On one side, this solution was highly optimized for the learning scenarios at RWTH Aachen University and its faculty and course structure. But one would need serious preparation and comprehensive understanding of the organizational structure of another university and its applied e-learning scenarios, to transfer or re-use this data warehouse infrastructure.

6.3 Analytics engine for data analysis

The second component of the analytics infrastructure was the development of the analytics engine which consisted of smaller optimized algorithms and methods which processed the raw data and produced analytics results. In principle, there are two ways how the analytics engine can handle the analytics process and produce results that can be displayed back to the user: real-time data analysis (streamed data processing) on demand, or do batch data analysis (batch data processing). The first way will always deliver the latest data and analytics results back to the user, at the expense of being overly complex and "resource hungry". The batch data processing, on the other hand, will not always provide the latest data and analytics results but can be designed to be an extremely efficient way to process large amounts of data at a comparably smaller operational cost. Additionally, it provides freedom and control to decide when (and the time intervals between two batches) the data analysis and processing can start and how much resources can be allocated to the batch analytics.

Considering both approaches regarding the applied e-learning scenarios within the learning platform, and the number of active courses per given day (Figure 20). On daily basis, there more than three thousand courses which have different learning and teaching within them. On some days, the numbers reach more than nine thousand courses.

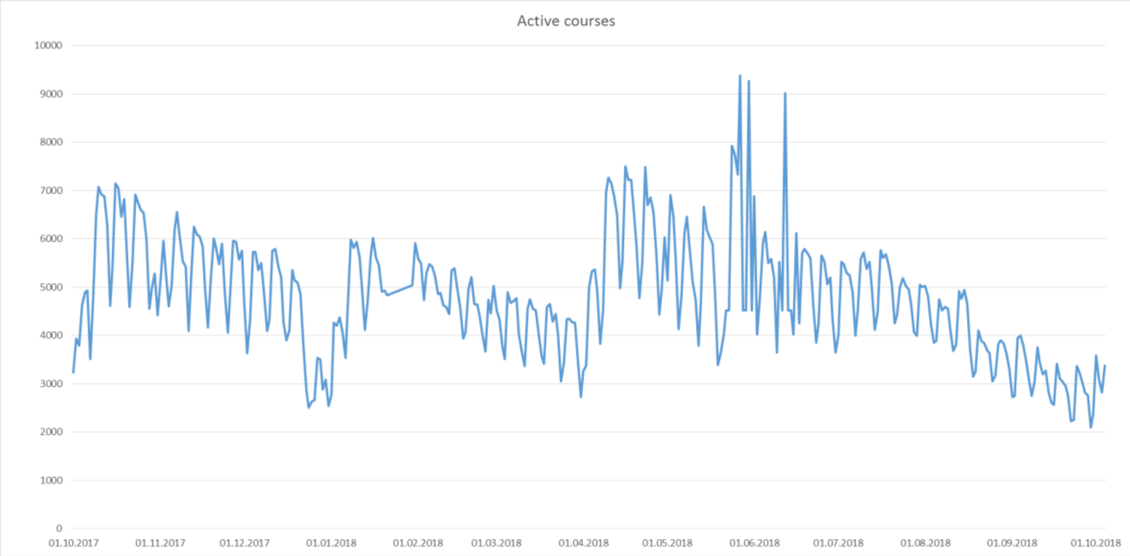

Figure 20. Number of active courses per day on the learning platform

If one makes an assumption that only in 10% of the active courses that analytics results will be used concurrently, that means that in 300-1000 courses altogether tens of millions of events would have to be aggregated and analyzed. On the other hand, based on previous experiences and previous experimental results, there is no real necessity for real-time analytics, but there is a necessity to deeper analyze the results after they are presented back to the user (Anna Lea Dyckhoff, 2014). Moreover, in the collected and elicited requirements for the different user groups, there was not an explicit requirement about having real-time analytics. Considering all these arguments, I opted for batch processing of the data with the analytics engine. However, for data exploration purposes and comparison I retained features that allowed for loading and manipulating the results data on-demand. The analytics engine was developed and deployed as a background service which was triggered automatically after each successful import of newly available raw data. It was developed in C# and .Net and uses the Windows Communication Foundation and Web API services for providing data and communication interfaces between the different components of the analytics infrastructure. The process itself consisted of three steps: (1) getting (loading) the raw data from the data warehouse in the main memory, (2) analyzing the data with the analytics engine, (3) building and formatting the analytics results and saving them in the results data warehouse.

(1) Loading raw data

As mentioned in the previous section, the amount of raw data and user events amounted to millions of events (or on average, a couple of GBs per day) which posed a challenge of how to deliver (parts of) the raw data to the analytics engine and its algorithms and methods for analysis. Querying and transferring large amounts of data is expensive in terms of memory consumption, disk space (in relation to the transaction logs and other logging methods that ensure the health and integrity of the data), it takes time to query and search for an exact data set or data chunks especially when querying for concurrent method calls that run in parallel, and it takes considerable time to transfer large chunks of data across a network. The data delivery methods for raw data were optimized with two strategies: by using stored procedures, and by using parametrized methods that delivered the same raw data set only once for multiple analytics methods and algorithms (the same strategy was used in the results database). Stored procedures are prepared SQL statements (or code) which are only compiled once and stored in an executable form which makes them very efficient. Additionally, they are cached and there is almost zero invocation overhead and can include multiple SQL statements which can be executed with a single call and decrease the number of different (new) database operations. The parametrized methods used the stored procedures to load the needed (and suitable) raw data by selecting data which was needed for each analytics method and indexed it in the main memory. If some additional querying or manipulation on the data was necessary, it was executed while the data was in the main memory because the main memory was faster by several magnitudes compared with the disk drives. Figure 21 sketches the execution flow and the parallel execution of the threads while analyzing the raw data. In earlier versions of the raw data warehouse and the data loading methods, this was done with multiple SQL statements which put extra load and time for delivering the raw data, but with subsequent optimizations this raw data loading strategy was disregarded because it was considerably slower and more resource-hungry[3].

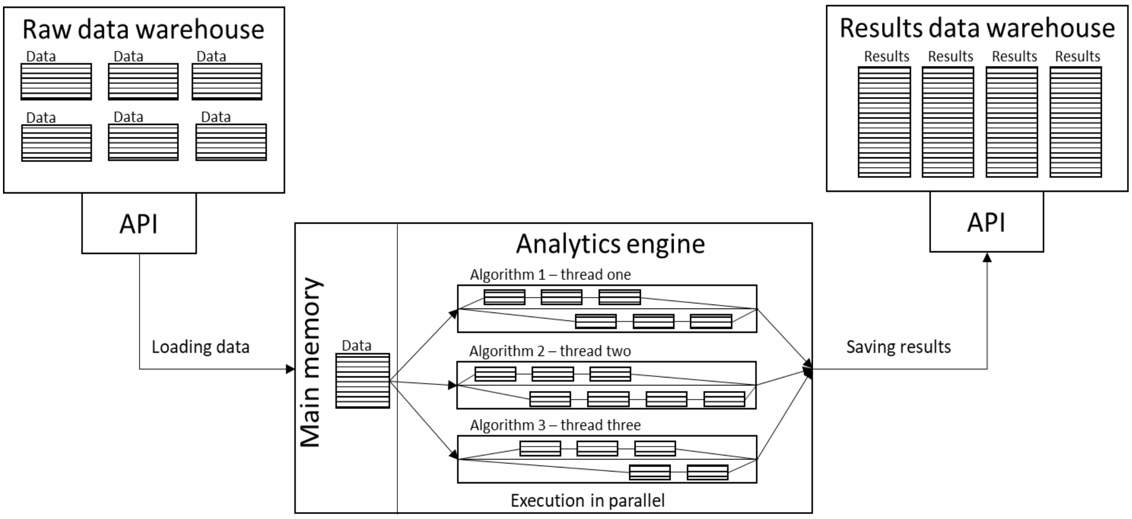

Figure 21. Parallel execution of the analytics engine service

(2) Analyzing data – the analytics engine

After the raw data was loaded in the memory, the analytics engine initiated its analysis. For each corresponding abstraction level from the results data warehouse (platform-based activities, course based read (view) activities, and course-based create activities), a set of analytics methods and algorithms were implemented. The algorithms and analysis methods for these three abstraction levels were executed in parallel while sharing the raw data as input. The analytics engine was, in essence, a background service which used (or rather re-used existing) multiple threads to analyze the raw data in parallel. This multi-threaded approach enhanced the performance and throughput of the code, thus reducing the time necessary to analyze the data. Each thread had a data-loader method which loaded a chunk of raw data and provided it to another method which called upon further optimized and parametrized static methods which analyzed only this chunk of data and

[3] This approach can be used for prototyping and development, or on systems that do not have a lot of RAM memory available, because the bulk of the analysis and loading is divided between the different chunks of data delivered by the SQL queries which run on the disk drives.

returned a result (either intermediary or a final result) which was later re-used, or saved within the results data warehouse. Whenever a thread has finished execution for the current data chunk, it picked up new data and continued analyzing the data. If, by chance, the thread failed to execute, or encountered an error, the intermediate results were discarded, the error was saved, and another thread tried to re-analyze the raw data chunk. The analysis of the data was executed until all of the available raw data was analyzed.

An analytics helper class was designed which contained the optimized parametrized static methods which analyzed the data. They were shared across the different execution threads. The optimized worker methods looked for specific patterns within the data with the help of regular expressions, pattern matching, and text analysis. These static parametrized methods were refactored in several iterations with the purpose of performance optimization and the introduction of performant data structures such as dictionaries, hashsets, hash tables, and indices. The refactoring greatly optimized the performance and sped up the throughput of each static method, especially when they were coupled with one-directional data streamers. These methods were organized in modular and extensible units which made it straightforward to maintain and upgrade them accordingly without changing the overall logic and structure of the analytics engine.

(3) Saving the results

After the analytics engine finished with analyzing the loaded data, another method organized the obtained results to prepare them for the Results data warehouse. Initially, they were updated individually (or in-between calculations) to increase their consistency and reliability, but the saving process also took a lot of time. In later iterations and revisions, the results were formatted in the column-friendly fashion of the tables within the results data warehouse, and they were inserted as a bulk with selected update SQL statements (via a stored procedure). The results from the analytics engine and the data analysis are used as basis for the prototypes and the user interface evolution.

6.4 User interface evolution and prototypes

The last part of the sustainable infrastructure was the design and the formative evaluation of the user interfaces intended for the stakeholders. The user interfaces consisted of learning analytics indicators which represented and visualized the learning analytics results obtained from the raw data through the analytics engine. The user interface was developed with html5, CSS, JavaScript which included the Highcharts[4] charts and data visualization library. The results were delivered by a set of methods from the Web API developed for the data warehouse that stores the analytics results. The different user interfaces are created as different single page applications which run on the same web server and have access to the analytics results. The pages that contained the analytics indicators and were intended for different stakeholder groups, were created as standalone, but they could also be embedded in other systems. For students and teaching staff stakeholder groups, the interfaces were provided within the course rooms on the learning platform. In the following sections, I summarize the development and structure of the results API, the development and evolution of the user interfaces resulting from the collected user requirements, and the results of the conducted formative evaluations of the user interfaces.

6.4.1 Web API design

The designed and implemented RESTful Web API was designed with multiple routes for handling traffic and data for the different stakeholders. Each stakeholder group received a distinct route and each route depending on the stakeholders' needs contained a set of publicly exposed endpoints (Although, an API route can be defined without any endpoints. The number of routes

[4] https://www.highcharts.com

does not influence the performance of the server-side logic or code). The API endpoints bridged the persistent storage that stored the analytics results and the user interface, because they specify the location of the resources with unified resource identifiers (URIs) to which HTTP requests are dispatched, and an appropriate response is delivered back to the client. The endpoints delivered the data in a pre-defined JSON readable format which was natively supported by the visualization library (Highcharts).

Course Analytics API route: This route delivered analytics results for the teaching staff stakeholder group. Figure 22 depicts the pattern in which the HTTP requests should be constructed and dispatched to the server.

```
[Host URL]/Course API/ {Controller Name}/ {Course ID}/ {Parameter 1}/ {Parameter 2}
```

Figure 22. Course API routing structure pattern

The first parameter (Course API) identified the route to which the request should be directed to. The second parameter (Controller Name) identified the controller that handled a particular data model. In this case, for each module that is available in the course room on the learning platform, I designed a separate data structure and controller. The third parameter (Course ID) was used as an id to identify for which course room to load analytics data for the particular module. The fourth and fifth parameters were designed as wildcards parameters. Depending on what kind of value they held (dateTime or keywords), they were used as additional filters to further filter the analytics results. The fourth and fifth parameter within in the request/path was optional.

Platform Analytics API route: This route that delivered results for the stakeholder groups of the administration and the development team. Figure 23 depicts the pattern in which the HTTP requests should be constructed and dispatched to the server.

```
[Host URL]/Platform API/ {Controller Name}/ {Faculty}/ {Department}/ {Semester}/ {Course Type}
```

Figure 23. Platform API routing structure pattern

The first parameter (Platform API) identified the route to which the request should be directed to. The second parameter (Controller Name) identified the controller that handled a particular data model. In this case, I identified different domains and areas of interests of the platform and organized them into different data structures and controllers. Additionally, for each module available in each course I also developed a separate data structure and controller. The third parameter (Faculty) was used as a filtering parameter to load data from courses that belong to a specific faculty. The fourth parameter (Department) was also used as a filtering parameter to load data from courses that belong to a specific department from a specific faculty. The fifth parameter (Semester) was used as a filtering parameter to load data from courses depending on their semester. The sixth parameter (Course Type) was used as a filtering parameter to load data from courses that are from different types (lectures, labs, seminar, etc.). The third, fourth, fifth, and the sixth parameter within in the request/path were optional parameters.

6.4.2 Analytics indicators visualization strategies and development

The interfaces of the different developed prototypes consisted of different learning analytics indicators. The first implementation of the indicators was based on the requirements analysis and the matched indicators with personas described in section 5.2.8 from chapter 5. The data visualizations strategies were based on knowledge derived from the work of Iliinsky and Steele (2011), and the work of Abela (2014) which provided practical guidelines and concrete suggestions about visualization, and the visual properties of the data that needed to be encoded. Abela developed a mapping strategy that put the charts and visualizations into four categories depending on what a specific chart should accomplish. The chart/visualization can show the *relationship* between data; the chart/visualization can show a *comparison* between different data entities; the chart can provide a *distribution* of the data over temporal or spatial properties, and lastly, the chart can show the *composition* of the data. The mapping strategy is visualized in Figure 24 can be used as a chart selector guide to identify what kind of chart is most optimal for visualizing the data for a particular analytics indicator. The developer has to consider the different characteristics of the dataset at hand, the number of dimensions and variables for each indicator, and whether there is a temporal or spatial aspect of the data (Abela, 2014).

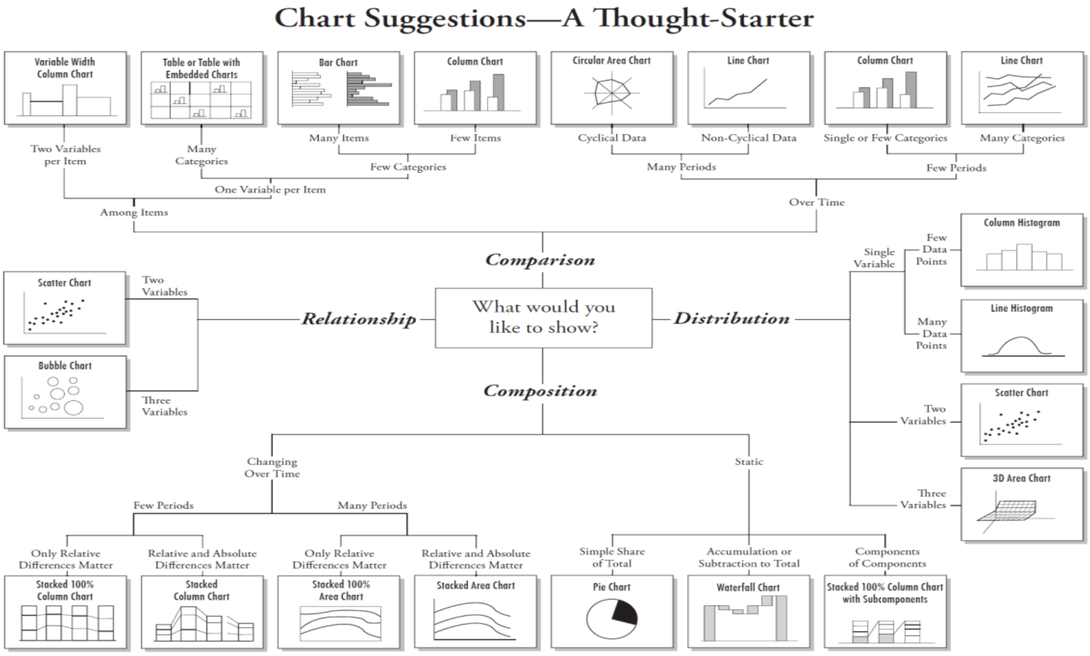

Figure 24. Workflow to select a suitable chart for information visualization from (Abela, 2014)

The visualized data has different visual properties, depending on the nature of the visual properties of the data that needs to be visualized. The data properties of the analytics results were considered and included in the final design of the selected charts because many visual properties can be used to encode several data types, as can be seen from Figure 25. The resulting analytics data in most cases had temporal characteristics, and special attention was placed on the position, the layout, and the axes. In the context of the implementation work in this thesis histograms, bar charts, line graphs, time series, pie graphs, treemaps and heat maps were used as comparative and quantitative formats for the visualization layouts (Iliinsky & Steele, 2011).

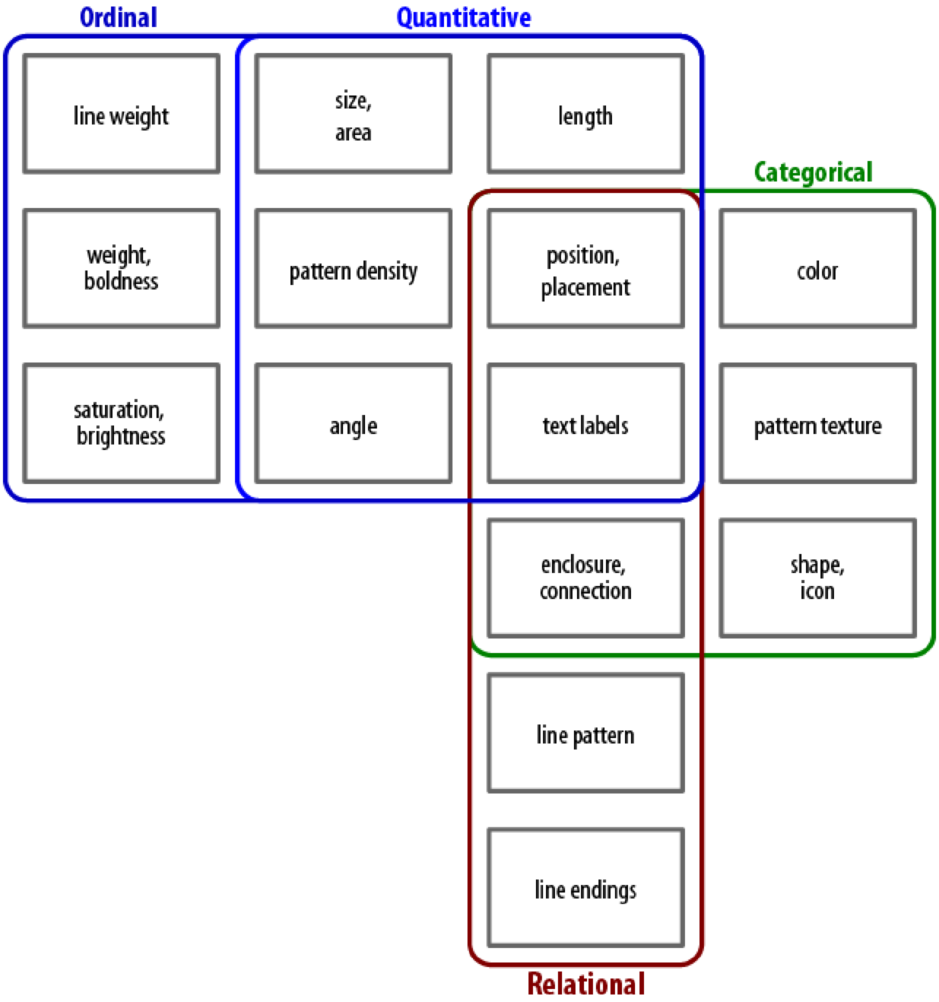

Figure 25. Visual properties grouped by the types of data they can be used to encode data (Iliinsky & Steele, 2011)

6.4.3 Insights learning analytics prototype (v1)

The user interface provides a set of indicators which visualize the learning data for the scenarios and goals of the users belonging to the stakeholder groups of teachers and students. The learning analytics prototype was designed to automatically detect the role of the user within a course room and adjusted the learning analytics indicators accordingly. The dashboard page was designed to be responsive and adapt to the screen size of the user, and the prototype itself was developed to be capable of being embedded on the learning platform. This characteristic paired with the automatic detection of the users provided a simple way to control access and the delivery of analytics results to specific course rooms on the platform (instead of providing it for the entire learning platform). The dashboard for both the teachers and student contained up to ten indicators. The visualizations in the prototype were developed to be interactive (with filtering and zoom-in functionality) and enabled the user to focus on specific parts of the visualization. The reason behind this split can be inferred from their goals and perspectives because the two user groups have similar, or overlapping goals, they have different perspectives. The indicators and the visualizations were based on the requirements and guidelines with regards to the questions they answer, the data analytics literacy of the two user groups, and the visual properties of the data with the appropriate visualizations. Figure 26 shows the “Insights” learning analytics prototype seamlessly integrated within a course room on the learning platform. This version of the prototype was also used for a semester-long case study evaluation with around fifty courses at RWTH Aachen University.

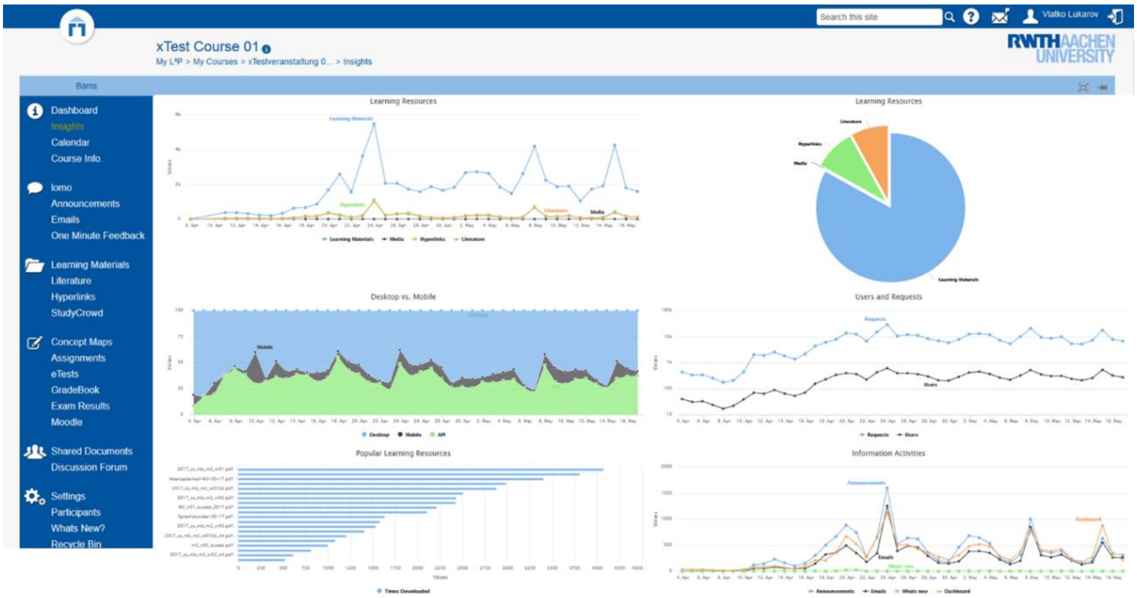

Figure 26. Insights v1 embedded in a course room on the learning platform

The visualizations in the prototype were simple and responsive, thus the cognitive effort upon the user was low, and she could concentrate on the visualizations themselves, instead of concentrating on understanding the interface. The interactions with the visualizations were always coupled with informative tooltips (bound to mouse-hover events) to provide feedback about what is displayed at a given time and place.

Example indicators

In this part, I provide examples of indicators and their interpretation which have been implemented in the “Insights” learning analytics prototype. Figure 27 contains three indicators from an individual course room intended for the teaching staff. The first indicator (called Learning Resources) is a two-dimensional line chart which represents the behavior of the students within the course room modules responsible for distributing the learning resources to the students. The goal is to provide an overview to the teaching staff about when and how the students show up/interact with the learning resources in the course room, and whether the student activities comply with the premise of the teaching staff about continuous learning. Similar indicators were built for the different modules for information distribution to the students, the different modules for formative assessment, and for the different modules that provided collaboration tools. The indicators which visualized the different collaboration activities within the modules were also provided to the students as well, to provide an opportunity to gain a quick overview (and act as a small nudge to motivate them to participate in these collaborative activities).

The second indicator (Desktop vs. Mobile) uses an area chart coupled with time series to show with what kind of devices the students use to access the platform (and potentially learn). The two main devices are desktop and mobile. The third aspect is using synchronization applications which use the API of the learning platform to directly synchronize learning resources from the course room onto their computers. The idea behind is to see whether the teaching staff needs (or should adapt) their learning resources so they are more compatible with the devices on which the students use them. Moreover, if many students do not go to their course room through the web interface, but only synchronize their learning resources directly, they might be missing out on some teaching activities in the other modules on the learning platform.

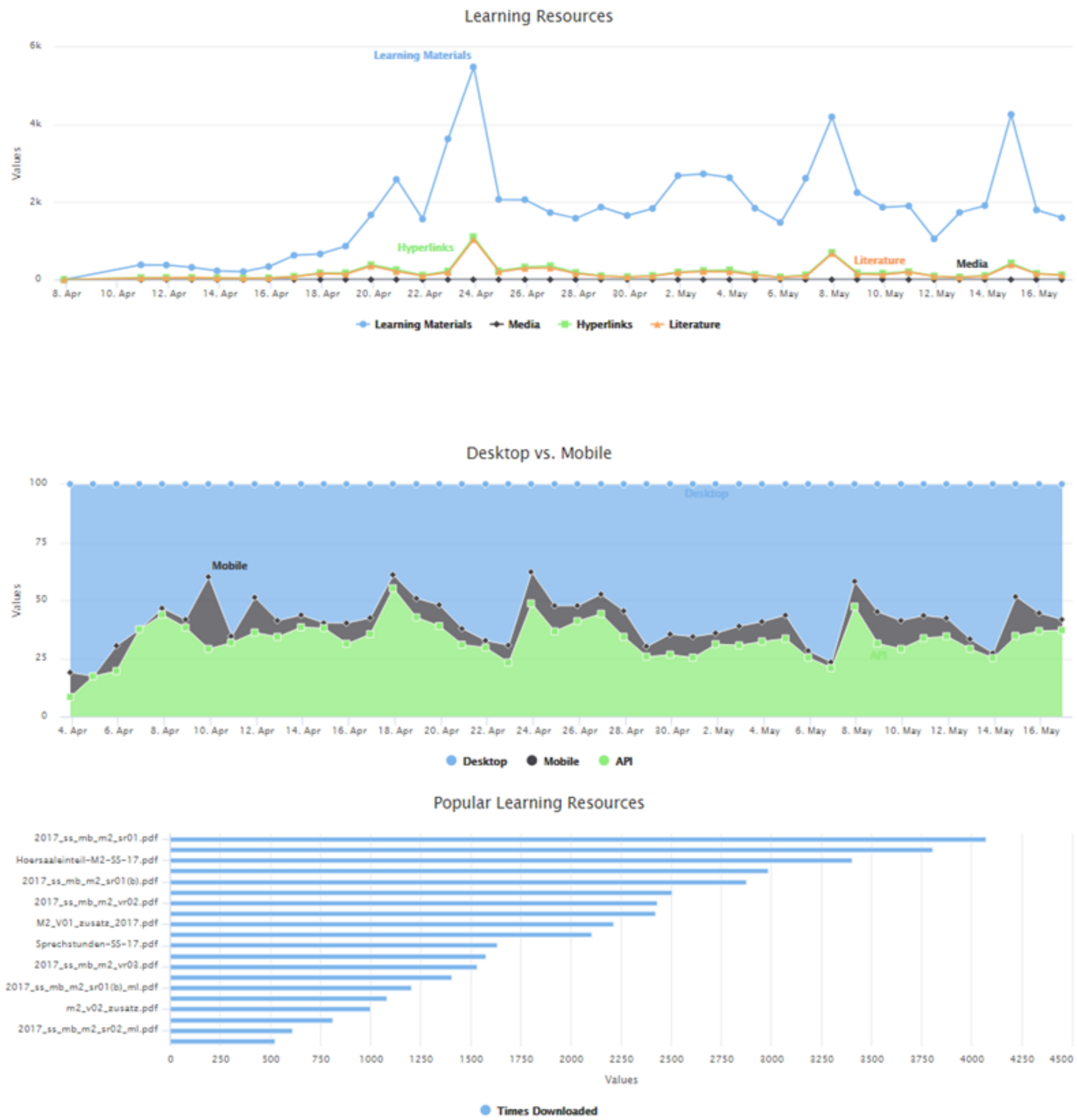

Figure 27. Three indicators displaying results about learning resources and desktop and mobile distribution

The third indicator (Popular learning resources) uses an ordered bar chart to display which of the provided learning resources within the course room are the most used/most downloaded by the students. This indicator does not differentiate between the different types of resources, just shows the most used learning resources (with names) in descending order. The goal is to provide to the teaching staff an overview which particular learning resources are most used (and in turn), they can identify which ones were not used, and if necessary, develop an appropriate strategy for intervention. Another example of an indicator intended for the teaching staff is represented in Figure 28. The visualization is a logarithmic line chart that shows the correlation between a number of students (users) and the number of requests they generate within the course room. The nonlinear logarithmic scale of the chart was necessary to show the correlation between the number of students and number of requests because there is a large range between the two quantities on a given day. The goal of the visualization is to show the number of students that show up in the course room and conduct different learning activities (or interact with the different modules and learning resources) during the semester. This indicator is also provided to the students to show them how many of their peers are coming to the course room on the learning platform.

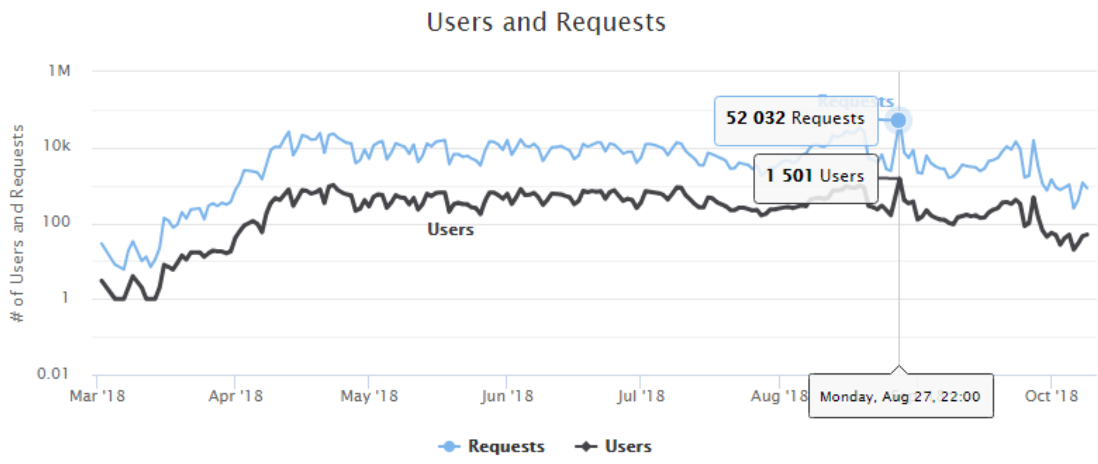

Figure 28. Number of users and their requests in a course room

The next indicators were developed for in the students' dashboard prototype and were included in it. Figure 29 shows three indicators which used formative assessment information to show the amount of work the students need to invest in the assignments so that they can qualify for the exam or see their progress through the course. Three visualizations were developed for the first indicator to show the number of assignments each student must submit, in relation to the overall number of mandatory assignments. In the following prototype iterations (and with feedback from the evaluation), the third visualization was selected (Gospodinova, 2018).

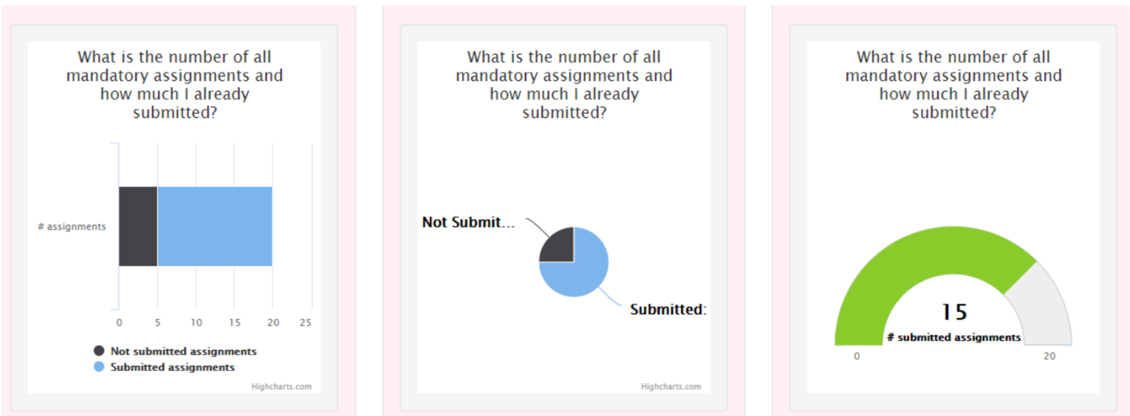

Figure 29. Indicator prototypes for formative assessment

The next indicator available for the students was the one that shows the comparison of the total amount of time spent on learning activities within the course room, with the average of the rest of the class. The class average was taken, and two additional indicators were developed where they showed the students' time spent with the median and the mode per month during the semester. Figure 30 shows three different visualizations for the same indicator by using a line chart, area chart, and a column chart. In the following prototype iterations (and with feedback from the evaluation), the column chart was selected (Gospodinova, 2018).

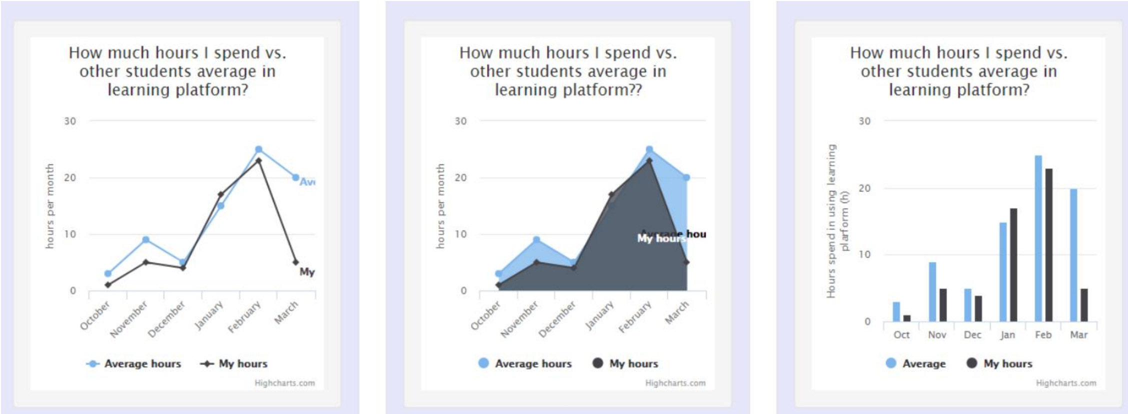

Figure 30. Indicator prototypes for effort and activity correlation

Formative evaluation

The “Insights” analytics prototype was evaluated with users from both stakeholder groups (students and teaching staff). The received results of the formative evaluation can be examined into details in the works of Mentiu (Mentiu, 2018). I provide a summary of the salient points of the evaluation she conducted during her master thesis. Overall, the two dashboards from the “Insights” analytics prototype were used in controlled experiments with 21 users. 18 of the users were students (bachelor and master students), and three teaching assistants. The experiments followed the think-aloud evaluation strategy in which the user tried to verbalize her intentions and interactions with the interface. The entire user sessions followed a predefined structure and the evaluator led each user through the entire evaluation session. The participants received a small introduction about the prototype, its goals, and purpose, and informed the participants that the session will be recorded (audio, video, and a screen capture of the interactions). The third part of each session was a usability questionnaire followed with a small informal discussion about the prototype and the study (Mentiu, 2018). The purpose of the informal discussion was to relax the participants in the evaluation, and this was used as an opportunity to gather personal opinions and qualitative feedback. The evaluation with the students was divided into two parts because after the first ten sessions the raw feedback received from the users was getting repetitive. The evaluator decided not to continue with the evaluations and focus on improving the user interface and the indicators themselves. After the feedback was analyzed, the user interface with the indicators was updated, and the second part of the evaluation with students was conducted (Mentiu, 2018).

The analyzed results from the conducted evaluation fell into two main categories: Functionality and Utility, and User Interface improvements. When evaluating with students, one major discovery was the low level of data literacy and understanding of the visualized data. Additionally, many of them stated that they did not consider analytics as something they would use on regular basis during their studies (Mentiu, 2018). Another potential problem was the dashboard with too many indicators in one place. The users were simply overwhelmed with all of the available indicators and requested some help and documentation to help them understand the user interface and the analytics indicators. Additionally, when designing for users who are not experts in the field, most of them would not discover hidden interaction possibilities or interface elements that can be manipulated. Hence, the developers should provide explicit visual cues and hints that additional interactions are available with the visualizations and data. The most prominent request from the feedback was the possibility to drill-down even further into the data and see more correlations with other activities and more details about the visualizations (Mentiu, 2018).

Concerning the user interface improvements, the feedback showed that the users appreciated the use of icons and buttons which have a familiar design from other well-established interfaces (e.g. Close button for a window from the interface of an operating system). Another major point was the use of language, labels, and descriptions within the interface because the user perception of the language varied across the different participants in the evaluation (Mentiu, 2018). When designing an analytics tool for a wide audience, the designers and developers should consider that there is a major difference in vocabulary, and knowledge among the different users from different fields of study. In other words, no technical language that can be understood only from domain experts, or computer science students. Considering the visualization strategies, the feedback from the evaluation provided conflicting data visualization strategies. There are users who are very comfortable with pie-charts, while there are users who would prefer bar charts. One way that can solve this problem is to provide the user with a possibility to choose her preferred visualization style within the analytics indicator (whenever possible) (Mentiu, 2018).

6.4.4 Insights learning analytics prototype (v2)

The second version of the "Insights" learning analytics prototype is an updated version of the previous prototype, with the notable difference, that this prototype was explicitly updated and intended just for the teaching staff stakeholder group. This decision was to bind the scope of this Ph.D. thesis regarding the evaluation and validation of one stakeholder group. The complete development, deployment, and evaluation and validation for all stakeholder groups is a tremendous undertaking which is well beyond the scope of a single dissertation. The second version of the "Insights" learning analytics prototype automatically detected the role of the user and was also responsive and adapted to the screen size of the user's browser. The visualizations were based on the requirements of the teaching staff, and they were updated based on their feedback from the case study and the feedback and results from the conducted formative evaluations. One notable difference compared with the previous prototype was the introduction of "correlation indicators". These indicators visualized correlations of the teaching activities and their effect on the students' online behavior regarding these teaching activities. Another difference in this version of the "Insights" prototype was the number of available indicators. Namely, the number of presented indicators was dependent on the course room setup and the activated modules within the course room. This means that courses that used videos as part of their learning scenarios would receive indicator(s) concerning the lecture videos; if in the course there are assignments, indicators concerning the assignments would be provided. Hence, one user can have different looking analytics dashboards in different course rooms. The new indicators and their visualizations in the prototype were also simple and responsive and as such, they kept the cognitive effort upon the user low. This means that the user could concentrate on the visualizations themselves, instead of concentrating on understanding the interface. The interactions with the visualizations were always coupled with informative tooltips (bound to mouse-hover events) to provide feedback about what is displayed at a given time and place.

Example indicators

The structure of the user interface of the second version of the "Insights" analytics prototype remained very similar to the previous version. Only the indicators themselves were updated, or new indicators were developed. The first indicator from the new "correlation indicators" is the indicator presented in Figure 31, that shows the correlation of student activities with regard to a certain teaching activity from the teaching staff. A teaching activity can be an upload or edit of learning resources (for example lecture slides, or lecture videos), sharing hyperlinks, or distribution of scientific literature to the students. Basically, any explicit teaching activity within the learning resources modules on the learning platform. These activities are represented as columns in the chart, while the student activities are displayed as a continuous line to make it easier for the user to detect trends in the students' usage behavior. The goal of this indicator to show the relationship between the teaching activities and students' activities, and possibly show the teaching staff the influence they have on the students' motivation and their interaction with the provision of learning resources.

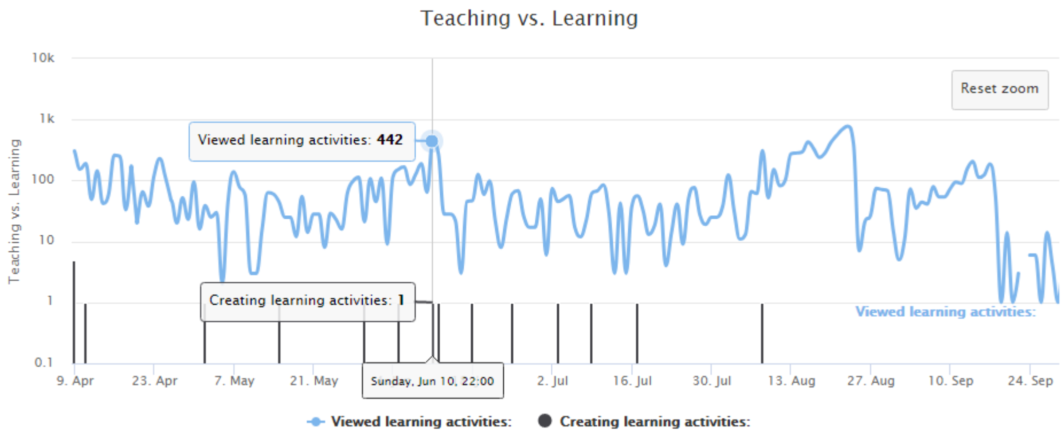

Figure 31. Indicator showing a correlation between teaching and learning activities

Similarly, the next indicator in Figure 32 shows the students' interactions within the information distribution modules, with regard to the teaching activities which distribute information. The teaching staff can distribute information back to the students about upcoming course activities, events, and change in the course by posting a course announcement, or they can contact the students by sending an e-mail from the learning platform. Here, like in the previous visualization, the teaching activities are shown as columns, while the student activities are displayed as a continuous line to make it easier for the user to detect trends in the students' usage behavior. The goal of this indicator is to show whether the students open and receive the information in a timely manner.

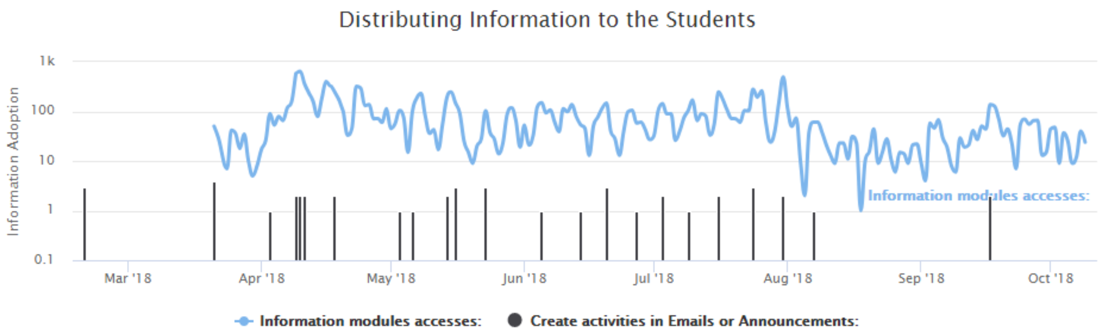

Figure 32. Indicator showing correlations in information distribution activities

The indicators in Figure 33 shows the correlation of student behavior within the learning resources while completing electronic tests as part of the implemented scenarios in their respective courses. These indicators are not available in every course room, but only in the course rooms that have one of the two electronic test systems integrated into the learning platform, eTests by Moodle or Dynexite[5]. These electronic test modules are usually used independently within a course, meaning that if one course uses Dynexite, it does not use the eTests by Moodle, or vice versa. These two systems are an external system, and as such, there is no detailed log data within the collected raw data. Instead, only data that a student accessed the electronic test module is provided. However, this data is enough to show what exactly the students are doing within the learning resources modules, whenever they have a scheduled electronic test. The goal of the indicators is to show whether the students use the provided learning resources, and relative to that when do they complete the electronic tests. Also, the teaching staff can observe within a day when exactly the students are completing the provided electronic tests.

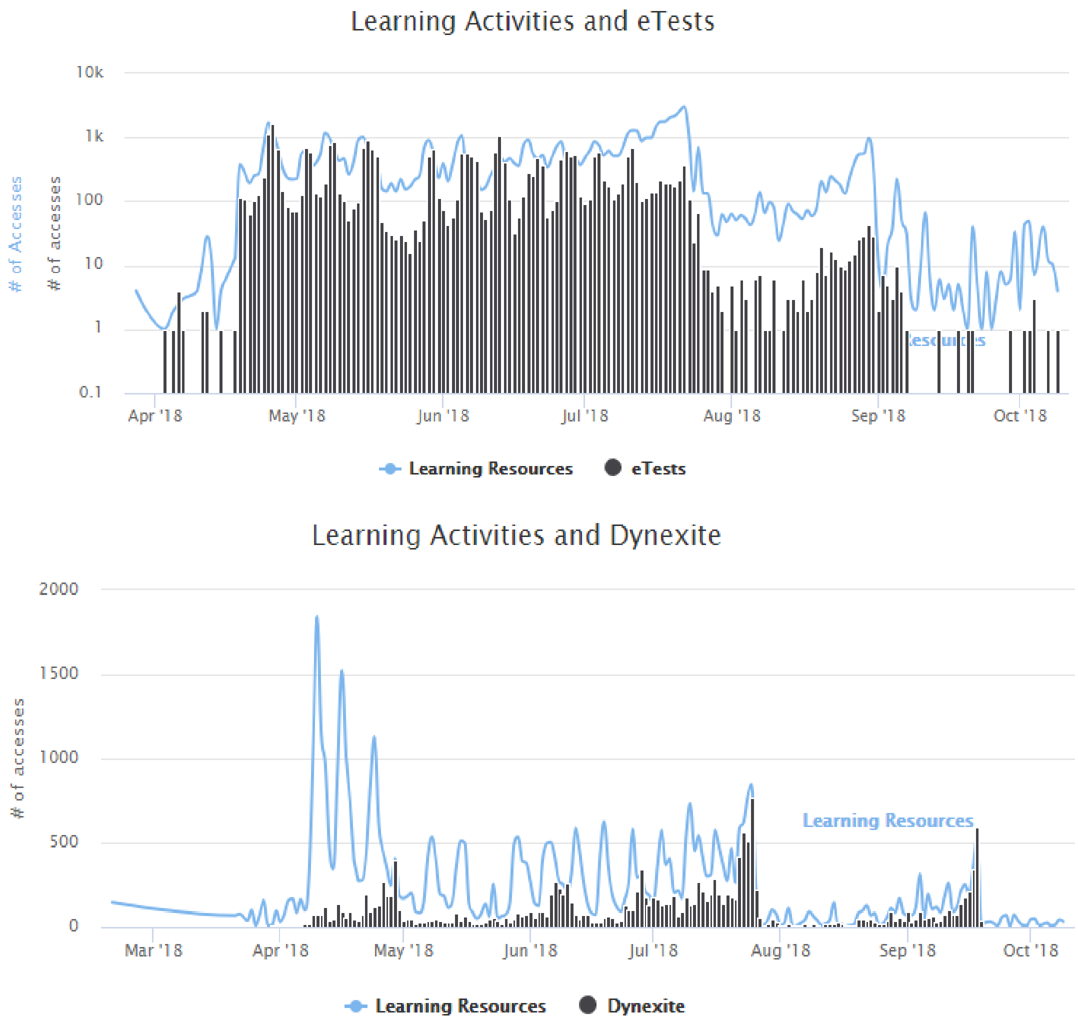

Figure 33. Correlation between e-test and learning resources activities

The indicators in Figure 34 show the correlation of student behavior within the formative assessment modules while completing regular assignments as part of the implemented scenarios in their respective courses. These indicators are not available in every course room, but only in the course rooms that have the Assignments module activated and actively used to manage the exercise submission and correction processes online. The goal of the indicators is to show whether the students use the provided learning resources, and relative to that when they work on their

[5] Proprietary e-test system from the RWTH Business School

submissions for the assignments. Also, the teaching staff can observe within a day when exactly the students are completing the submissions (for example waiting to do it on the last day of the submission, instead of doing it earlier). In the visualization on top, the correlation of creating and grading assignments from the teaching staff is displayed as columns, while the creation of submission activities from the students is displayed as a continuous line. This indicator can also be used as part of the students' indicators because it shows transparency and continuous engagement from the teaching staff to provide immediate feedback back to the students in a timely fashion. The bottom indicator shows that the students engaged in continuous learning before creating submissions for the provided assignments.

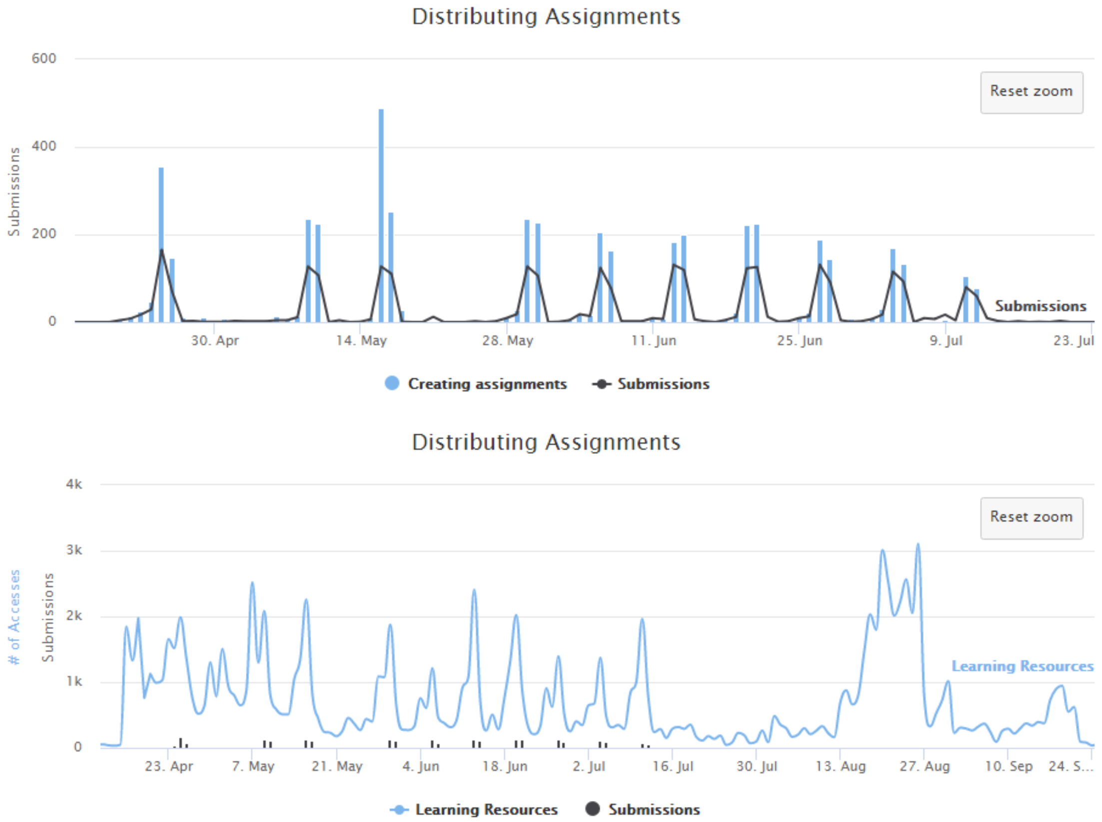

Figure 34. Correlation of assignments, submissions, and learning resources

The last presented indicators Figure 35, and Figure 36 are learning analytics indicators concerned with visualizing the student activities about engagement with the video lectures provided by the teaching staff. The video lecture usage shows the overall usage of all of the video lectures provided in the course room. These indicators are in the course rooms that have the media library module activated and actively used provide videos to the students. Furthermore, there are a lot of courses that do not use the media library for providing videos, but they simply use the learning materials module to upload all video learning resources. This user behavior was considered as well, and the student interactions with these video learning resources were also considered within the creation and data of these indicators. The most watched videos is a derivation of the indicator of most popular (most accessed learning resources) that displays the most watched (accessed) videos in descending order. The main goal of these two indicators is to provide feedback to the teaching staff whether the learning videos are extensively used by the students. This feedback is important because creating high-quality slides which are coupled with a video production creates a tremendous amount of additional work for the teaching staff to provide lecture videos as learning resources.

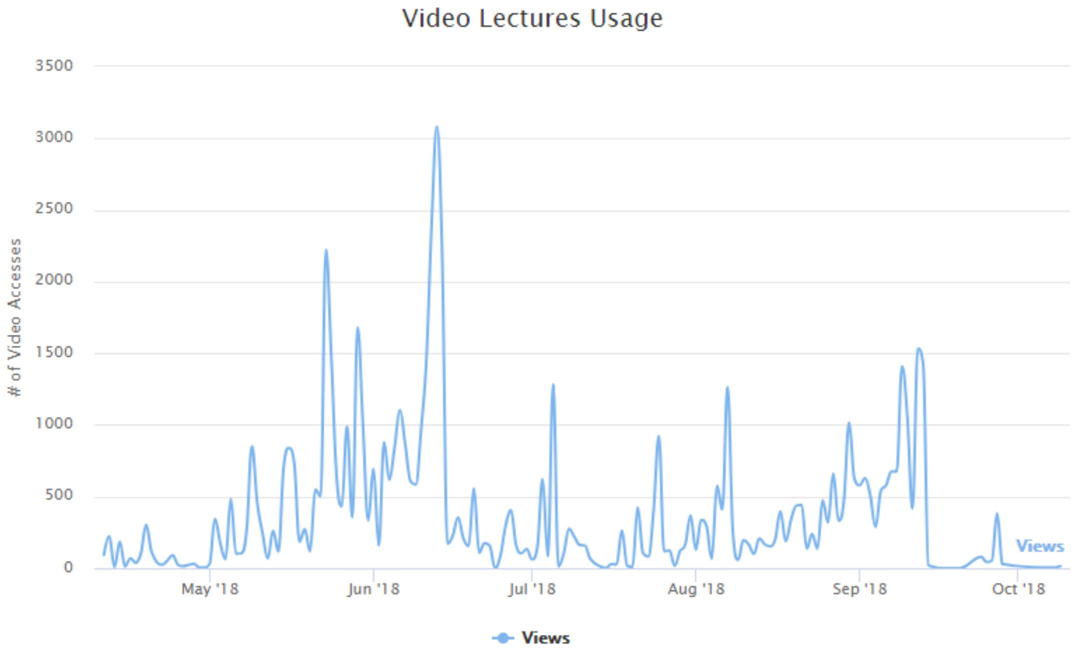

Figure 35. Video lectures use over time

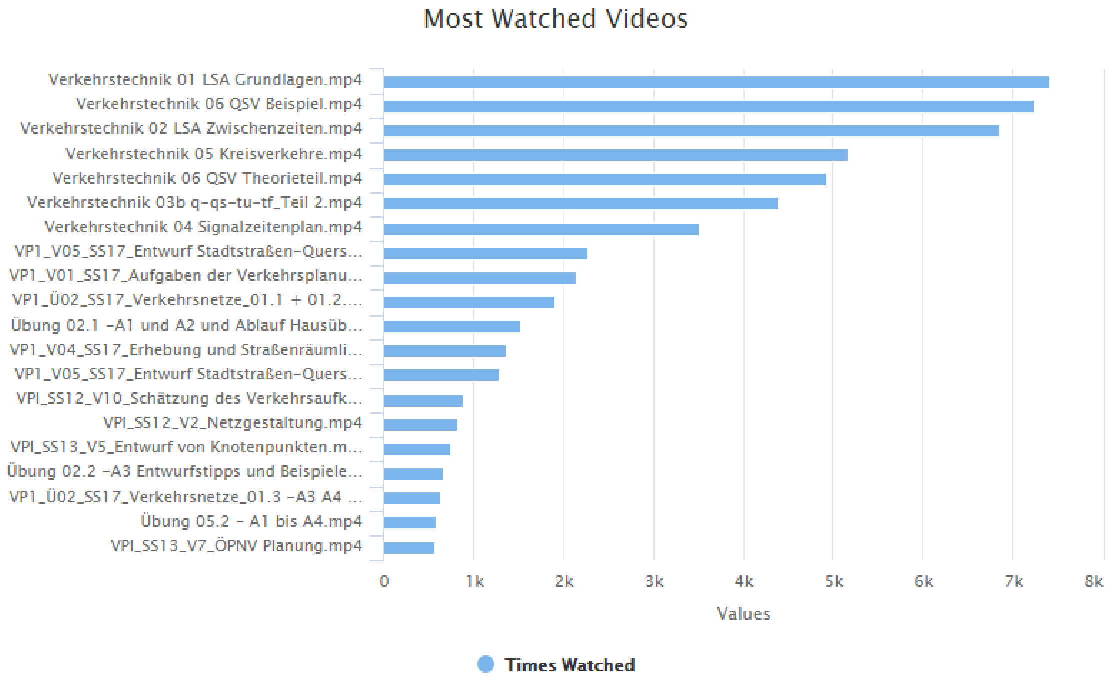

Figure 36. Most watched lecture videos in a course

6.4.5 Aix Analytics learning analytics prototype

The user interface provides a set of indicators which visualize the learning data for the scenarios and goals of the users belonging to the stakeholder groups of administration and the development team (Lukarov & Schroeder, 2017a). The interface was divided into six main categories. The first category of indicators covers usage statistics from the platform itself and the different modules that exist in each course room. The second category covers individual faculty's overall activities, within the different modules in each course room over the different semesters. This category allows for a comparison between courses from different faculties and semesters, or comparisons of different semesters from the same faculty. The third, fourth, fifth, and sixth category take a different perspective and cover the different modules available in each course room on the learning platform. The grouping characteristics within the indicators from these categories are the different faculties at RWTH Aachen University. The visualizations in the prototype were developed to be interactive (with filtering and zoom-in functionality) and enabled the user to focus on specific parts of the visualization. The reason behind this split can be inferred from their goals and perspectives because the two user groups have similar, or overlapping goals, they have different perspectives. The indicators and the visualizations were based on the requirements and guidelines with regards to the questions they answer, the data analytics literacy of the two user groups, and the visual properties of the data with the appropriate visualizations (Lukarov & Schroeder, 2017a).

The visualizations in the prototype were simple and responsive, thus the cognitive effort upon the user was low, and she could concentrate on the visualizations themselves, instead of concentrating on understanding the interface. Furthermore, the visualizations themselves were interactive (filtering and zoom-in functionality) which enabled the user to focus on specific parts of the visualization and drill-down into the data to get a deeper understanding of the data. The interactions with the visualizations were always coupled with informative tooltips (bound to mouse-hover events) to provide feedback about what is displayed at a given time and place. Overall, I developed around 30 different indicators for the six indicator categories (Lukarov & Schroeder, 2017a).

Example indicators

In this part, I provide examples of indicators and their interpretation which have been implemented in the AiX Analytics prototype. The first indicator presents how different faculties use the assessment functionalities on the learning platform. Figure 37 shows the usage over one year of the assessment modules available on the learning platform (Assignments, e-tests, Gradebook, and Exam Results).

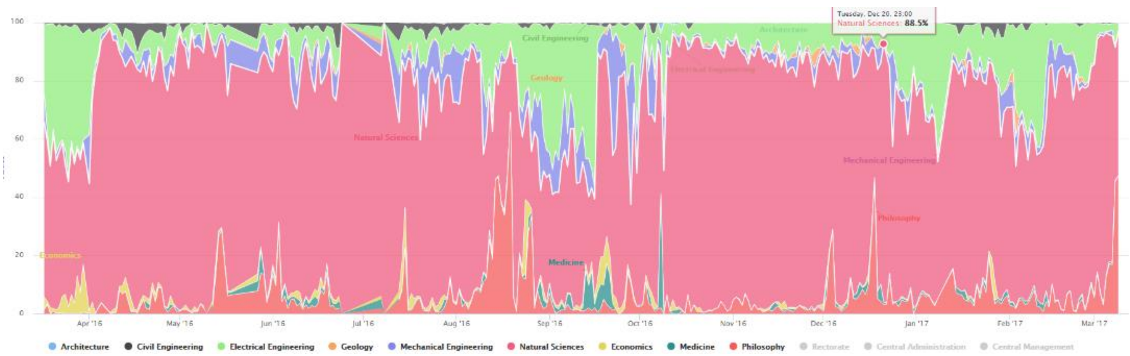

Figure 37. Assessment distribution among the faculties at RWTH Aachen University

As one can see that during the semesters (from April till July 2016, and October 2016 till February 2017) the largest and continuous usage of different assessment modules comes from the Faculty of Mathematics, Informatics, and Natural Sciences (Faculty 1). In the exam phases, the use of assessment modules increases because other faculties publish the results of the exams within the modules. This indicator helps the user to conclude that Faculty 1 uses the assessment modules for formative assessment as an integral part of their blended learning strategies. In comparison, the other faculties (for example Civil Engineering, Geology) are lagging. Moreover, Figure 38 shows an indicator that displays the usage of assignments at the Faculty of Mathematics, Informatics, and Natural Sciences and usage of assignments at the Faculty of Civil Engineering. One can clearly identify the weekly peaks on the left which correspond to weekly assignments, while on the right there are only peaks for two weeks in January.

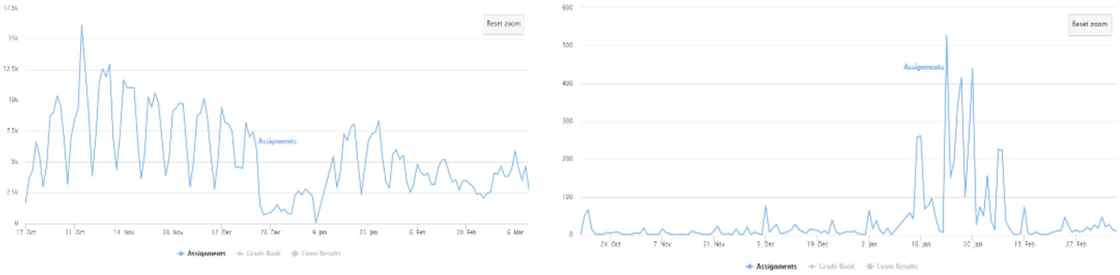

Figure 38. Assignments use - a comparison between two faculties

If one drills down to the most active courses at Faculty 1 by using the indicator that shows courses on the learning platform with the highest number of student activities and events (Figure 39), in the first ten courses, four of them belong to the department of computer science, and all of them use weekly (or bi-weekly) assignments.

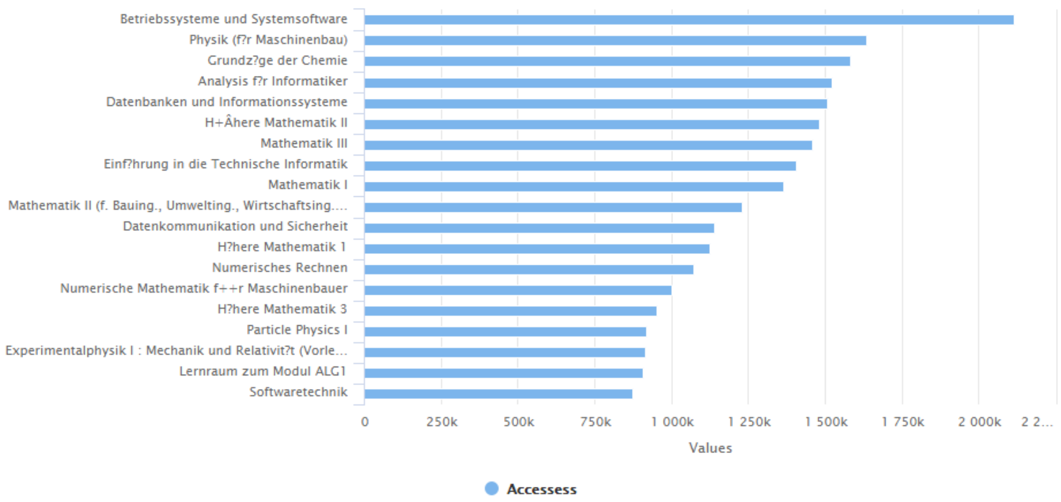

Figure 39. Most active courses from the Faculty of Mathematics, Informatics, and Natural Sciences

The fourth indicator example involves the use of different media in the learning processes and scenarios at the different faculties. If we look at the overall media use, distributed among the faculties, half of the usage comes from the Faculty of Mechanical Engineering (Figure 40). The Faculty of Mechanical Engineering is the biggest faculty at RWTH Aachen University, and these numbers are understandable and expected.

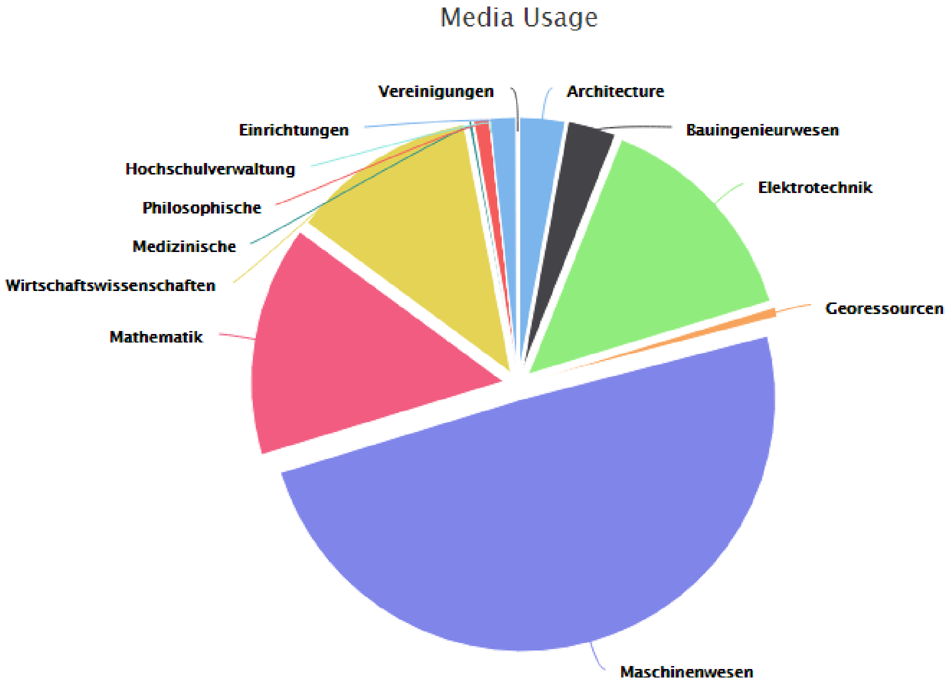

Figure 40. Overall media use among the faculties

The fifth indicator is more general and shows how many users and requests the learning platform has on a daily basis; with what kind of devices the users use the learning platform; and which parts of the learning platform are used over one year. Figure 41[6] represents a logarithmic representation of the daily requests and unique clients that there are on the platform every day. During the week, there are from 16 – 22 000 unique clients (users), and 1.5 – 2.5 million requests, while on weekends the numbers drop to 8-10 000 unique clients, and 0.5 – 1.2 million requests.

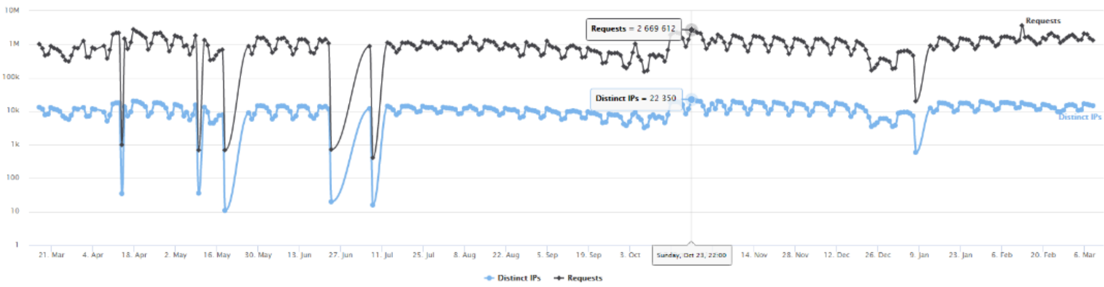

Figure 41. Unique clients and requests on the learning platform over a year

An additional indicator that shows the distribution of desktop and mobile clients on the learning platform. Figure 42 shows that on regular basis around 25 percent of the use comes from mobile devices. At the beginning of every semester, there are peaks (sometimes around 30-40% of the devices are phones or tablets) which can be explained with the fact that the students are looking

[6] The troughs which go to zero in the visualization are points in time when log data was not delivered for analysis.

and registering for courses and lectures and are trying to organize their coursework for the semester.

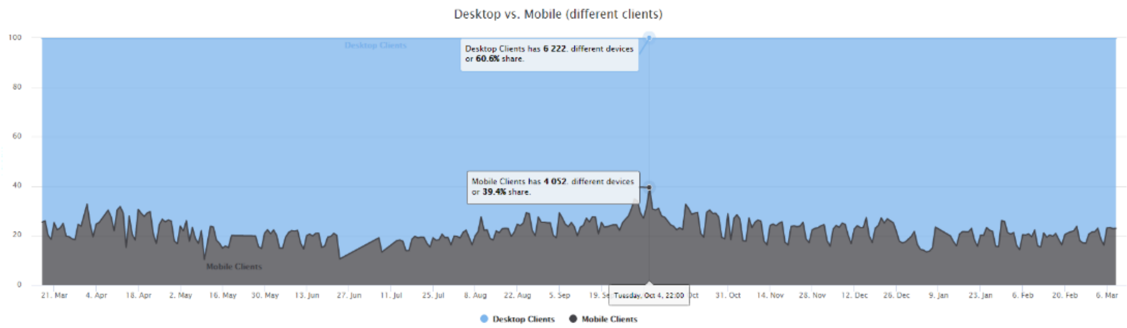

Figure 42. Desktop vs. Mobile devices on the learning platform

As mentioned in chapter 5 section 5.4.1, each course room of the learning platform consists of six domains which contain different modules: Organization, Information, Learning Resources, Assessment, Collaboration, and Management. Figure 43 shows the distribution of the aggregated usage of these six domains. The two module domains with the highest usage are the learning resources and information distribution, amounting to 70-80 percent of the use. They are followed by the collaboration and assessment domains, which each see usage ranging from four to ten percent each.

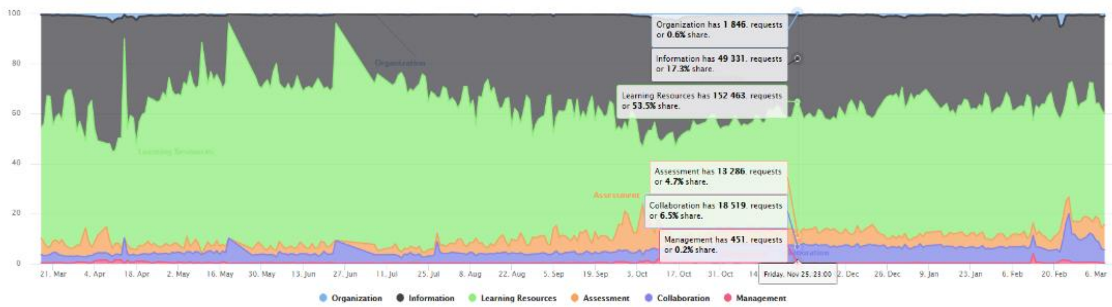

Figure 43. Course room domains use on the learning platform

Formative evaluation

The developed interface for the Aix Analytics learning analytics prototype was evaluated with experts by applying a set of heuristics, and by conducting cognitive walkthrough to pre-defined aspects of the user interface. Overall, four experts participated in the evaluation, one of them was from the learning analytics field, two experts were from the media informatics and computing field, and one of the experts was from the HCI field and Communication Sciences. The cognitive walkthrough was conducted with the persona for the administration group. The description of the persona was provided to the evaluators so that they can familiarize with the user group, the analytics prototype, and the task with a complete list of steps for successfully completing the task. During the walkthrough, the evaluation experts went through the steps for completing the task, while commenting and providing their thoughts for each step within the interface, and their feedback was written down and recorded. The second part of the evaluation consisted of applying a developed set of heuristics by the experts. They received instructions which representative visualizations to open so that all of them reviewed the same visualizations and visualization strategies. Their feedback and comments were also recorded and written down. After the evaluations were concluded, the feedback was analyzed and transformed into concrete steps and improvements which were implemented within the visualizations and the user interface.

The results of the evaluations provided feedback about the visualizations themselves and the applied visualization strategies respective to the different data dimensions. Additionally, the experts mentioned that the number of available indicators or visualizations within one view (indicator group) should not exceed ten visualizations, and the number of different visualizations should range from five to seven visualizations, to keep the cognitive load to the user to a manageable level. The evaluation results also showed that the interaction options should be streamlined, so that the selection of different filters, timespan, or different dates should be finished with the least number of clicks and effort from the user. As an additional point, if there is a table that contains the data visualized within the chart/visualization, this table filled with data should not be displayed by default but be available as an additional option. As a concluding remark, all of the interaction steps, filtering options, and visualization interactions should clearly have a “Back” button that reverses the last step at any given point within the analytics indicator.

6.4.6 Indicator implementation analysis

The rapid application prototyping and implementation led to the design and development of 71 learning analytics indicators within the three prototypes. They were developed and tested in initial and later stages across the different learning analytics prototypes. However, considering their fidelity and how well they fit in with the collected and assigned indicators for each persona in section 5.2.8, the complete set was not used for the several types of the conducted evaluation. Namely, the collected learning analytics indicators from the literature review and document analysis in section 5.2.6, 31 learning analytics indicators were implemented, out of which 20 indicators were used within the subsequent evaluations. Resulting from the cross-comparison with the assigned indicators from the personas and the collected indicators from the literature review and document analysis, 39 new learning analytics indicators were defined and implemented within the learning analytics prototypes, which used analytics results from the learning analytics engine. The complete list of implemented indicators and which ones were used in the different evaluations can be found in the Appendix.

Additionally, as part of the development process, the comprehensive set of collected indicators from the literature review were analyzed through the scope of the available data from the learning platform, the modularity of the implementation, and the data that would be necessary to implement the entire set of 272 learning analytics indicators. The analysis showed that 33 indicators could be implemented with the current dataset and analytics engine (meaning only the Web API that delivers the results has to be adapted). Further 22 indicators can be developed on the existing dataset with updates on the analytics engine and the results warehouse update. If more anonymized learning data is provided, further 45 indicators could be developed within the existing infrastructure with updates within the analytics engine and the results data warehouse. Moreover, a cross reference with the goals, indicators, and use cases from section 5.2.8 with the comprehensive indicator list, showed that most of them could be implemented with anonymous, and pseudonymized data. By further inspection of the indicators and what kind of data they need with relation to the implementation within this dissertation, 167 indicators can be implemented only with pseudonymized or personal data. Out of these 167 indicators, 53 of them can be implemented only with pseudonymized data, for 37 pseudonymized and personal data is necessary, while 72 of them can be only implemented with personal data. The complete of the different indicators and according to their implementation status is available in the Appendix.

6.5 Transferring implementation to other learning platforms

The learning analytics prototypes and infrastructure was built based on available data from the learning platform at our university (Lukarov & Schroeder, 2018). They were built on anonymized collected data from the learning platform following the concept of data minimalism (Bundesdatenschutzgesetz, 1990). Therefore, no personal data was collected that was not necessarily needed to provide analytics as a service, which in Germany can be very challenging due to stringent data privacy laws. The only openly identifiable element of the request was the course in which the activities were made. The collected raw data arrived in the form of seven different parameters (Figure 44) identifying a single HTTP request made to the learning platform.

Log Time	Client IP Address	Client Agent	Processing Time	Operation	URI	Result Code

Figure 44. Raw data structure from the data collectors

The learning platform in use at my higher education institution is a closed-source custom solution created for support and implementation of the different blended learning scenarios. However, many German (and EU) universities do not have the resources nor the expertise for developing a bespoke learning platform for supporting their students and teaching staff in their teaching and learning processes. There are many universities that use an open source learning platform, such as Moodle, or ILIAS. Both learning platforms provide activity logging and learning data collection which can be used as a basis for providing learning analytics services in their respective learning and usage scenarios. However, the data that is collected with their built-in data collection techniques cannot be used as-is, because it is highly personalized. Furthermore, the personal raw data is stored for an indefinite amount of time and can be used to pinpoint individual users and observe their various daily activities within their courses on the platform. These two aspects are not conformant with the current data privacy law rules and regulations.

Time	User full name	Affected user	Event context	Component	Event name	Description	Origin	IP address

Figure 45. Moodle log data activity structure

Figure 45 shows the structure in which the data collection methods are logging the users' activities on the Moodle learning platform. However, Moodle is a modular and open source platform and the data collection mechanisms can be changed and updated to be conformant with the data privacy laws and institutional regulations to provide data which can be used as a basis for providing data privacy conformant learning analytics. The developers and providers of learning analytics services should develop plugins (methods) that pseudonymize or completely anonymize the entries in fields "*User Full Name*", "*Affected User*", and "*IP address*" of the collected log data. Furthermore, they need to develop raw data management strategies that create, clean, and delete the collected personalized logs after the privacy transformation and delete the pseudonymized (anonymized) logs after the conducted analysis on the data. The pseudonymization of these fields will not remove the semantics of the logs, nor reduce their value for providing learning analytics, and the provided analytics will be on the same level as the analytics developed in the works of this dissertation. Additionally, the pre-processed raw data should be analyzed with a different application (and preferably on a different physical layer), so that the user experience and performance is unaffected by the computational-heavy data analysis. The learning analytics infrastructure and implementation can be a standalone web application which can be embedded in different courses on the developed learning platform, or in Moodle, the analytics results would be delivered via a Moodle plugin. As the last step towards providing analytics as a service, an automation process should be developed that automatically triggers every step from the process: data collection and pre-processing; the data analysis and saving the results; providing them in the appropriate courses on the learning platform; and removing them completely from the system with accordance to the pre-defined course lifecycle. By implementing this framework, fellow researchers can develop learning analytics components, experiment with them in different blended learning scenarios like the case studies in this research work, and potentially scale them up and provide them as a service in their institutions.

6.6 Conclusion

This chapter presented the design and implementation of sustainable learning analytics infrastructure, and two learning analytics prototypes for the different stakeholder groups. The "Insights" learning analytics prototype enables the teachers to gain a deeper understanding of how their teaching activities relate to the student behavior in the virtual course rooms, and whether they engage in continuous learning. With the same analytics prototype, the students can observe their behavior and their effort within the course, correlated with the success scenario requirements and with their peers. The "Aix Analytics" prototype provided indicators for the administration and the IT Staff to observe the learning platform usage patterns, the different implemented learning strategies and scenarios, and observe the distribution and use of learning resources and e-learning infrastructure among the different departments and faculties. Figure 46 provides the implemented learning analytics infrastructure and the underlying technologies used for server-side and client-side implementation.

The provision of the two learning analytics prototypes encompassed the development of two data warehouses, one for the raw interaction data, and the other for storing the results from the analytics engine. The analytics engine was optimized by using optimized data loading techniques and performant data structures which reduced the computational time for analyzing the raw data. The communication between the two data warehouses, the analytics engine, and the user interfaces were implemented in a standardized way via extensible RESTful APIs to ensure the logical and modular independence of the different analytics infrastructure components, thus making it easier to deploy and maintain the entire infrastructure.

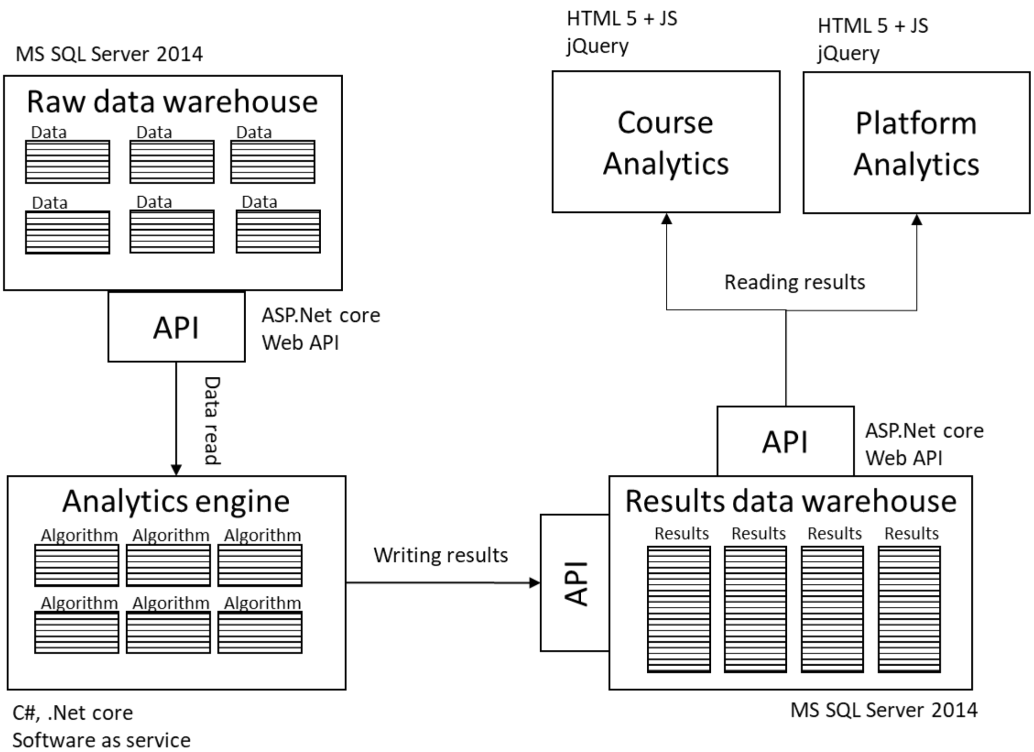

Figure 46. The implemented LA infrastructure with server-side and client-side technologies

Each of the analytics prototypes followed the learning dashboard design philosophy and provided a selected set of indicators and visualization based on the identified needs and requirements of each stakeholder group. These analytics prototypes were iteratively developed by applying rapid application prototyping and development, so that early iterations of the indicators could be evaluated and improved. The most completed prototype which went through many smaller iterations, and two major ones, was the “Insights” learning analytics prototype for the teaching staff. The entire infrastructure with this prototype would be tested and evaluated in a real deployment scenario, to assess the effect of the real-world application of learning analytics tool in blended learning scenarios. The method and the results are presented in chapter 7.

7 EVALUATION

Learning analytics practitioners need easy to use analytics tools which provide pleasant user experience (Anna Lea Dyckhoff, 2014). Moreover, these analytics tools should also have good utility and usefulness which ensure continuous usage and incorporation in their online activities. The provision and access of such tools should be seamlessly integrated within the users' online working environment and available at-hand without extra effort. These analytics tools should provide added value to learners, educators, and the administration (Scheffel et al., 2014). Learning analytics can help learners to reflect and plan their activities by becoming aware of their situation and actions within the learning processes, comprehend it and then project future activities which can result in continuous learning and achieving better results and change in their learning behavior. Learning analytics offers ways for learners to improve while taking courses with the help of learning data visualizations and analytics which inspire awareness, reflection, and action (Ferguson, 2012a; Scheffel et al., 2014). The added value for the teaching staff lies in the disclosure of hidden processes, the disclosure of the current situation within a course in comparison with the assumptions of the teaching staff, and the students' learning behavior and activities within the virtual learning environments. Learning analytics can provide invaluable assistance by providing information about the students' activities and interactions with the provided learning resources and course material, and whether the students are showing intrinsic motivation by engaging in continuous learning. Moreover, learning analytics can provide knowledge and support when the teaching staff is evaluating a course in terms of the applied didactic concepts, the chosen learning resources and resources delivery styles, their content, the formative assessment activities and the results of the information distribution, teaching and intervention activities during the course duration (Anna Lea Dyckhoff, 2014; Scheffel et al., 2014). The administration can receive substantial support from learning analytics within their decision-making processes for resource planning and have evidence-based overview of the different implemented blended learning scenarios at their institution, and in turn adjust their support for the development and improvement of the existing learning institutional infrastructure (of if necessary) provide support and funding for extending and building new services. Their data-driven decision making and investing effort and support where it is most needed and has the highest development potential can improve their accountability and the image of the higher education institution (Lukarov & Schroeder, 2017b). Lastly, the IT Staff and the development team can benefit from the knowledge about how different faculties use the various functionalities in their course rooms on the learning platform. They will know how exactly (when, how much, which devices) the users interact with the different modules in these courses. This way the development team can evaluate the modules on the learning platform, identify patterns of usage, potential problems, system load, and other parameters that could help improve the platform from a developer's/technical point of view (Lukarov & Schroeder, 2017b).

The research questions and the goals of this dissertation afforded the development of a sustainable learning analytics infrastructure which includes learning analytics prototypes for the different stakeholder groups. The infrastructure was built within the bounds of the theories presented in chapter 4; the analytics infrastructure and the resulting prototypes included the requirements and the institutional framework and setting presented in chapter 5. The iterative development of the analytics infrastructure, the development, and evaluation of the infrastructure and the prototypes are presented in chapter 6. Nevertheless, to achieve the goals of this dissertation and answer the research questions, the entire learning analytics framework with the resulting learning analytics prototypes has to be evaluated within real-world scenarios which include many participants within many courses. The evaluation methods should measure the utility, sustainability, and to which extent the needs, the goals, and the expectations of the different stakeholder groups. These evaluation methods have to continuously observe the usage of the learning analytics in real learning and teaching scenarios and should collect feedback about the integration and correlation of analytics usage and other teaching and learning activities. On the other hand, with the administration and the IT Staff, the evaluation methods must collect feedback about the integration and correlation of analytics usage and their other day-to-day activities. Conducting all these longitudinal studies with all of the stakeholders is a very complex and challenging undertaking that can take years to conduct, complete, analyze, and are well beyond the evaluation scope of a single dissertation. Therefore, for the longitudinal evaluation of the architecture and learning analytics prototypes, I focused only on one stakeholder group. The longitudinal studies for the other stakeholder groups can be organized and conducted with the same approach and method.

The evaluation methods should work within the context and collect data and feedback from multiple sources which focus strongly on qualitative data on

- the needs, goals, and expectations of the teaching staff whether they were properly and comprehensively collected;
- the place of delivery and provision of analytics tools, so that the teaching staff can use the analytics tools on a regular basis;
- incorporation of analytics in day-to-day teaching activities;
- teaching staff awareness of students' behavior and trends in online activities;
- reflections and correlations of their teaching activities;
- analytics competencies, analytics delivery, and acceptance strategies and teaching behavior change after the provision of learning analytics;
- a pleasant user experience;

Collecting feedback about all these aspects and determining the impact of learning analytics required extensive knowledge about designing the raw data collection mechanisms, and previous practical experience in development of educational software so that all the possible influencing factors on the users and the contexts are considered. Moreover, each course is a small learning eco-system and its context and applied learning scenarios must be considered when collecting feedback data. The evaluation type was matched with the main goal of this thesis and as such the evaluation was conducted in longitudinal studies with many courses with implemented automatic data collection from several data sources and the use of pre-defined qualitative survey methodology. This quasi-experimental evaluation setup enabled automated quantifiable data collection which could be correlated and corroborated with the collected qualitative feedback from the participants in the evaluation. The qualitative and quantitative data from the evaluation setup could provide answers to questions relating to the topics above. For example:

- *How the users are interacting with the learning analytics tool?* The assumption is that since the learning analytics tool is provided in their course rooms, the teachers would open it and try to use it to monitor the students' behavior. I also hypothesize that users with background previous knowledge would appreciate the visualizations and would expect functionality that will allow them to export the data and do additional analysis.

- *Are there any features that are meaningful to some users?* The analytics indicators were based on extensive results from requirements engineering and elicitation, and the assumption is that the design of each course would influence the users' interests in some of the indicators, while other indicators would be deemed as unnecessary, or not useful for them.
- *Are there users who use the tools while doing other teaching activities?* The provision of learning analytics tools as part of the e-learning services is not enough for learning analytics. It is crucial that the use and application of learning analytics tool do not differ from any other teaching activities and be part of them. Therefore, I assume that it is very important to provide the learning analytics tool within the course room on the learning platform so that the teaching staff would use it while conducting other teaching activities within the course room.
- *Do the users plan to change something in their teaching activities or courses after being presented with the tool?* I expect that the teaching staff would appreciate the immediate feedback and conveyed information about the students' behavior and activities within the course room and they would use this feedback to re-evaluate their teaching and reflect upon what could be done better in the learning scenarios and the learning processes based on the data provided by the indicators.
- *What were the reasons for not using the learning analytics tool?* My assumption of this claim is that there will be users who would not appreciate any new technology nor e-learning solutions provided in their course rooms. The answer to this question can provide information on what could be done better when provisioning learning analytics on a scale.

7.1 Evaluation methods and techniques

I chose to use case studies for evaluating the learning analytics infrastructure and the prototypes. A case study is a detailed examination of one or more specific situations within a specific real-life context. The case study is described by the following four key aspects: (1) In-depth investigation of a small number of cases, (2) Examination in context, (3) Multiple data sources, and (4) Emphasis on qualitative data and analysis. The evaluation consisted of two longitudinal case studies which ran over two following summer semesters at RWTH Aachen University. The first case study was on a smaller scale involving a small number of courses, and the second case study was on a larger scale involving hundreds of courses. The main steps of the evaluations were: selecting the courses, activating the "Insights" analytics prototype in them, conduct the case study's two-part survey, and analysis of the results and conclusions. The advantage of the case studies was in the fact that it was conducted in the field which afforded a great deal of freedom of the participants which meant that the results were closer to reality (in comparison with lab studies). However, the downside of this study setup was that there were a lot of external factors that could not be controlled, thus influencing the users and the results. The raw feedback data from the multiple sources was carefully documented and the final analysis and the coding of this data were conducted by using a grounded theory approach from Corbin and Straus (Corbin & Strauss, 1990). The coding and analysis approach had three stages:

1. *Open Coding*: First, the collected data from the multiple sources was closely examined and interpreted. For this reason, different data analysis methods were developed and applied to analyze it, so that different phenomena could be identified and labeled.
2. *Axial Coding*: In this stage, the relationships between the identified events and phenomena were discovered and correlated by identifying the connections and relationships between them. Each event and phenomenon were examined, and the process model of these relations was also created.
3. *Selective Coding*: In the last stage all of the identified categories and events which were central for answering the research questions, were summarized and converged within the central phenomenon of the study (the provision of learning analytics tools).

The following sections provide comprehensive information and details on the two case studies, the implemented data collection mechanisms, present the findings, and draw conclusions based on the results and the findings.

7.2 Pilot Case Study Summer Term 2017

This case study consisted of deploying an analytics prototype to a small number of courses (53 courses) on the learning platform. It fits well in a case study scenario because on the learning platform there are around 3000 courses per semester, and these fifty courses are ~1,5% of the number of running courses per semester, which means the sample size is small. The context of the study was the technology aspect of the implemented blended learning scenarios in real courses on the learning platform to grasp a more realistic understanding of how analytics would be used within the learning platform in blended learning scenarios. I built three types of data collection mechanisms to collect corroborating evidence and with the help of data triangulation clarify and support the observations and results of the case study. For this purpose, I collected anonymous log data on the usage of the analytics prototype, collected log data on the users' activities within these courses on the learning platform, and conducted a two-part survey. The survey consisted of questions that collected qualitative feedback about the analytics prototype and a seven-point Likert scale usability questionnaire based on the ISO 9241/10 international standard.

The case study goal was to develop an understanding of teachers' use and incorporation of analytics and statistics tools/interfaces in their day to day teaching activities within the learning platform. The idea was that the teaching staff could freely explore the analytics prototype while conducting their online teaching activities in blended learning scenarios. This information is important to explore and understand how teachers carry out teaching tasks with the learning platform and explain how they incorporated analytics within their day-to-day activities. Furthermore, the case study could help in proving that the teaching staff really used the prototype and describe the tools' usage patterns and activities within the course. The hypothesis of the case study was that teachers would use the analytics module on regular basis, while doing other teaching activities within the course, within one session on the learning platform. This assumption is derived from the fact that the analytics prototype would be seamlessly integrated within the course on the learning platform, and teachers would be compelled to use it to get a glimpse of what is going on in their course.

7.2.1 Case study setting and design

I randomly selected and contacted a wide audience of professors and teaching assistants of different faculties at RWTH Aachen University via email to offer them to participate in the case study. Professors and teaching assistants of 53 courses agreed to take part. 33 courses were a lecture connected with an exercise, 15 courses were practice-oriented laboratory courses, and five courses were seminar courses. Regarding the course distribution among different faculties, 17 courses were from the Faculty of Mathematics, Computer Science and Natural Sciences; Six courses were from the Faculty of Mechanical Engineering; Three courses were from the Faculty of Electrical Engineering and Information Technology; 11 courses were from the Faculty of Arts and Humanities; 15 courses were from the School of Business and Economics; and one course from the Faculty of Medicine. The number of students participating in each course varied from 20 students to 2200 students. In retrospection, one can conclude that although the number of courses is small in regard to the total amount of courses per semester at RWTH Aachen University, the courses were distributed among six faculties (out of nine faculties in total), the course types are the three most common course types at RWTH Aachen University, and according to the number of students per course, the sample size contains courses with a small number of students and very big courses with more than 2000 participants.

The analytics prototype “Insights” presented in chapter 6, section 6.4, was available for the study participants by the end of April 2017 as an integral module inside their respective courses on the learning platform. After activating the “Insights” module in their courses, all participants received via email instructions and explanations about the module, descriptions about the visualizations, what kind of data is visualized, and guidelines about possible (valid) interpretations of the represented data within the visualizations. I did not provide special instructions about when and how they are supposed to use the module, but rather try to incorporate in their day-to-day activities. They also received information that the module activities would be observed by automatic logging tools, and towards the end of the pilot phase, they would be given an online survey about their experiences with the “Insights” module. The survey itself was non-binding, meaning that the participants were not obliged to fill it in. The “Insights” module was never deactivated from the courses, so the teaching staff could still use it in their (now old) courses. However, only the period between the availability of the “Insights” module and the end of the semester is considered for the analysis of this case study.

7.2.2 Results and Discussions

The case study results provided sufficient amount of raw-data and feedback indications to conclude that the teaching staff used the “Insights” module. The results would be based on the collected data from the three identified data sources and selected for the case study in section 7.2.

Overall, during the time of the case study, 40 courses from the 53 that had agreed to participate in the case study have used the “Insights” module at least at one occasion. However, this means that 13 participants had not used the “Insights” module at all. The usage[7] frequency of the “Insights” module during the study was not evenly distributed. There were five courses where the teaching staff used the “Insights” module only once; in nine courses the teaching staff used the “Insights” module less than five times; in 10 courses the teaching staff had used the “Insights” module between six and ten times; and in 15 courses the teaching staff had used the “Insights” module over 10 times during the duration of the study. In the five courses with most usage frequency, the usage numbers ranged from 43 to 103 times (occasions) on which the “Insights” module had been used. The usage frequency of the “Insights” module during the study is summarized and presented in Table 13.

Table 13. Usage frequency of the "Insights" module in SS 2017

Number of Courses	Number of uses
15 courses	More than ten times
10 courses	Between six and ten times
9 courses	Less than five times
13 courses	No usage detected

I also conducted descriptive statistics of the mean usage frequency with standard deviation and calculated the median and the mode. The average usage frequency is 15.5 times per course with a standard deviation of 22, the median is seven, while the mode is one. In this case, the average usage frequency and the standard deviation do not depict the real outcomes, because the standard deviation is larger than the mean and the coefficient of variation of 141, show that the usage frequency data is spread across widely around the mean. In this case study, the median and the mode are more descriptive and suitable for the analysis, because the median separates the higher half from the lower half of the usage frequency data, while the mode shows the value that appears

[7] Usage in the context of this study means that a teacher has opened the “Insights” module in her course on the learning platform, and afterwards performed another action or activity within the course on the learning platform. This interpretation helps in excluding noise within the raw data from accidental clicks and usage.

the most within the usage frequency data. In other words, the mode appears within the lower half of the usage frequency data, and at least in half of the courses, the teaching staff has used the “Insights” module on multiple occasions.

In Figure 47 can be observed how the usage frequency of the “Insights” module developed over the course of the case study. On the *x-axis* is the timespan of the study, while on the *y-axis* is the number of different courses from which the “Insights” module was accessed over the given time-period. After the initial peak of usage when the “Insights” module was available to the teaching staff, regular weekly peaks of the module’s usage can be identified, troughs on the weekends, and the activities in the “Insights” module decrease towards the end of the semester.

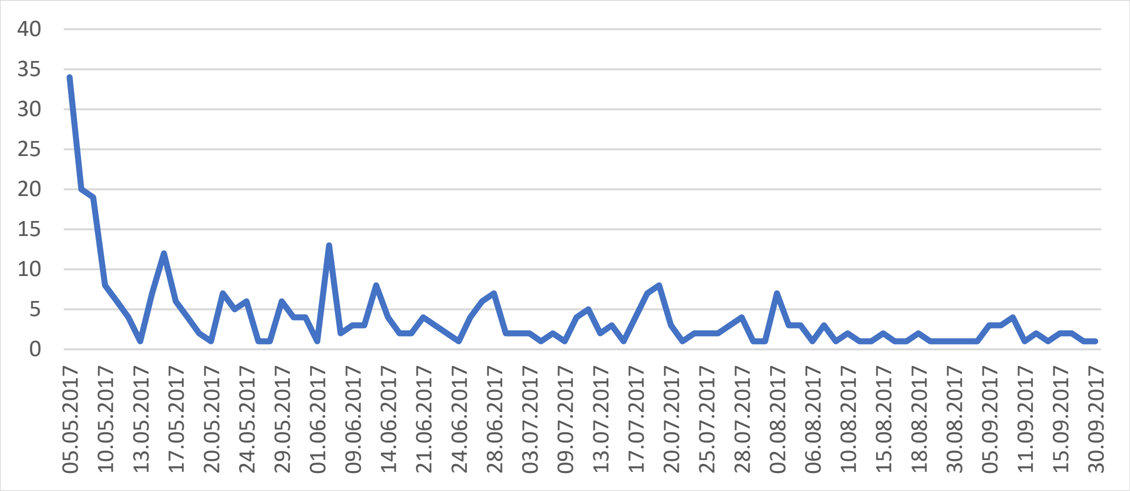

Figure 47. Number of courses that used the "Insights" module in SS 2017

This can be attributed to the fact that the semester in German universities is divided into “Semesterwochenstunde” or SWS, to indicate the time students need to spend on a course per week and to measure the teaching load of the teaching staff. For this reason, I analyzed the distribution of the tool’s usage within the course over the semester divided into weeks to understand the usage distributed over different courses spread over the weeks of the semester. Figure 48 shows the “Insights” usage per course distributed over the weeks of the semester.

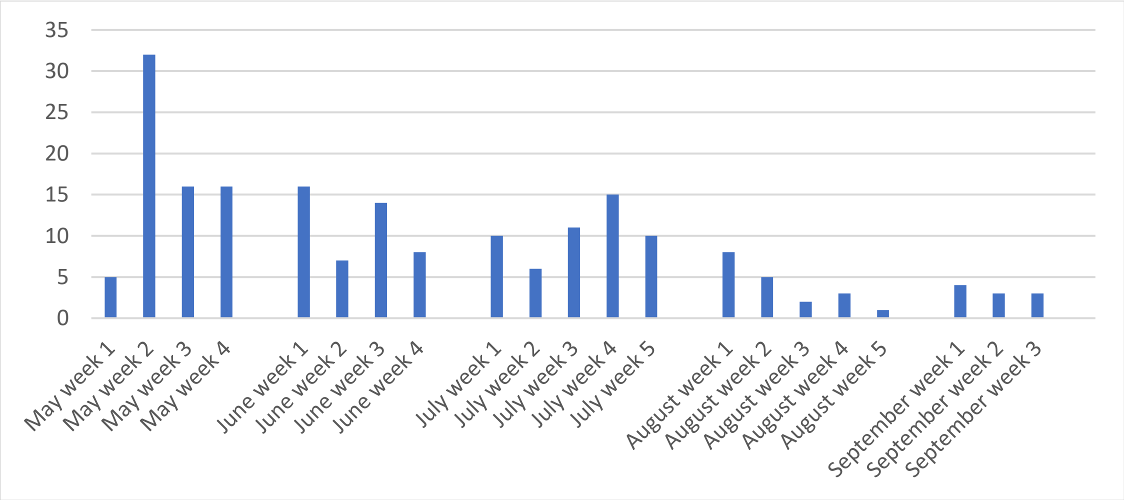

Figure 48. Number of courses per weekly summaries SS2017

In the first week of May is the largest number of courses in which the module was used. This is because this was a new feature in the course announced to them via email, and most probably they opened it to take a look at the new module in their course rooms. The next three weeks showed a consistent number of courses (15-20) in which the "Insights" module had been used. In the second week of June was the excursion week (a holiday week in which there are no lectures), and the trough is evident because afterward the number of courses increases above ten courses. The course lectures end by the second week of July and the exam phase starts immediately after the end of the lectures and the exam phase lasts in some cases until the end of September (which is also the end of the semester). What I find interesting in the weekly distribution of the "Insight" module usage is, although the lectures had ended, the number of courses in which the "Insights" module had increased in the last three weeks of July. In August and September, the number of courses in which the "Insights" module was used, steadily decreased.

The next analysis that I conducted was to observe whether there was a similarity between the usages of the "Insights" module among different courses. The reason behind this observation was to investigate whether in the peaks there were many single courses with incidental usage (only once), or there were many courses with intentional usage of the "Insights" module. For this purpose, I collected the usage of the "Insights" module for every individual course and plotted the usage data for each course in a line chart, Figure 49, 50, 51, 52 represent a combined line chart for the usage of the "Insights" module over time for each individual course per month. I divided this representation per month, to make it easier to read and understand the visualizations, and to correlate it directly to the weekly distributed analysis, presented in Figure 47. On the *x-axis* are presented the days of the month, while on the y-axis are the number of times the "Insights" module has been used on each day within a course.

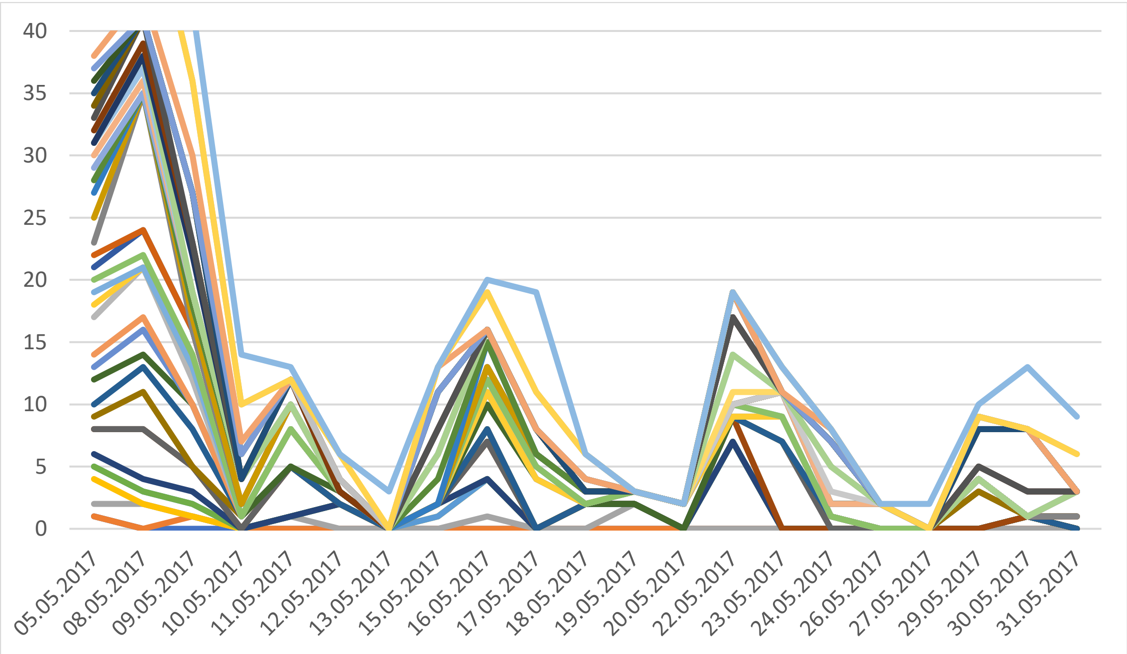

Figure 49. Use of "Insights" per individual course in May

As one can see in Figure 49, the peaks in May together with the number of courses correspond with the number of courses presented in Figure 48, and the number of accesses of the "Insights" module in most cases is bigger than one. A similar observation could be made about the usage of the "Insights" module in June, with the notable peak at the beginning of June. As mentioned before, the second week of June was the excursion week and there were no lectures. The peak usage was right before the start of the student holidays, followed by a trough, and then again peak towards the end of June, as presented in Figure 50.

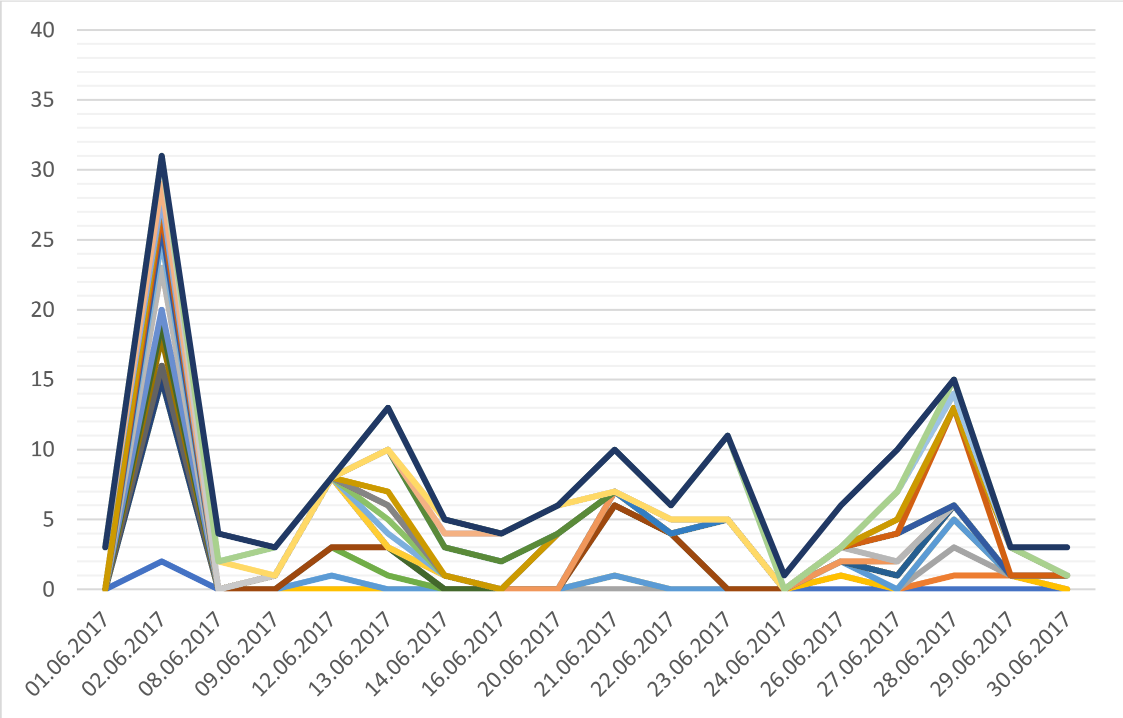

Figure 50. Use of "Insights" per individual course in June

Towards the end of the semester, the usage of the insights module per course had a decreasing trend, as the recurrence of the "Insights" usage fell in terms of usage per single course, and the number of courses. Figure 51 represents the situation in July which corresponds with a smaller number of courses at the beginning in which the module had been used, with an increase towards the second half of the month. As mentioned before, in the second week of July the lectures had ended, and the exam phase had started.

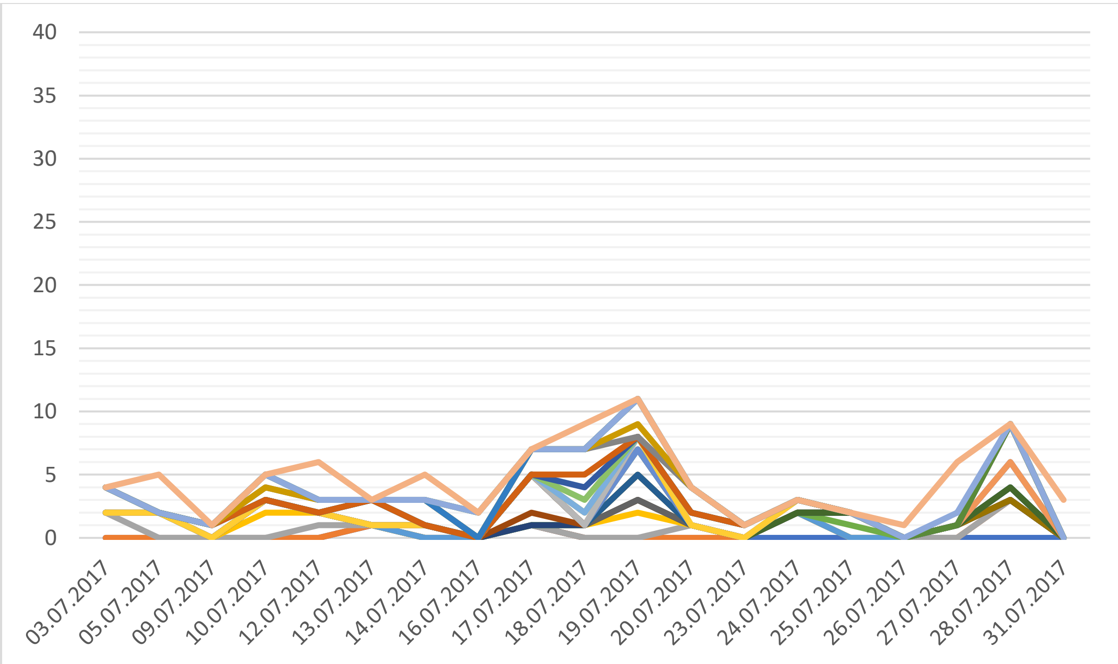

Figure 51. Use of "Insights" per individual course in July

Lastly, Figure 52 represents the usage of the "Insights" module per course in August and September. There was an increasing trend in the usage of the module from the third week of July, which peaked at the beginning of August. This is the time-frame where most of the exams were placed, and afterward, there is a distinct decrease in the overall usage of the "Insights" module across all courses, the exception being a couple of courses that have extensively used the module on several occasions.

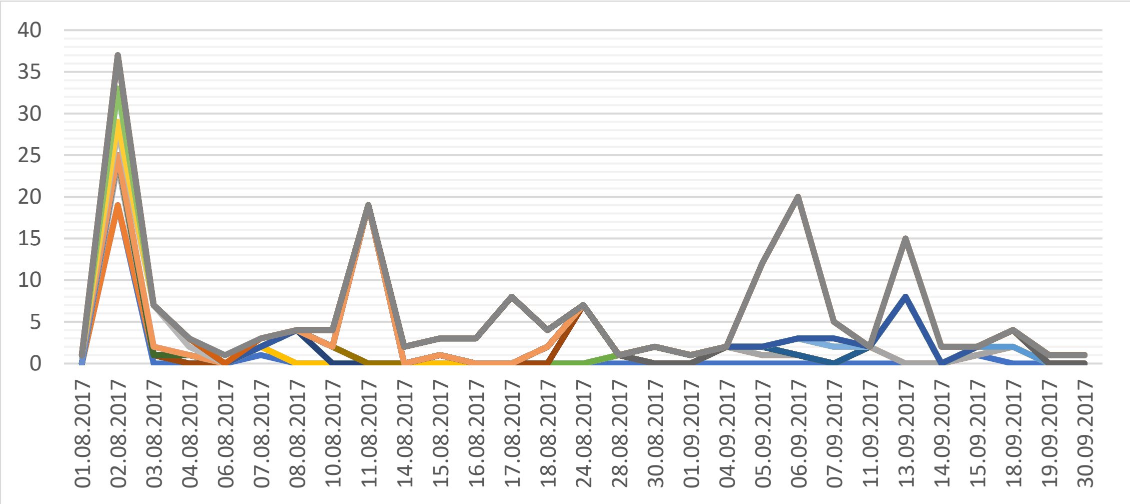

Figure 52. Use of "Insights" per individual course in August and September

The last analysis of the collected usage data of the "Insights" module was conducted to identify whether the teaching staff logged in on the learning platform to use the "Insights" module explicitly, or they used it as part of their day-to-day teaching activities. For this purpose, the collected log data from the learning platform used for providing analytics and insights about the students' activities within the course was re-purposed and re-analyzed to identify and aggregate the teaching activities within the course rooms on the platform. The main idea behind this analysis was to discover whether, within the same session[8] in the course on the learning platform, the teaching staff has performed teaching activities while they have used the "Insights" module. The teaching activities covered with the analysis were chosen based on their relevance and influence on the learning processes within the course. Hence, bureaucratic, and technical activities within the course room that have no influence over the students and their learning, were disregarded and were not analyzed. The teaching activities that were taken into consideration for the analysis were divided into four major groups: *information distribution activities*, *course organization activities*, *distribution of learning resources*, and *formative assessment activities*.

Information distribution activities cover activities within the Announcements and Emails modules within the course on the learning platform. The main purpose of these modules is to enable the teaching staff to contact their students to distribute important course information, send out reminders about course events, resources, or different activities (including formative or summative assessment). The teaching staff can distribute this information either by posting an online announcement or send out an email to the students. The correlation of the log-data analysis showed that in 17 courses when the teaching staff used the "Insights" module within the same session, they have posted an announcement, and 17 courses have used the "Insights" module within the same session when they had sent an email. Combining the two course lists to remove

[8] The term "session" is used in the context of an interaction session, when the user has logged in onto the system via a web browser and has performed different activities and interactions within the system.

redundancy, overall the teaching staff of 26 courses had distributed various course information with the modules for information distribution, within the same session when they had used the "Insights" module.

The course organization activities cover activities within the Calendar and Survey modules within the course on the learning platform. The main purpose of the Calendar module is to organize and manage the course calendar by allowing the teaching staff to import course event dates from the campus management system and creating custom course events. The main purpose of the survey module is to enable the teaching staff to conduct in-course surveys with the students. The correlation of the log-data analysis showed that in two courses when the teaching staff used the "Insights" module have also created or edited course events in the calendar and that in one course the staff has created/edited a survey in the same session when using the "Insights" module.

The learning resources distribution modules cover activities that distribute learning materials, learning media and lecture videos, online resources as hyperlinks, and scientific literature to the students. The main purpose of these modules is to support the teaching staff in implementing their teaching scenarios so that the students have access to a versatile set of learning resources. The correlation of the log-data analysis showed that in 22 courses when the teaching staff had provided or uploaded learning materials has also used the "Insights" module. In four courses they have uploaded or embedded lecture videos, and in three they have provided online resources as hyperlinks. Combining all course lists, overall the teaching staff of 24 courses had provided learning resources within the same session when they had used the "Insights" module.

The formative assessment activities cover modules that provide formative assessment features within the course room. This enables the teaching staff to publish assignments within the course, then the students can submit solutions, and afterward, the teaching staff can correct the assignments and publish the scores within the Assignments module on the learning platform. The correlation of the log-data analysis showed that in six courses when the teaching staff provided or edited an assignment has also used the "Insights" module. In three courses, the teaching staff used the "Insights" module when correcting student submissions. Combining the two course lists, resulted in a total of seven courses where the teaching staff had performed activities within the formative assessment modules within the same session when they had used the "Insights" module. To conclude, all the courses in which the teaching staff has performed a teaching activity whilst using the "Insights" module was also compiled and the result is 36 courses.

Survey Results

The two-part anonymous survey provided qualitative feedback about the "Insights" module, and a usability questionnaire based on the ISO 9241/10 standard (Figl, 2009; ISO, 2010) was performed to assess the usability of the prototype. In total, eight participants have filled in the survey. The first part of the survey collected feedback about what were the positive aspects of the "Insights" module; what were the negative aspects or experiences with the "Insights" module; which features, and visualizations of the "Insights" module were useful the most to the participants; and what would the participants wish to see in the "Insights" module to better fulfill their needs and expectations. According to the answers, the possibility to have an overview about which learning materials are mostly used over time within the course room; the possibility to see whether the students have used the provided media and the lecture recordings provided by the teacher; and the possibility to observe how the students' behavior developed over time in the course room are among the positive aspects provided by the "Insights" module. On one occasion, the teacher could infer with certainty when and how the students worked on the assignments and their submissions. The negative experiences with the tool were mainly concerned with the data representation and visualization. The answers included statements about glitches in the zooming functionality and unfitting representations of the data on the charts' axes; and the lack of help and description of the visualizations. One participant also had problems with accessing the tool because the module was available only within the RWTH Aachen network. An interesting claim as a negative experience was that the participant's fears were confirmed that the students always studied and looked at learning resources just before the exam. According to the answers, the highlighted features of the "Insights" module were the ones that showed analytics and information about activities within the learning resources modules. The participants could see which the most popular learning materials and resources within the course room were. One participant mentioned, that his expectations about the students' behavior were confirmed and that he can use the tool to adapt his learning offerings and teaching behavior. As possible improvements, the participants in the survey suggested a provision of help and guidelines about how to interpret the visualizations; smoother and clearer visualizations with better zooming functionality. There was also requested the possibility to be able to combine and export the visualized data for offline analysis and usage, and to provide the tool available outside RWTH Aachen network.

The goal of the second part of the survey was to collect feedback and information about the usability of the "Insights" module. Usability in an interface ensures that the interactive system is easy to learn, effective to use, and enjoyable from the user's perspective. Usability involves optimizations of the interactions and the look-and-feel of the interface to enable the users to carry out their activities. Usability can be broken down in these goals: effective to use (*Effectiveness*), efficient to use (*Efficiency*), have good utility (*Utility*), easy to learn (*Learnability*), easy to remember how to use (*Memorability*). In the survey there were a set of questions that covered each of these goals, to collect qualitative feedback from the participants. The questions themselves were created based on to ISO 9241/10 standard. The seven-point Likert scale ranged from "Strongly Disagree" to "Strongly Agree" (1-7) and was used as a ranked order across all 16 questions to receive more consistent results for the different usability goals. The raw results of the survey are ordinal data because the Likert scale uses order (or rank) and one cannot consistently and correctly define the distance between the categories. Therefore, inferential statistics based on means and standard deviation, or analysis of variance are not appropriate methods for analyzing the results (Agresti, 2002). For ordinal data, it is recommended to use methods that preserve the ordering of the data so that there is no loss of power, such as computing the median and the mode. In table 14 the analyzed results of the usability survey are presented. The table columns represent the five usability goals, while the rows represent the median and the mode of each question from the survey.

Table 14. Results of the usability survey at the end of the case study in SS 2017

	Effectiveness		**Efficiency**		**Utility**		**Learnability**		**Memorability**	
	Median	**Mode**	**Median**	**Mode**	**Median**	**Mode**	**Median**	**Mode**	**Median**	**Mode**
Q1	4.5	4	5	4	5.5	5	5.5	6	5	4
Q2	5.5	6	5.5	6	5.75	6	6	7	5	5
Q3	6	6	6	6	6	6	6	7	6	6
Q4	-	-	6	7	-	-	-	-	6	7

The results, in general, show that the participants in the survey have positively rated the tool on the five usability goals because the lowest score on a question was four, the highest seven. Out of the five categories, the ones that fared the best was *Learnability* with the highest median and mode of the answers, followed by *Utility*, *Efficiency*, and *Memorability*. The goal that fared the poorest was *Effectiveness* with the lowest median and mode of the answers.

The representation of the two-part survey concludes the presentation of the outcomes and results of the conducted case study for the "Insights" analytics prototype. In the next section, I discuss and analyze the results and outline the main findings and conclusions from the results with regards to the hypothesis of the study and indicate its limitations.

7.2.3 Case study findings

The goal of the case study was to find out whether professors and teaching assistants in blended learning scenarios within a higher education institution would use learning analytics tools on regular basis and incorporate them in their day to day teaching activities within the learning platform. The underlying hypothesis of the case study was that teachers while doing other teaching activities within the course on the platform, they would also use the analytics prototype within the same session on the platform. The case study itself can be labeled as quasi-experiment because the participants were not randomly assigned to the conditions of the tool; the usage of the “Insights” module was not randomly sampled among the teaching staff because as a study in the field the environment, conditions and factors could not be controlled. However, the presented results in section 7.2.2 show that the experimental design provided enough control and provided substantial evidence that the goal of the study was fulfilled.

The fact is, in 40 courses out of 53 the teaching staff has used the “Insights” module at least once during the semester, and in 25 courses the teaching staff used the “Insights” more than five times during the semester. This means that almost in half of the courses, the teaching staff explicitly used the “Insights” module on multiple occasions. This statement is corroborated with the median and mode of the number of usages, because the mode was equal to one, and the median equal to seven. The mode appears within the lower half of the usage frequency data, and at least in half of the courses, the teaching staff has used the “Insights” module on multiple occasions. The weekly distribution of courses in which “Insights” module that had been used also showed regular weekly peaks and troughs on the weekends, or on the lecture-free week in June. The aggregated analysis also showed that towards the end of the semester the number of courses in which the “Insights” module was used, steadily decreased which was also expected. What was unexpected was the fact that in the weeks right after the lectures ended, the number of courses in which the “Insights” module was used started increasing. One possible explanation for this could be that the teaching staff wanted to observe and evaluate the students’ behavior over the span of the entire semester within the course room on the learning platform.

The findings presented in the previous paragraph were summative because they dealt with the overall number of courses over the given amount of time. This means that there was a possibility that within the weekly peaks there could have been many courses with incidental usage (only once, although such requests and usage was filtered out from the raw data); or there were many courses with intentional usage of the “Insights” module. The “Insights” usage per individual course represented the aggregated data in Figure 49, Figure 50, Figure 51, and Figure 52 that in every month the number of courses (each represented with an individual line on the line chart) corresponds with the number of courses with the weekly distribution. Furthermore, the number of usages of the “Insights” module per course is in most cases (especially in the peaks) larger than one by a big margin. This can be interpreted as usage of the “Insights” module on regular basis in many individual courses and as such associates well with the first part of the goal of this case study, namely that the teaching staff would use analytics tool on regular basis within their course on the learning platform. The results of the analysis also showed that the teaching staff in 36 courses have performed various teaching activities whilst using the “Insights” module within the same session on the learning platform. In 24 courses the teaching staff had performed activities that provide various learning resources (materials, slides, media, and hyperlinks) to their students. In 26 courses the teaching staff had performed activities that distributed various course information to their students, and in seven courses they corrected assignments or provided new ones within the same session. This is a strong indicator that the teaching staff used the “Insights” module as part of their teaching activities and confirm the second part of the goal of the case study and confirms its underlying hypothesis. Considering the results that confirmed the goal and the hypothesis of the study, it is safe to conclude that the correct place for providing learning analytics solutions and visualizations in blended learning scenarios is the course room on the learning platform. Nonetheless, this corroborated outcome does not provide evidence of whether the teaching staff understood, observed, or even acted upon of the visualizations and analytics results while using the “Insights” module. These findings show only that the teaching staff used the “Insights” module on regular basis.

The two-part anonymous survey collected information and its analysis provided qualitative feedback about the features and the user interface of the “Insights” module. In total, only eight participants from the case study decided to participate and provide their feedback and answers to the survey. In comparison with the usage frequency and number of participating courses, the response rate was comparatively low. The results from the qualitative feedback showed that the teaching staff used the visualizations and analytics to get an overview about how the learning resources were used and be more aware of the student behavior in the different modules of the course room on the learning platform. Despite the inconsistencies with the data representation and the lack of help and documentation to guide them through the interface, they were aware of what was happening in their course, and their assumptions about intermittent learning were confirmed by the “Insights” module. The suggested improvements in the tool were directed towards improving the data visualization and interface, rather than providing new and different analyses and data. To better understand their perspective about their qualitative feedback in the survey, analysis on the session duration, showed that the time spent on the “Insights” module ranged from 60 seconds to seven minutes. Additionally, in relation to the time of day (in the morning, or in the afternoon) when they had used the “Insights” module is (almost) normally distributed, with a slight advantage to the afternoon. The teaching staff had relatively short sessions while using the “Insights” module, and during these short sessions, they tried to get an overview or detect trends within the analyzed learning data. This also implicates the design of the visualizations, including the data representation, which can also be confirmed by their feedback which concentrated on improving the provided visualizations and analytics representations. The usability survey showed that the “Insights” module was easy to learn to use and had good utility. However, the effectiveness of the “Insights” module was rated the poorest. This indicates that the “Insights” module needs to consider the feedback for improvements and become better at providing learning analytics in the course rooms on the learning platforms.

7.3 Pilot Case Study Summer Term 2018

This case study consisted of deploying an analytics prototype to a large number of courses (400 courses) on the learning platform. In comparison with the previous case study, the idea was to examine and evaluate the entire learning analytics infrastructure and the “Insights” prototype and inspect whether the infrastructure would scale to support a large number of courses. The context of the study was the technology aspect of the implemented blended learning scenarios in real courses on the learning platform to grasp a more realistic understanding of how analytics would be used within the learning platform in blended learning scenarios. For this case study, I also built three types of data collection mechanisms to collect corroborating evidence and with the help of data triangulation clarify and support the observations and results. For this purpose, I collected anonymous log data on the usage of the analytics prototype, collected log data on the users’ activities within these courses on the learning platform, and conducted a two-part survey. The survey consisted of questions that collected qualitative feedback about the analytics prototype, and a seven-point Likert scale usability questionnaire based on the ISO 9241/10 international standard (Figl, 2009; ISO, 2010).

This case study had two goals. The first goal was to develop an understanding of teachers' use and incorporation of analytics and statistics tools/interfaces in their day to day teaching activities within the learning platform. The idea was that the teaching staff could freely explore the analytics prototype while conducting their online teaching activities in blended learning scenarios. This information is important to explore and understand how teachers accomplish teaching tasks with the learning platform and explain how they incorporated analytics within their day-to-day activities. The second goal was to actively test and observe the analytics infrastructure in terms of performance, scalability, and maintainability while hosting and serving analytics results to the "Insights" prototype in the courses. This case study could also help in demonstrating that the teaching staff really used the prototype and describe the tools' usage patterns and activities within the course. The hypothesis of the case study was the same as the one from the previous case study, that teachers would use the analytics module on a regular basis while doing other teaching activities within the course. This was derived as a result from the previous study where the analytics prototype was seamlessly integrated within the course on the learning platform, and teachers were compelled to use it to get a glimpse of what was going on in their course.

7.3.1 Case Study Setting and Design

The second goal of this case study influenced the selection because it was necessary to select a large number of courses in which to activate the "Insights" analytics prototype. For this reason, I developed a randomized selection process for courses which considered different factors which selected courses in which the prototype would be activated. The randomization took into consideration courses which have regular use by different (large) number of students; courses that use any form of formative assessment during the semester, such as, assignments or electronic tests; and the last factor was courses that use different kinds of media (videos), literature, or relied heavy on student collaboration in their e-learning scenarios. The reasoning behind this randomization was to select a versatile set of different courses in which to activate the "Insights" prototype. This randomization process identified 400 courses (out of 1950 active courses) which constitute around 20% of the total active courses for the summer semester in 2018, and the tool was automatically activated the "Insights" module in these courses. 302 courses were lectures, out of which 66 were courses connected with an exercise, 33 courses were just exercise courses, 28 were practice-oriented laboratory courses, 18 courses were seminar courses, three courses were language courses, and there were 16 courses for which there was no information about the course type. Regarding the course distribution among different faculties, 108 courses were from the Faculty of Mathematics, Computer Science and Natural Sciences; 89 courses were from the Faculty of Mechanical Engineering; 38 courses were from the School of Business and Economics; 38 courses were from the Faculty of Civil Engineering, 37 courses were from the Faculty of Electrical Engineering and Information Technology; 32 courses were from the Faculty of Arts and Humanities; 29 courses were from the Faculty of Georesources and Materials Engineering; 11 courses were from the Faculty of Medicine; one course from the Faculty of Architecture; eight courses were affiliate courses to the university, while for ten courses there was no information available. The number of students participating in each course varied from 18 students to 2800 students. In retrospect, there were courses from all the faculties, and they belonged to most of the types of courses offered at RWTH Aachen University. According to the number of students per course, the sample size contained courses with a relatively small number of students and very big courses with more than 2500 course participants.

The analytics prototype "Insights" was available for the study participants by the end of April 2018 as an integral module inside their respective courses on the learning platform. After activating the "Insights" module in their courses, I did not provide special instructions about when and how they are supposed to use the module, but rather try to incorporate in their day-to-day activities. They received information within the module that their activities would be observed by automatic logging tools, and towards the end of the pilot phase, they received an online survey about their experiences with the “Insights” module. The survey itself was non-binding, meaning that the participants were not obliged to fill it in. The “Insights” module was never deactivated from the courses, so the teaching staff could still use it in their (now old) courses. However, only the period between the availability of the “Insights” module and the end of the semester is considered for the analysis of this case study.

7.3.2 Usage analysis results

The case study data collection methods provided enough raw data from the usage to conclude that the teaching staff used the “Insights” module. The results would be based on the collected data from the three data sources identified and selected for the case study in section 7.3. Overall, during the time of the case study, 294 courses from the 400 courses in which the “Insights” prototype was activated, the teaching staff have used the “Insights” module at least once. This means that 106 courses, the teaching staff had not used the “Insights” module at all. The usage[9] frequency of the “Insights” module during the study was not evenly distributed. There were 177 courses where the teaching staff used the “Insights” module less than five times, out of which in 56 courses it was used only once; in 75 courses the teaching staff had used the “Insights” module between six and ten times; and in 42 courses the teaching staff had used the “Insights” module over 10 times during the duration of the study. In the ten courses with most usage frequency, the usage numbers ranged from 20 to 62 times (occasions) on which the “Insights” module had been used. The usage distribution is shown in Figure 53 and in table 15.

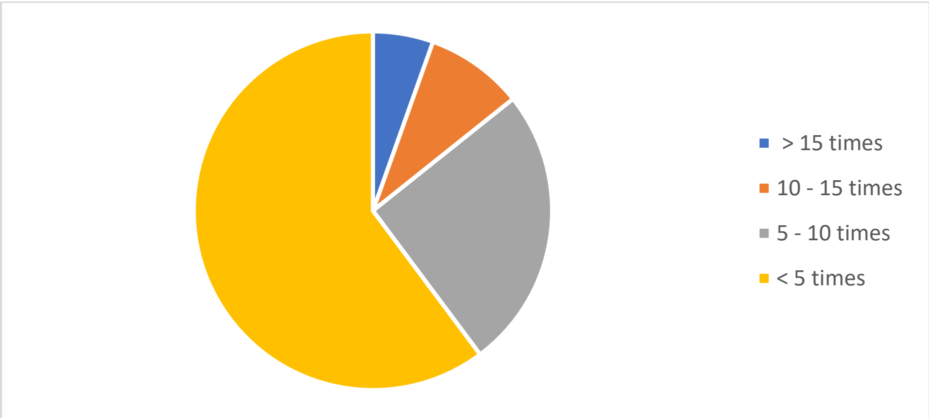

Figure 53. Summary of "Insights" uses per course

Table 15. Usage frequency of the "Insights" module in SS 2018

Number of courses	Number of uses
42 courses	More than ten times
75 courses	Between six and ten times
177 courses	Less than five times
106 courses	No usage detected

I also conducted descriptive statistics of the mean usage frequency with standard deviation and calculated the median and the mode. The average usage frequency is 5.5 times per course with a

[9] Usage in the context of this study means that a teacher has opened the Insights module in her course on the learning platform, and afterwards performed another action or activity within the course on the learning platform. This interpretation helps in excluding noise within the raw data from accidental clicks and usage.

standard deviation of 6.27, the median is 3, while the mode is 2. The average usage frequency and the standard deviation do not depict the real outcomes, because the standard deviation is larger than the mean which means that the usage frequency data is spread widely across around the mean. In this case study, the median and the mode are more descriptive and suitable for the analysis, because the median separates the higher half from the lower half of the usage frequency data, while the mode shows the value that appears the most within the usage frequency data. In other words, the mode appears within the lower half of the usage frequency data, and at least in half of the courses, the teaching staff has used the “Insights” module at least on three occasions. A large number of courses in which the “Insights” module was used, and the different number of usages in each individual course afforded for analyzing the distribution of the population. The number of uses suggested that the distribution will be skewed and there will be unusual observations (courses with a large number of accesses). Figure 54 represents the box and whiskers plot and visualizes a summary of the data. As one can observe, the data is skewed towards the left, and that in three-quarters of the courses (220 courses), the number of times in which the “Insights” module was used is less than seven times. Between the two quartiles (147 courses), the usage of the “Insights” module ranges from two to seven times. In the fourth quartile (74 courses) the usage of the “Insights” module ranged from seven times to 62 times.

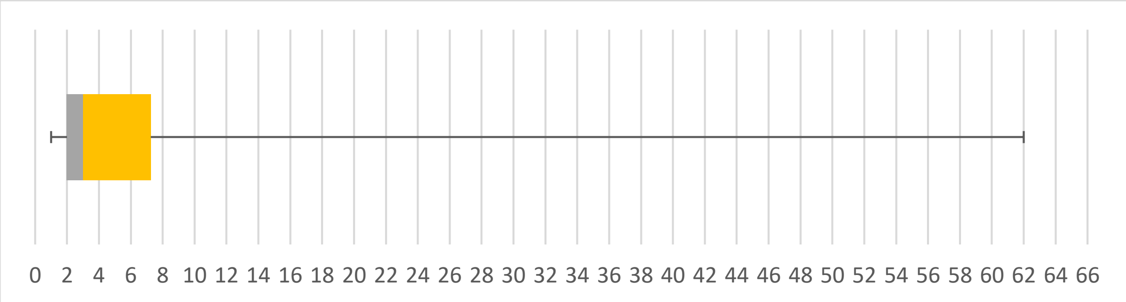

Figure 54. Distribution of the "Insights" use per course (skewed)

As part of the summative and descriptive statistics, I also distributed the “Insights” module usage across the faculties to see whether there is a significant difference between the courses in which the “Insights” prototype was activated, and in the courses, the prototype was used. As can be seen from Figure 55, in six of the faculties, more than three quarters from the courses in which the prototype had been activated, was used by the teaching staff. The biggest difference between the number of activated courses and in which the learning analytics module was used was in the courses belonging to the Faculty of Georesources and Materials Engineering and the Faculty of Arts and Humanities. In the courses of these two faculties, less than half of the courses in which the prototype had been activated, was actually used by the teaching staff. This, however, is no indication of how often or how extensive was the use within the modules.

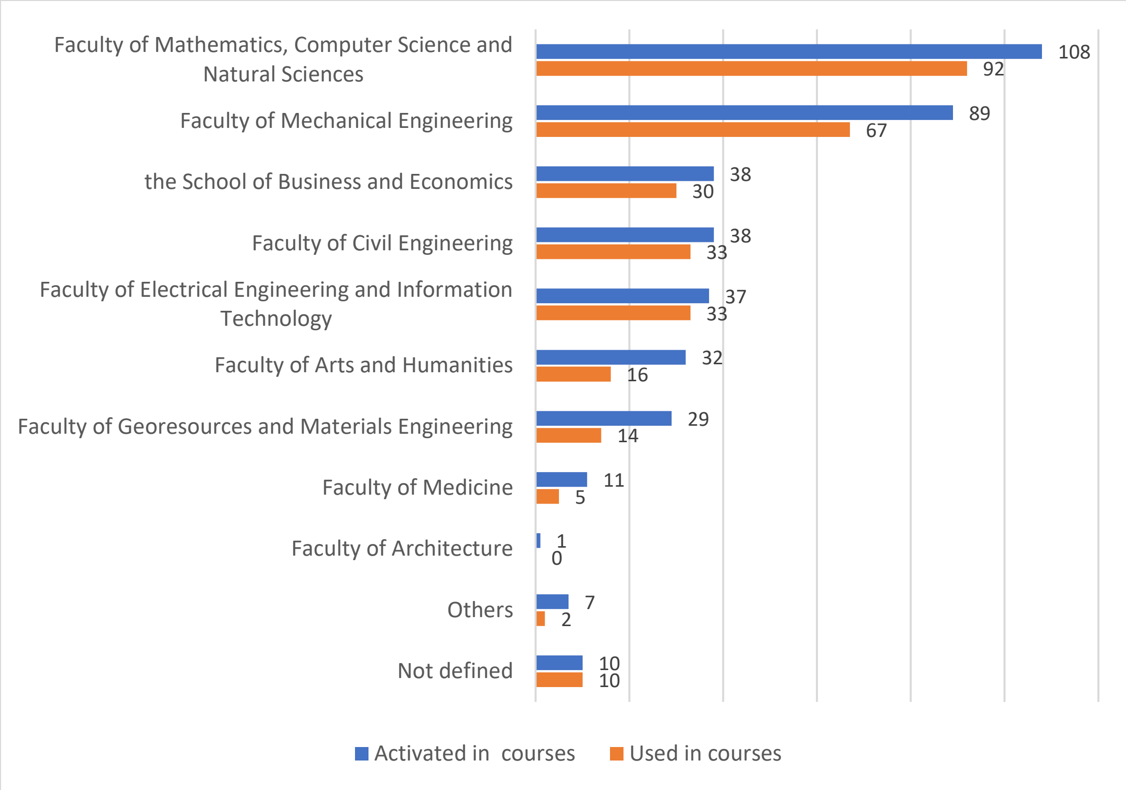

Figure 55. Courses activated vs. Courses used per faculty

In Figure 56 can be observed how the usage frequency of the "Insights" module developed over the course of the case study. On the *x-axis* is the timespan of the study, while on the *y-axis* is the number of different courses from which the "Insights" module was accessed over the given time-period. After the initial peak of usage when the "Insights" module was available to the teaching staff, regular weekly peaks of the module's usage can be identified, troughs on the weekends, and the activities in the "Insights" module decrease towards the end of the semester. The biggest peak of usage can be observed on the day the two-part anonymous survey was released and the teaching staff from all of the courses in which the learning analytics module was activated. In Figure 57, one can see the correlation of courses and the accesses to the "Insights" module. The large difference between the number of courses and the number of accesses of the "Insights" module can be observed which indicates that on a given day, the module was used multiple times in multiple courses.

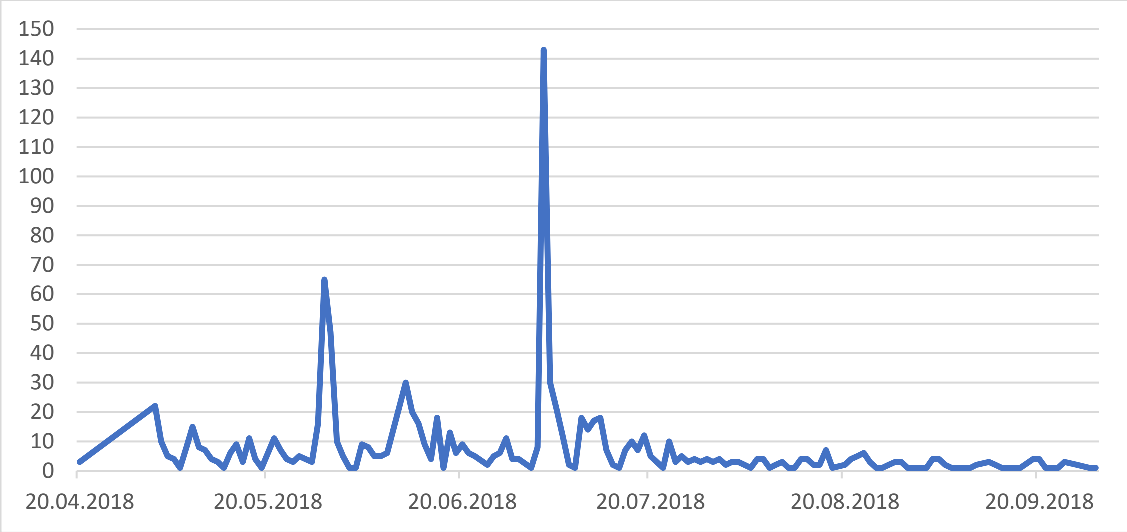

Figure 56. Number of courses in which the "Insights" module was used in SS2018

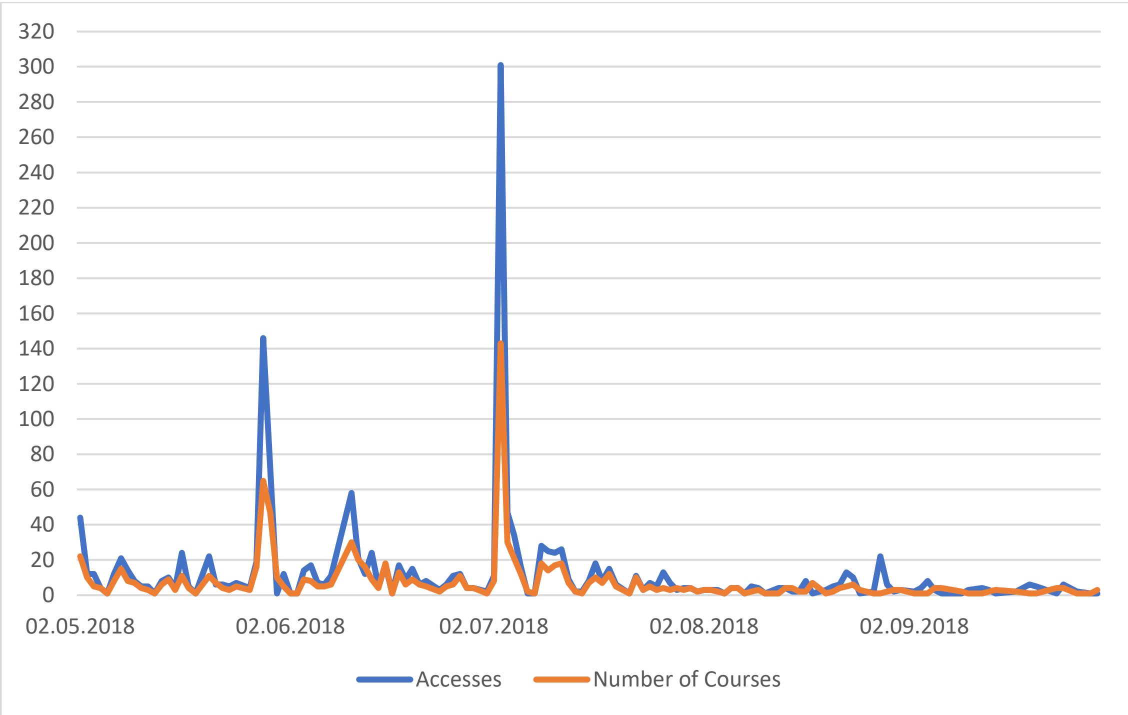

Figure 57. Correlation of number of courses and number of uses in SS2018

This can be attributed to the fact that the semester in German universities is divided into “Semesterwochenstunde” or SWS, to indicate the time students need to spend on a course per week and to measure the teaching load of the teaching staff. For this reason, I analyzed the distribution of the tool’s usage within the course over the semester divided into weeks to understand the usage distributed over different courses spread over the weeks of the semester. Figure 58 shows the “Insights” usage per course distributed over the weeks of the semester.

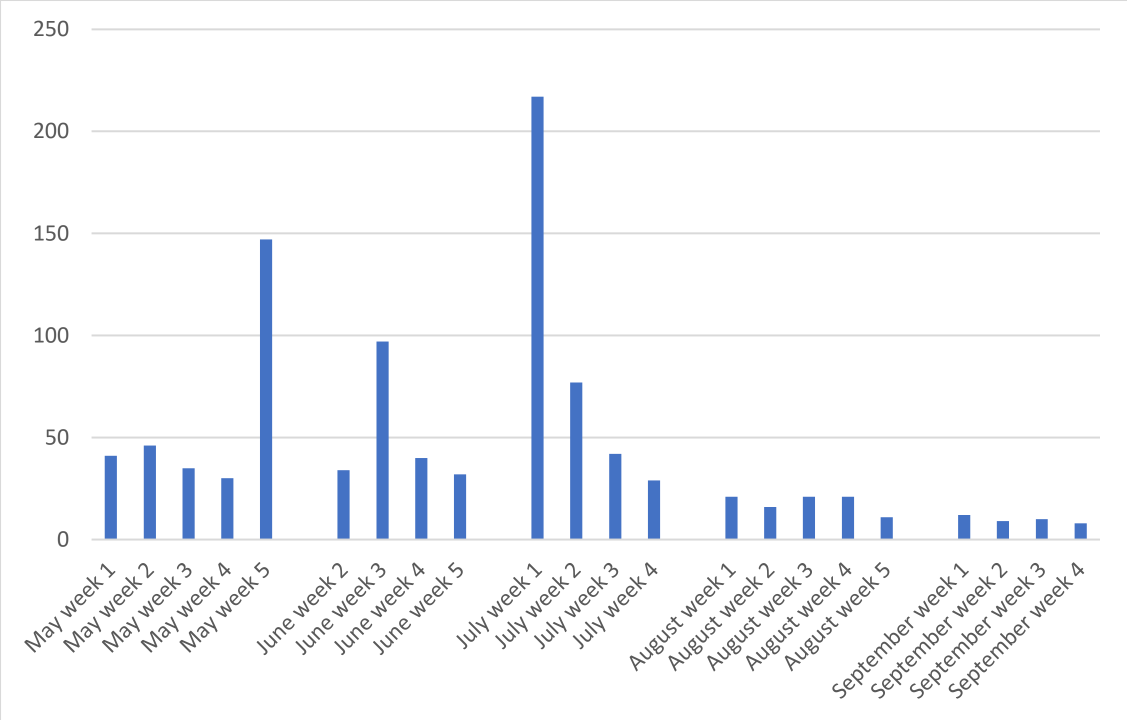

Figure 58. Number of courses per weekly summaries SS2018

In the first week of July is the largest number of courses in which the module was used. This is because this was when the survey was provided via e-mail to the teaching staff and most probably many opened it to look at the "Insights" module in their course rooms. The last week of May was the excursion week (a holiday week in which there are no lectures), and there are two major peaks of usage at the end of May and in the second week of June. The number of courses in which the learning analytics module was used ranged from 30 to 80, with the two peaks at the end of May and the first week of July. The course lectures ended by the second week of July and the exam phase starts immediately after the end of the lectures and the exam phase lasts in some cases until the end of September (which is also the end of the semester). After the lectures ended, the number of courses in which the "Insights" module was used steadily decreased.

Just like in the previous case study, I inspected whether the teaching staff logged in on the learning platform to explicitly use the "Insights" module, or they used it while conducting other daily teaching activities. This inspection was based on the collected log data from the learning platform used for providing analytics and insights about the students' activities within the course. This collected log data was re-purposed and re-analyzed to identify and aggregate the teaching activities within the course rooms on the platform. The analysis methods were designed to discover whether, within the same session[10] in the course on the learning platform, the teaching staff had performed teaching activities while they used the "Insights" module. The teaching activities covered with the analysis were chosen based on their relevance and influence on the learning processes within the course. The teaching activities that were taken into consideration for the analysis were divided into four major groups: (1) *information distribution activities*, (2) *course organization activities*, (3) *distribution of learning resources*, and (4) *formative assessment activities*. In total, in 190 courses the teaching staff has performed at least one teaching activity whilst using the "Insights" module. The distribution of teaching activities correlated with the usage of the "Insights" modules follows in more details:

[10] The term "session" is used in the context of an interaction session, when the user has logged in onto the system via a web browser and has performed different activities and interactions within the system.

(1) The teaching staff can distribute this information either by posting an online announcement or send out an email to the students. The correlation of the log-data analysis showed that in 68 courses when the teaching staff used the “Insights” module within the same session, they have posted an announcement. When sending an email in 68 courses, the teaching staff had used the “Insights” module within the same session. The combination of the two course lists to remove redundancy showed that in 112 courses the teaching staff had distributed various course information with the modules for information distribution, within the same session when they had used the “Insights” module.
(2) The correlation of the log-data analysis showed that in seven courses when the teaching staff used the “Insights” module have also created or edited course events in the calendar and that in 17 courses the staff has created/edited a survey in the same session when using the “Insights” module. After removing the duplicates, in 22 courses the teaching staff had conducted course organization teaching activities and had used the “Insights” prototype.
(3) The correlation of the log-data analysis showed that in 140 courses when the teaching staff had provided or uploaded learning materials has also used the “Insights” module. In five courses they had uploaded or embedded lecture videos. Moreover, in five courses the teaching staff had provided online resources as hyperlinks, and in four courses they had provided scientific literature to the students. The combination of all four course lists resulted in 141 courses in which the teaching staff had provided learning resources within the same session when they had used the “Insights” module.
(4) The correlation of the log-data analysis showed that in 31 courses when the teaching staff provided or edited an assignment had also used the “Insights” module. In 11 courses, the teaching staff used the “Insights” module when correcting student submissions of the provided assignments. In 24 courses the teaching staff had provided results of the formative assessment strategies via the gradebook. Combining the three course lists resulted in a total of 46 courses in which the teaching staff had performed activities within the formative assessment modules within the same session when they had used the “Insights” module.

7.3.3 Survey Results

The two-part anonymous survey provided qualitative feedback about the “Insights” module, and a usability questionnaire based on the ISO 9241/10 standard was performed to assess the usability of the prototype. The survey was distributed automatically via an e-mail to all of the teaching staff who had courses in which the “Insights” module was activated. Overall, 1382 persons from these 400 courses were contacted, and the survey was filled in 64 times. 18 participants from 18 courses declined to participate in the survey and requested via e-mail/telephone that the “Insights” module should be deactivated from the courses because they felt uncomfortable observing the students’ behavior in the course rooms. The survey had a slightly different structure and had additional questions when compared to the one from the previous study. The survey design was branching to collect feedback from users who had used the learning analytics module, and from users who had not used the learning analytics module. In the following paragraphs, the qualitative feedback from the survey is summarized and presented to ease its interpretation and each question serves as paragraph heading.

(1) Have you used the Insights analytics module in your course room? On this question, 38 responses were “No”, and 26 responses were “Yes”. This question was the branching question. The answer “No” branched towards the question that collected feedback about the reasons for not using the module, and then the survey was finished for the user.
(2) (if No) Would you like to share with us the reasons for not using the Insights module? The feedback from this question was mixed. Six responses did not contain any

information nor feedback about why they would not use the module. In 21 responses there was feedback that they were not aware of the existence of such a module in their course rooms. Six from the participants responded that they knew that there was a learning analytics module in their courses, but they either did not have suitable learning scenarios or thought that their courses would not be suitable for applying/using learning analytics in them. Four of the participants responded that they knew about the learning analytics module in their courses, they knew what it is, but they said that such information would not bring them anything new. One of the answers contained a statement that he was not particularly keen on usage statistics, so he would not use it because the accuracy of such statistics would be biased. Another answer contained a statement that the person already knew everything about his students and was not sure how this module could help him. There was one answer that plainly stated that the participant would not use diagrams and charts in his courses.

(3) (If Yes) What were the most positive aspects of the Insights module? The answers to this question provided views that were in line with the overall intention of providing the analytics module in the course rooms. The results were aggregated into five clusters of responses. In four responses the good overview of all students' activities and the design of the charts were provided as feedback. In seven responses the possibility to view and review the accesses and downloads of the learning resources modules and their correlation with other student activities in other modules. Moreover, the users appreciated the possibility to see which learning resources and how often they are used, which helped them to predict how many students were actually present in the course room, and when did the students use them. The next cluster consisting of seven responses revolved about the opportunity to be able to consider, view, and review the students' behavior within the course room on the platform, and outside of the course reviews and evaluation forms from the students at the end of the semester. Another positive aspect was the chance to help them identify and comprehend how hard the students had used the course room, or at which point in time the students had accessed which type of information and identify which learning resources were the most interesting for the students. One response exemplified that the "Insights" module provided a good presentation of the students' behavior in the course. The penultimate cluster of responses collected responses revolving around the possibility to see information and data about the course room usage, for example, how many students really went inside the course room, or how many had accessed the learning resources and whether the students really had prepared themselves for the lectures. The last cluster of responses consisted of only two answers which stated that the most positive aspects were that (in fact) there is finally a module that provides analytics and statistics to the teaching staff.

(4) (if Yes) What were your negative experiences in the Insights Module? (name two-three aspects) The responses for this question resulted in three clusters of responses and several responses that could not be coded into the clusters of responses. The biggest cluster had seven responses which focused on the layout and design of the "Insights" module. The users felt that there were too many analytics indicators displayed on one page. They felt that the number of indicators made it difficult to understand all of them. There was one suggestion to enable the possibility to select a single indicator and concentrate on it. One of the responses also stated that the colors of the visualizations could be better if they were different. The second cluster contained five responses on which the underlying topic was the lack of help and documentation concerning the indicators in the analytics module. These users would have preferred to receive (detailed) descriptions and explanations of the charts and the visualizations. The third cluster of responses contained four responses in which the underlying negative experiences were the completeness and the trustworthiness of the visualized data. The users had serious doubts whether the data really comes from the course in which it was displayed, or why there were activities in modules which were not used by the teaching staff in their learning scenarios. One particular response was that the user wondered why there would be a lot of activities from the students in the "Hyperlinks" resources module when he explicitly did not use this module to provide learning resources. In three responses there were no negative aspects, while two mentioned that the module was sometimes slow, and two did not provide any response to this question. The first of the three responses which did not fit in any of the clusters was that the user interface of the module was only provided in English and that it would be better if it were in German. The second response was that it is impossible to select an individual learning resource and analyze it in more details. The third response was that the provided analytics and statistics did not help her to improve the learning offerings in the course room, because of the way she used the course room.

(5) ***(if Yes) Have you discovered, or learned anything from the data visualizations? Which features would you like to use?*** The responses to this question converged into three homogenous response clusters. However, this question also had the biggest number of empty responses (six responses). The biggest cluster contained nine responses which were concerned with the discovery of the students' learning behavior in the course room and their interaction with the learning resources. The users learned whether the students used the learning resources and if yes, when and which learning resources they had used over time. One response goes farther with an emphasis on the correlation between the teaching activities and its influence on the student behavior. Another response focused on whether the students prepare for the lecture, and which topics they found most relevant for the lecture, and this information would help them in estimating how many of them are actually involved in the lecture during the semester. The third response contained a finding that the students would not download the documents but read them directly in the browser on regular basis. A fourth user discovered that the Exam Results module was used on regular basis, although the module was inactive. He wondered whether there is an error in the data because the students had used the module although nothing was displayed there. The second cluster contained seven responses concerning the discovered knowledge and indicators about the usage of the learning resources in the course room. In four of these responses, the names of the most interesting indicators were mentioned directly, and these indicators were the ones that showed the access and download of the provided learning resources. One response contained that a new finding was that the provided learning resources were not fully used by the students. Another response indicated that the user already had a similar expectation about the usage from the provided usage analytics and results. The third cluster contained three responses that they have not discovered/learned anything new. In one response, it was stated that it would not help him to better understand the learning behavior of the students.

(6) ***(if Yes) Have you considered changing your teaching activities or improving your learning resources based on the visualizations?*** The responses to this question were clustered into three clusters. The first cluster contained 16 responses in which the participants have answered "No". Four of them contained clarification behind their decision. For two of them, they would not change anything for the current semester but could consider some changes for the next semesters. The third response said, although she could make conclusions from the analytics prototype, she would not make any changes because the module is too new. The second cluster contained four responses in which the users would consider initiating some changes in their learning resources and activities. One response mentioned that they took the analytics results into consideration with regards to the low in-person attendance of the course and that they would revise the lecture organization. Another response mentioned that since many learning resources were not used, they would be reviewed and analyzed for their relevance to the lecture.

Third response mentioned that after the end of the semester, they would use the results in their internal review process. The third cluster also contained four responses and revolved around the notion that they might use the analytics results if the feature was available for longer periods of time (on regular basis). One response mentioned that, if there is an export function to export the results and then he can analyze them, he would trust the analysis. As such, the feature is too new and not trustworthy. Another response mentioned that he would use this feature in the future to analyze the students' learning behavior, and their reaction to the content so that he could make his own conclusions.

(7) ***(if Yes) What needs to be changed in the Insights module to better meet your needs?*** The last question's responses were clustered in three clusters. The largest cluster contained nine responses and all of them were a request for extending the existing indicators with additional capabilities or adding new indicators. Four responses were requests for adding functionality that would enable the users to export the raw data so that they can analyze it themselves in more details. Another request was to be able to dive deep within the learning resources data to follow up on exactly how and when the students were interacting with it, and the possibility to show usage of different folders within the learning files. Another feature request was to enable the user to remove unwanted or unnecessary indicators from the interface, or automatically to hide the modules that were deactivated or the modules that were sparsely used. Another suggestion was that there should be created a Help and Documentation page that describes in more details the meaning of each indicator, and on what kind of data it was built upon, and that the interface should be available in German as well. The second cluster contained six responses about improving the user interface and reducing the number of indicators. Furthermore, three answers suggested that the indicators should be rearranged to further improve the module itself. One suggestion was to allow user grouping of the indicators, and selection to quickly receive an overview from the analytics module. The third cluster had three responses and all of them claimed that nothing should be changed. There were two responses that did not belong to any cluster and suggested that the data sources and the analysis methods should be checked and evaluated because according to them they did not represent the students' behavior accurately in the course rooms on the platform. Lastly, there were five empty responses to this question.

The goal of the second part of the survey was to collect feedback and information about the usability of the "Insights" module. Usability in an interface ensures that the interactive system is easy to learn, effective to use, and enjoyable from the user's perspective. Usability involves optimizations of the interactions and the look-and-feel of the interface to enable the users to carry out their activities. Usability can be broken down in these goals: effective to use (*Effectiveness*), efficient to use (*Efficiency*), have good utility (*Utility*), easy to learn (*Learnability*), easy to remember how to use (*Memorability*). The complete list of questions and to which of these goals they belong to is presented in table 16. The questions themselves were created based on to ISO 9241/10 standard. The seven-point Likert scale ranged from "Strongly Disagree" to "Strongly Agree" (1-7) and was used as a ranked order across all 16 questions to receive more consistent results for the different usability goals.

Table 16. Usability survey questions as part of the case study evaluation in SS 2018

Effectiveness	Q1	I can effectively complete tasks with this system.
	Q2	The presented information helps me to effectively perform my tasks.
	Q3	It is easy to find the information I need.
Efficiency	Q1	I am able to perform tasks efficiently with the Insights module.
	Q2	Whenever I made a mistake, I was able to quickly correct it.
	Q3	The Insights module behaves as I expect it to.
	Q4	I did not notice any inconsistencies in its use.
Utility	Q1	The Insights module has all the features I expect it to have.
	Q2	I would use the Insights module systematically for a long time.
	Q3	Overall, I am satisfied with the Insights module.
Learnability	Q1	It was easy to use the Insights module.
	Q2	It was easy to learn how to use the Insights module.
	Q3	I feel comfortable using the Insights module.
Memorability	Q1	I did not notice any inconsistencies when I used the Insights module.
	Q2	The user interface of the Insights module is pleasant.
	Q3	I like the user interface of the Insights module.
	Q4	I did not notice any inconsistencies in its use.

The raw results of the survey are ordinal data because the Likert scale uses order (or rank) and one cannot consistently and correctly define the distance between the categories. Therefore, inferential statistics based on means and standard deviation, or analysis of variance are not appropriate methods for analyzing the results (Agresti, 2002). For ordinal data, it is recommended to use methods that preserve the ordering of the data so that there is no loss of power, such as computing the median and the mode. In Table 17 the analyzed results of the usability survey are presented. The table columns represent the five usability goals, while the rows represent the median and the mode of each question from the survey.

Table 17. Results of the usability survey at the end of the case study in SS 2018

	Effectiveness		**Efficiency**		**Utility**		**Learnability**		**Memorability**	
	Median	**Mode**	**Median**	**Mode**	**Median**	**Mode**	**Median**	**Mode**	**Median**	**Mode**
Q1	4	4	4	4	3	3	6	7	5	5
Q2	4	5	4	4	5	4	5	7	4.5	3
Q3	4	3	5	6	5	6	4.5	4	5	5
Q4	-	-	4.5	7	-	-	-	-	4,5	7

The results from this second round of the usability survey are different in comparison from the previous case study. They are noticeably lower and present a pragmatic picture of the usability of the “Insights” module. Out of the five categories, the ones that fared the best was *Learnability* with the highest median and mode of the answers, followed by *Memorability, and Efficiency*. The goals that fared the poorest were *Effectiveness* and *Utility*. The representation of the two-part survey concludes the presentation of the outcomes and results of the conducted case study for the “Insights” analytics prototype.

7.3.4 Findings about the usage analysis

According to the presented results, almost twenty percent of the courses in which "Insights" module was activated the teaching staff explicitly used the "Insights" module on multiple occasions. This statement is corroborated with the upper quartile of the number of usages per course ($Q_3 = 7$), meaning that at least in one fourth of the courses in which the "Insights" module was activated and used (75 courses), the teaching staff has used the "Insights" module on seven or more occasions. In relation to the total number of courses in which the "Insights" module was activated, one comes at the ratio of ~1/5, or in one of every five courses, the module was used multiple times. The usage data of the "Insights" module there are regular weekly peaks and troughs on the weekends. The analysis also showed that towards the end of the semester the number of courses in which the "Insights" module was used, steadily decreased. What was unexpected was the fact that before and after the excursion week (last week of May) the usage of the "Insights" module spiked on both occasions. One possible explanation for this could be that the teaching staff wanted to observe and evaluate the students' behavior over the span of a couple of weeks of no lectures within the course room. The findings concerning the usage of the module are summative because they deal with the overall number of courses and the usage in them over the given amount of time. This means that there was a possibility that within the weekly usage peaks there could have been many courses with incidental usage (although such requests and usage were filtered out from the raw data). However, the usage of data analysis and the correlation of a number of courses and usages per day, showed that the "Insights" module was used intentionally especially whenever there were peaks in the number of different courses. This finding associate well with one of the goals of this case study, namely that the teaching staff would use analytics tool on regular basis within their course on the learning platform.

7.3.5 Teaching activities and using analytics results

The results of the analysis also showed that the teaching staff in 190 courses have performed various teaching activities whilst using the "Insights" module within the same session. In 140 courses the teaching staff had performed activities that provide various learning resources (materials, slides, media, etc.) to their students. In 112 courses the teaching staff had performed activities that distributed various course information to their students, and in 31 courses they corrected assignments or provided new ones within the same session. In 22 courses the teaching staff has arranged and created different course organization activities by adding various course events during the semester. This is a strong indicator that the teaching staff used the "Insights" module as part of their teaching activities and confirms the second part of the first goal of the case study and its hypothesis. Considering the results that confirmed the first goal and the hypothesis of the study, it is safe to conclude again (like the goal of the previous case study) that the correct place for providing learning analytics solutions and visualizations in blended learning scenarios is the course room on the learning platform. The teaching staff would use learning analytics tools and results in their teaching activities while conducting their learning scenarios. Nonetheless, this corroborated outcome does not provide evidence of whether the teaching staff understood, observed, or even acted upon of the visualizations and analytics results while using the "Insights" module. These findings show only that the teaching staff used the "Insights" module on regular basis. The evidence about understanding was extracted from the anonymous two-part survey.

7.3.6 Survey findings

The two-part anonymous survey collected information and its analysis provided qualitative feedback about the features and the user interface of the "Insights" module. In total 64 participants from the case study decided to participate and provide their feedback and answers to the survey. In comparison with the usage frequency and number of participating courses and considering the number of people that were contacted, the response rate was lower than the expectations (~20%

of the number of courses in which the “Insights” module was used). All answers were anonymous and voluntary (meaning that the participants could choose which questions to answer and which questions to omit).

Part 1. Qualitative feedback

The main reason behind why the participants had not used the “Insights” module was the lack of information and knowledge that such a module existed. However, there were also users who were aware that there is a learning analytics module in their course room, but they did not have the suitable learning scenarios which could benefit from the analytics results. Moreover, there were also members who did not believe that learning analytics could help them nor bring new knowledge about their students. Concerns about the accuracy and reliability of the analytics results were also raised, and there were personal contacts requesting to turn off the analytics module from their course because of data privacy reasons, or plainly refusing to use/apply analytics in their courses.

The responses from the qualitative feedback showed that the teaching staff mostly used the visualizations and analytics to get an overview about how the learning resources were used and be more aware of the student behavior in the different modules of the course room on the learning platform. The teaching staff used the analytics module to learn more about the student behavior in the course rooms and comprehend what kind of behavior the students had with the learning resources, whether they used it regularly, or at which points in time they accessed which type of information in relation to the learning resources. Moreover, the teaching staff used the learning analytics module to evaluate how were their learning resources appreciated by the students which helped them to predict how many students were actually present in the course room. The analytics module also facilitated them to foresee whether the students really had prepared themselves for the lectures, and ultimately, whether the students actually were engaging in continuous learning. The findings of the most positive aspects of the “Insights” module were supported with the findings concerning the discoveries made with the help of the module. The most prominent findings were the discoveries about the students’ learning behavior in the course room and their interactions with the provided learning resources. These findings helped the teaching staff to observe whether the students prepare for their lecture; helped them predict how many students were continuously involved in the course; identify whether their teaching activities have effects on the students’ behavior; and how did the students exactly use the learning resources as part of their learning. What was an unexpected discovery in the feedback, was that the teaching staff had not discovered anything, nor learned something new because they expected the observed behavior from their students.

The “Insights” learning analytics tool encouraged (albeit unexpected in some ways) change in teaching activities, behavior, and improvement of the learning resources based on the findings from the survey. In this regard, although many participants acknowledged interesting findings and new knowledge about student behavior and their learning resources, they would not immediately change their learning resources or their teaching activities. The notion that the module was “too new” and that it could not be completely trusted with the results was a reason the teaching staff would not change anything in their lecture. If the module were available for prolonged periods of time (multiple semesters), they would consider the presented analytics results, and then maybe act upon them. However, there were clear results that the teaching staff would like to consider the new findings and analytics results more carefully and afterward would instigate changes in their learning resources and teaching activities. Another discovery was the fact that the findings from the module did initiate or helped in the course review processes after the end of the lectures. In these review processes, they would use the results to review and improve the teaching activities, the course structure, and organization, and revise and improve the learning resources for future iterations of their courses. This is a bit surprising because the future iterations of the courses would potentially benefit from the learning analytics results. Overall, the learning analytics module did encourage awareness, reflection, and initiated activities based upon the results in a measured and paced manner.

The lack of help and documentation and the color scheme and style of the visualizations were considered as negative experiences with the "Insights" learning analytics module. The users felt that they needed explanations and descriptions of how to use the visualizations. Trustworthiness and correctness were other aspects that for some users was a point for concerns because they thought that the data was incorrect and incomplete because they saw usage in modules in which they have not conducted any teaching activities. However, they did not think about that maybe students checked (or expected) for additional resources there and did not find any resources there. What was unexpected from the feedback was the lack of specific/new indicators to further analyze student behavior, or that the analytics module could not help them because of the way their course had been organized was considered as a negative experience with the module. The last part of the findings is concerned with feedback about which changes, and improvements would better meet the needs of the teaching staff. The teaching staff requested new indicators and visualization features to further analyze the collected learning data. Moreover, they wanted an export feature which exports the raw data so that they could analyze the data with greater granularity on their own. Moreover, it would be beneficial to make the learning analytics dashboard customizable, so that the user could remove unwanted or unnecessary indicators. Lastly, when the learning analytics module would be deployed in all of the course rooms on the learning platform, it should be both in English and in German and have a good Help and Documentation section which explains the interface and helps in interpreting the analytics indicators.

To better understand their perspective about their qualitative feedback in the survey, I also conducted an analysis on the session duration of the usages. The analysis showed that the time the teaching staff spent on the "Insights" module ranged from 45 seconds to nine minutes. In relation to the time of day (in the morning, or in the afternoon) when they had used the "Insights" module is (almost) normally distributed, with a slight advantage to the afternoon (like the previous case study). The teaching staff had relatively short sessions while using the "Insights" module, and during these short sessions, they tried to get an overview or detect trends within the visualized data. Their feedback about concentrating on improving the provided visualizations and analytics representations and in connection with the length of the usage sessions implicates the design of the visualizations and the data representation on the dashboards, the number of indicators displayed, and the level of details and granularity from the learning data.

Part 2. Usability Survey

The usability survey results of this case study are different and overall lower in score than the results of the usability survey of the previous case study. *Learnability* and *Memorability* fared the highest, just like in the previous case study. This means that the users learned the module quickly, even when some of them used it infrequently. The lowest results were for the goals of *Effectiveness* and *Utility,* although the module was upgraded and improved after the first case study. This still indicates that there is a room for improvement concerning the right functionalities and analytics indicators so that the teaching staff could do what they need to do with the analytics module. The effectiveness could also be improved in terms of how good the "Insights" analytics module is at providing analytics results to the teaching staff. One major concern which was related to the effectiveness of the module is the trustworthiness and the correctness of the analysis and the presented information based on the feedback from the first part of the survey. Another clarification of the poor results could relate to the fact that users had doubts, and second thoughts about interpreting the visualized data, as pointed out in the requests for help and documentation to help them better understand and interpret the visualizations of the data.

7.4 Conclusion

The conducted evaluation, the obtained results and the analysis of the findings led to the conclusion that the "Insights" learning analytics module helps the teaching staff in reflecting about their teaching activities, their learning resources and activities, and the students' online behavior in their courses on the learning platform. The results and the findings of the two longitudinal studies are summarized by the specific themes introduced at the beginning of this chapter. The results from both studies confirmed that the place for delivery of learning analytics results and learning analytics tools is the learning platform of the higher education institution. The integrated way of provisioning learning analytics service in the same place where the teaching staff conducts the rest of the online teaching activities, enables them to use these tools and services on regular basis. This was also evident in the usage analysis results showed, the teaching staff used the "Insights" module in the same way as they used other modules within their courses and incorporated it in their day-to-day teaching activities, thus making the activities in this module as part of the teaching activities.

The "Insights" module helped in increasing the teachers' awareness of the students' behavior and they were able to detect trends in online activities, reflect upon them and correlate them with their previous knowledge and experiences and devise some activities based on them. The reflection and action are heavily dependent on the results of being aware and understanding activities of people in one's context. Awareness helps in deciding the next steps for reaching certain goals (Dourish & Bellotti, 1992). The "Insights" module provided information and different indicators in a learning dashboard which summarized the students' online activities and behavior inside the course room on the learning platform. This information must be presented well and summarized so that the users could quickly grasp and understand what is going on the course room. This is crucial because the analysis of the length of the user sessions ranged from 45 seconds to nine minutes, and if one factors the repetitiveness of the usage concerning the median and the quartiles, the total usage in time per person is in the range of 30-60 minutes per month (in some courses even less). Hence, the number and selection of indicators have to be selected appropriately. The "Insights" module already does some pre-selection of indicators, so that only relevant ones are shown, and the loading times should be fast. In the regard of performance, the entire infrastructure was performant and responsive and delivered the analytics results in a matter of seconds. From the results, one can observe that the users (which provided feedback) became aware of what is going on in their courses and tried to understand them and associated these newly learned facts with their previous knowledge. This is clear because there are distinct indications that they were looking for knowledge that confirmed their assumptions about their students and courses, or they tried to look for data outliers, or in some cases errors in the analytics to explain inconsistent and unlikely students' behavior within the course room.

Specific examples of awareness, reflection and action can be found in the qualitative feedback from the survey where the awareness of the degree of usage of the learning resources led participants to further examine the learning resources regarding their relevance to the courses, or the awareness of the non-existing continuous learning and course preparation to review and adapt the course design to increase the attendance during the semester. They appreciated the immediate feedback from the indicators and the conveyed information about the students' behavior and activities within the course room and they acted on this information to re-evaluate their teaching and reflect upon what could be done better in the learning scenarios and the learning processes. This also showed that awareness could lead to further activities with regards to reflection and action. Moreover, a quick observation in the findings shows that the "Insights" module helped in inspiring or identifying actions which could create a difference and improvements in the learning resources or review the courses in relation to the students' behavior and the learning resources.

Overall, it can be concluded that the "Insights" module had an impact on reflection and action in the cases where the analytics results were meaningful to the users that used them, and the users felt that there is an actual benefit for them. The meaningful visualization and data were strongly dependent on the user and his interests, the courses and their learning scenarios, and based on the findings from the survey, these factors varied across the different users from the different courses. In many cases, the offered analyses and indicators fulfilled the task to initiate reflection within the users, but there were also cases where this was not the case. Getting what is meaningful and useful to every single user and every single course is a difficult challenge that was not completely alleviated with the several iterations and requirements engineering methodologies. According to the findings, although the analytics indicators were based on extensive requirements engineering and elicitation there were indicators on the analytics dashboard that were deemed as unnecessary, or not useful for the teaching staff. One remedy of this situation could be the possibility to let the users define and choose the appropriate indicators and metrics to better fit their goals. On the other hand, whenever there were information and indicators that showed something that the users considered wrong, or incorrect, would simply label the tool as not trustworthy or valid and question the tool and its validity.

Considering the results from the usability studies, the users had a pleasant user experience with the "Insights" module. Nevertheless, the results from the formative evaluation of the tool and the findings of both case studies confirmed the assumption that the competences concerning data analysis, statistics, and interpretation of analytics results among the ordinary users are relatively low, and they massively differ from user to user. There were users who had background knowledge about analyzing data and they requested export functionality so that they can further analyze the data. On the other hand, there were requests about detailed explanations, and help and documentation functionalities to help the users better understand the visualizations and the indicators. This was also confirmed in the formative evaluation stages when there were users who struggled with two-dimensional charts and visualizations and their connection to the additional knowledge represented by them. This issue can pose difficulties because if the visualizations themselves are challenging and generate high cognitive load, the users would actually focus on the visualizations, and not on what the visualizations are trying to show them. The analytics competences heavily influence the delivery methods and, in a way, limit the data visualization strategies and methodologies for the purpose of achieving a higher acceptance rate. However, a possible negative effect would be that the experienced users would like to have more sophisticated indicators to keep them interested. A deployment and acceptance strategy with proper Help and Documentation section ought to address this discordant situation.

The obtained results of the conducted evaluation with the two case studies provided comprehensive and versatile descriptions of the interaction dynamics of a learning analytics module in blended learning scenarios in a higher education institution. Both case studies were in their essence quasi-experiments because of the lack of control during their duration. As such, in both studies, there can be an existence of other hypotheses, and different explanations and interpretations of the observed results. Moreover, both case studies were not immune to the three major concerns with any case study research method: the research/methodological rigor, the subjectivity of the researcher, and the external validity and generalizability of the results. Considerable effort was invested in designing both so that they can be easily reproduced, and the methodological rigor to base the acceptance the hypotheses of the studies based on quantitative data. As mentioned in the description of both case studies, the data collection for analysis was automatically collected, cleaned, and analyzed, to remove the influence of human error. However, the methods of how to analyze and interpret the results from the study in relation to the goals of the studies and their underlying hypotheses were still subjected to the personal interpretation of the researcher. The setting of both studies and their implementation can be reproduced and carried out again with different courses, and with many courses (1000+). However, since the environment and conditions are not controlled, it is not possible to predict the behavior of the participants, without actually observing their behavior. Hence the results would be limited to describing the phenomenon at hand, rather than predicting the future behavior of the participants. In other words, there is no quantifiable certainty that by repeating both case studies the results would be the same as they are now. One way to remedy this situation is to involve repeated observations of the same variables (or participants) over a longer period of times so that enough longitudinal data could be generated and analyzed.

Considering all courses from different faculties, different course types, learning scenarios, the “Insights” module proved to be widely suitable for deploying LA tool at scale in blended learning scenarios. The thorough documentation of the two case studies can be taken at face value and applied in another higher education institution that provisions blended learning scenarios in their learning processes. The collected feedback together with the findings from this evaluation was particularly useful to finalize the entire sustainable learning analytics infrastructure and deploy it as a service at RWTH Aachen University. The next chapter presents the finalized process for scaling up learning analytics, by summarizing all steps and measures with the most important guidelines, decisions, and recommendations that can help future researchers, practitioners, and institutions in their strive for better learning experiences enhanced with learning analytics.

8 LESSONS LEARNED FOR SCALING UP LEARNING ANALYTICS

The content of this chapter presents the comprehensive study of the research topic of 'Scaling up Learning Analytics in Blended Learning Scenarios in Higher Education' through the work in this dissertation. Following the design-based research method introduced in chapter 3, the entire process of scaling up analytics and the process of developing and deploying learning analytics on scale is described in this chapter. The structure of the chapter follows the sub-research questions defined in chapter 1 and consists of four bundles that are comprised of requirements engineering and elicitation methods and guidelines including the requirements themselves; privacy guidelines and approaches how to handle these issues in practical scenarios; provide strategies about provision and acceptance strategies for introducing learning analytics as an integral e-learning service for the stakeholders in blended learning scenarios in higher education, technical specifications and recommendations concerning learning data management, analysis algorithms, deployment and provision strategies for learning analytics and the technical implementation of the hardware and software infrastructure; and lastly provide experiences, knowledge and suggestions about research approaches and methods for conductive formative and summative evaluation of the learning analytics prototypes and implementation. The outcomes of this research can be taken and applied to similar projects in this research and problem domain, and as such can help future projects in the design, implementation, and provision of learning analytics services.

8.1 Collecting requirement for learning analytics

Learning analytics services and systems are interactive software systems with strong emphasis on data analysis. Their designs are influenced by many factors among which are the choices and requests from the users, the implementation and usage scenarios, and the technical infrastructure. One of the biggest challenges of building learning analytics services (infrastructure) is deciding precisely what to build, which means that a comprehensive set of requirements has to be specified. This set of requirements includes detailed descriptions of how the system should behave, adequate descriptions of the system properties and attributes, the behavior of the system under various conditions (environmental, technical, and legal), and the users' view of the system coupled with the human-computer interactions. In general, there are four levels of software requirements applicable to the software perspectives of the research field of learning analytics: user requirements, business requirements, functional requirements, and non-functional requirements. During this dissertation, different requirements elicitation strategies were incrementally applied for gathering, analyzing, and specifying the requirements for developing and deploying learning analytics as a service.

In the course of this dissertation, the following requirements engineering methods were applied: the innovation strategy and market research technique Outcome Driven Innovation, exploratory data analysis, literature review, and document analysis, brainstorming sessions, surveys and questionnaires, and interviews. The combination of the applied requirements engineering methods provided good results in terms of capturing the user and functional requirements concerning learning analytics tools and systems. Furthermore, the entire requirements engineering process uncovered and provided additional requirements and users' requests that are closely related to the general provision of e-learning solutions, services, and the connected goals, expectations and ideas for the improvements of these services which were according to the research field, not in the scope of learning analytics. This is traceable within the results of the applied innovation strategy with the exploratory data analysis with both the teaching staff and students (sections 5.2.3 and section 5.2.4), and the semi-structured interviews conducted with stakeholders from the administration and the IT staff (section 5.2.5). On the other hand, the literature review and document analysis results which were exclusive to the research field of learning analytics and strongly focused on it, which provided very detailed results and requirements about designing and developing learning analytics prototypes, tools, and learning analytics indicators (section 5.2.6). However, this requirements elicitation technique provided scarce and vague results about developing and deploying learning analytics requirements on a scale, and effective match of these prototypes, tools, and indicators with the actual needs of the stakeholders. What is more, the collected requirements compared to the results of the ODI studies and the interviews diverged in what stakeholders needed, and what the researchers assumed the stakeholders needed. This identified discrepancy was not accidental, because there was (and still is) not enough practical experience and professional evidence of applied analytics at scale in blended learning scenarios in higher education, and the collected indicators and requirements were basically reflections and perspectives of the researchers in the field. I solved this problem by iteratively combining the collected data and knowledge to create personas, developed their corresponding use cases, and assigned learning analytics indicators for each persona and use case (sections 5.2.7 and 5.2.8). The outcomes of these three techniques streamlined and organized the results from the different elicitation methods into a comprehensive set of requirements which were used in the design and implementation of the technical aspects of the solution (presented in chapter 6). This incremental refinement and synergy of the requirements collected from the different elicitation strategies was confirmed and assessed with the results of the formative evaluation (sections 6.4.3 and 6.4.4) and the summative evaluation (chapter 7) by showing that the functional and user requirements were suitably captured for the teaching staff stakeholder group. Consequently, the requirements for the rest of the stakeholder groups can be assessed and evaluated in the same fashion and achieve analogous positive results.

Overall, one can conclude that capturing the needs of the stakeholders in the context of learning analytics is no small feat, especially when the development of learning analytics tools relates to providing them as an integrated e-learning service at scale. This aspect influences the provision of such tools because what the stakeholders need, and what the researchers think that the stakeholders need from the tools can be (and in this case are) two different things. Hence, it is crucial to use multiple and different requirements elicitation methodologies and preferably in several iterations, and it is strongly advisable that an innovation technique from the business market research field is applied to identify and leverage the possible needs and potentials for creating and developing learning analytics tools and services. Doing literature reviews and analysis alone does not provide the perspectives nor the actual needs and directions in which learning analytics services should be developed. Moreover, the elicited requirements must be evaluated and validated with empirical data to increase their objectivity and relevance. The validation and analysis methods and processes during the requirements elicitation discovered general and unknown knowledge and perceptions about the online parts of the blended learning scenarios within the learning platform (knowledge that could not be captured in any other way). The final comprehensive requirements should be a result of several incremental iterations and reviews and specified in a pre-defined structure called Software Requirements Specification (SRS) for software projects which are afterward used as a basis for the design and implementation of the technical aspects of the interactive system which will be offered as an integrated e-learning service.

8.2 Data privacy regulations and institutional preparation

The developed sustainable learning analytics infrastructure together with the "Insights" prototype was deployed as a service to hundreds of courses. The other developed prototypes were as well designed and implemented to be deployed and used in a productive environment. The technical infrastructure and the e-learning services this infrastructure provides should be deployed, controlled, and operated within the existing institutional rules and regulations concerning data privacy, the state, federal, and European data privacy regulations and laws. On the other hand, the existing e-learning services and infrastructure that need personal and sensitive data to work correctly affect the stakeholders' right to privacy and introduce legal and institutional issues and challenges. The storage, management, and processing of personal data is a requirement not just for learning analytics tools and services, but for all the provisioned e-learning services (specifically the learning platform and the campus management system). This means that a higher education institution must create institution-wide official rules and regulations that sanction the use of technology and e-learning services within the learning processes. In this case, as part of this dissertation, and during its duration, an official process was initiated to develop an official legal framework that encompassed all systems and services that stored private and sensitive data. The university with its internal governing bodies (the university's government, rectorate, and its senate) created the regulations for the protection of personal data in multimedia applications and use of e-learning methods at the RWTH Aachen University. This official document outlined the rules which apply to all services and e-learning solutions and e-learning processes that use and are processing personal and sensitive data within the university for the purpose of scientific training, which also includes future (new) e-learning services, such as learning analytics services. The texts of this official document were drafted by the institution's legal department with technical consultations and support from me, and the data privacy officer to make sure that the regulations were in accordance with the state and federal data privacy laws in Germany, and would be in accordance to the (back then still newly, and not yet in power) General Data Protection Regulation (GDPR) of the European Union.

The regulations itself consisted of 14 paragraphs which outlined the scope of the regulations, defined the affected persons, the basic principles and rules that this regulation covered, and outlined the responsible body and its duties towards the affected persons. Furthermore, the regulations held clear distinctions and clarification among the different types of personal data and included a separate paragraph that handled research and use of personal data within this research; the consent of the affected persons; and how long such data could be legally and safely stored. The detailed description of the process and the contents of the official regulations can be observed in section 5.3. The most important practical advantage of the existence of these regulations is the fact that it can be used as legal basis for developing, deploying, or improving existing e-learning services with new features that can handle personal data. There is an existing process which must be triggered and conducted with the data privacy officer of the higher education institution whenever a new technology-based service or e-learning service (tool or system) is provided on an institutional level which is based on the official regulations that handle the protection of personal data. This process is the creation of the official documentation of the procedures and operations for handling personal data (Verfahrensverzeichnis) with a description of the official documentation of the procedures and operations for handling personal data (Verfahrensbeschreibung) for the developed and deployed e-learning application. After the GDPR came into effect, there was an update in this process by updating the official documentation for the procedures and operations for handling personal data. The updated documents contained the similar information as the previous official documents, but with an update and reference to the new wording of the GDPR regulation whose essence was incorporated into the state and federal data privacy laws.

This official registration process of the newly developed e-learning service should be conducted and initiated from the technical lead of the development team who is responsible for the scientific lead and technical implementation and support of the service. The official documents that are filled in as part of this process describe and document the collection, storage, and processing of personal data. The documents themselves should be signed and kept in a secure place and signed copies should be sent to the data privacy officer. In turn, the data privacy officer is required by law to review these documents, and by request publicly provide these documents (or parts of these documents) and make them available to everyone in an appropriate manner upon request. These documents follow a structure predefined by the state data privacy law, and they contain detailed information about the institution/organizational unit that processes personal data, on which legal bases this data is processed, the types of personal data, the affected persons, the authorized personnel who has access to the data, and the complete description of the technical infrastructure and organizational structure.

The official regulations provided the legal and the technical boundaries in which the learning analytics data management strategies can be implemented. The following practical findings outline the identified and applied strategies for privacy conformant data management for scaling up learning analytics. The strategies cover the data collectors, the storage of this data including the technical infrastructure, and the creation of data expiration policies. The data collection mechanisms should not be built to collect all the available learning data, but only the necessary learning data. The identification of the necessary learning data can only be done by analyzing the blended learning scenarios and identifying which parts of these scenarios can be enhanced with learning analytics. This means that the requirements elicitation strategies should not only concentrate on developing learning analytics tools and services, but also encompass a more holistic approach and include the analysis of the applied learning scenarios. This also helps in identifying which indicators will leverage these learning processes, and in turn, identify which data is necessary for the implementation of the indicators. The data collectors must collect the data and store it on a dedicated logical location (which can also be a separate physical location or virtualized separate entity on top of a much larger hardware infrastructure) on the premises of the higher education institution, preferably within the department responsible for the provision of IT services for the university. This data warehouse should also implement upon the paradigm of data minimalism and data expiration within a suitable data model which can deliver scalable performance and a separate logical entity should hold the analytics results which are delivered by the analytics engine that analyzes the learner's data. The access to and from the data warehouse should be standardized and implemented via two-step authentication and scalable and extensible RESTful application programming interfaces which transport the data into JSON or XML format. The uniformed and standardized extensible access enables the practical implementations of the data privacy conformant measures and provides a standardized way of future implementations concerning new changes in the official rules and regulations about data privacy on institutional, state, federal, and EU regulations.

As a concluding remark, the institutional preparation for learning analytics encompassed and revealed more administrative work and practical preparations (both legal and practical) than it was initially expected in the preparatory stages of this dissertation and the provided experiences, research, and pieces of information from the research community and the published work. Nevertheless, this long, and tedious preparation process resulted in a legal framework that provides a basis for scaling up learning analytics in blended learning scenarios in higher education.

8.3 Provision of learning analytics tools and services

The deployment and provision of a new e-learning service on top of the various existing e-learning services in a higher education institution with more than 45 thousand students and more than five thousand persons in the academic staff is a challenge that needs extensive preparation and organization. One question that is strikingly omitted in the publications in the research field and community is where and how should learning analytics dashboards, tools, and services be provided to the users. The developers work vigorously on the design and development of research prototypes, learning analytics dashboards and tools, conduct studies and research with them, but there is rarely a focus where to provide them as a service. Most of the learning analytics prototypes are usually stand-alone applications and theoretically, they should be provided to the end users as another application or service. However, the simple existence of another online system or platform does not warrant success. As part of this dissertation, a distinct effort and care were committed into investigating the place and manner of providing learning analytics services to the different user groups and the development of acceptance strategies and technical provision mechanisms for learning analytics services.

The developed sustainable learning analytics infrastructure was designed to be deployed on dedicated hardware, and the user interfaces were designed to be responsive and to fit on various screen sizes and to be embeddable within other applications (in this case, to be embeddable on the learning platform). During the requirements engineering processes, the question about where the learning analytics results should (or tools) be provided was also researched, but there was no conclusive answer or solution identified. Instead, I used previous experiences from another study conducted at RWTH Aachen University, where the analytics prototype was provided as part of the course room on the learning platform (Anna Lea Dyckhoff, 2014). Additionally, in the two conducted case studies I explicitly observed the usage of the prototype in parallel with other already established user activities and interactions on the learning platform. The findings of both case studies (sections 7.2.3 and 7.3.4) confirmed the hypothesized statement that the teaching staff would use the learning analytics prototype on a regular basis while doing other teaching activities. This is a strong indication that for the teaching staff, the learning analytics dashboards, tools, and services should be provided within the learning platform. The provision of analytics for the other stakeholder groups was not investigated but considering the results of the two longitudinal studies and considering the daily activities and interactions on the learning platform, one can be safe to assume that for the students' user group the learning platform would also be the place where the learning analytics tools (dashboard, or services) is provided. However, for the IT staff stakeholder group, and the administration, a separate and stand-alone dashboard might be more fitting to their day-to-day activities because their work and scope of interest are broader and outside of the activities on the learning platform.

The subsequent practical challenges in the provisioning mechanism for learning analytics are the deployment, technical support and maintenance strategies, with the pledge to provide a scalable, responsive and performant system, with warranted consistency and correctness. The developed learning analytics infrastructure consists of four main components (see section 5.4.2) which communicate with each other in a pre-defined and structured way. This decoupling allows for changes and updates in each component without affecting the rest of them. If deployed as a service, the downtime should be reduced to a minimum (ideally zero downtime), and the deployment and maintenance processes and schedule should be designed and implemented with the goal of minimizing the influence on the availability of the learning analytics infrastructure. During the two case studies, the maintenance measures were conducted based on calculations which reduced the effects of the deployments which provided hotfixes, and new features. The deployment and maintenance process allowed for deployments which affect any component, or any combination of the components, or complete deployment (affecting all components). There was a small watchdog service that scanned for user activities, as well as system activities and based on its results, a deployment on any of the components was scheduled and then conducted. The technical deployment methods were well integrated within the built-in services of the technical infrastructure, ensuring smooth upgrade of the existing components. The deployments themselves were automated, meaning that within the code management system there was a specifically written script which was triggered to deploy a specific stable version to the production server at night (at 23:00), or in the early morning (04:00). The deployment itself never took more than three minutes including the automated tests as a post-deployment step. This ensured that practically, there was no downtime for the users participating in both case studies, thus reducing the risk of the non-responsive or unavailable system which can positively affect the user experience. Other aspects which are connected to the provision of learning analytics as a service and have a profound effect on the user experience are the correctness and consistency of the system, and its performance and responsiveness.

In the evaluation results from the two case studies (section 7.2.2, and section 7.3.3) in the qualitative feedback questions there was feedback questioning the trustworthiness of the tool and the correctness of the data. A correct system is a critical requirement for the users to trust the system, whose trust would most likely turn into action and impact. Both developed learning analytics prototypes and the analytics algorithms went through several iterations which helped in reducing errors and glitches. The UI of both prototypes was streamlined to reduce the number of errors and interactions which could have a negative effect on the user. Moreover, the data that was used for the visualizations was double-checked and mapped with the course id of each course room it was generated from to increase the precision and accuracy of the results of the analytics indicators. The design and placement of the indicators on the dashboard were also evaluated and examined to better fit the course room and the active modules in it and followed the ordering and filtering from the available modules within the course room and. The filtering and zooming functionalities within the indicators were always activated and reset in the same fashion and with the same gesture interaction. Based on past experiences and studies where performance and responsiveness were a negative influencing factor on the user experience, a considerable effort was invested in ensuring reliable performance of the analytics infrastructure, and the responsiveness of both learning analytics prototypes. The design of the raw data warehouse enabled efficient and quick incremental transfer and manipulation of the copious amounts of raw data on daily basis. The persistent storage optimization with refactored stored procedures and the parallel execution of the analytics algorithms which used performant data structures enabled the analysis and calculation of results available for each individual course. From the users' perspective, the delivery of the results and the visualizations was instant and if the "Insights" needed a couple of seconds more to load the data, previews, and transitions within the indicators were provided. The immediate delivery was essential, because based on the usage sessions' lengths (sections 7.2.2, 7.3.2), and the number of other create activities on the platform, and on top of the users' daily routines and other responsibilities, the users do not have time to wait upon the delivery of analytics results (or any other slow and unresponsive interface for that matter). Furthermore, the review, zooming and panning functionalities were calculated on the client side (no callbacks to the server to load additional data) which provided a smooth and streamlined user experience while exploring and using the indicators on the dashboard.

As with the initiation of any new service, one has to take into account that the majority of users will need a lot of support in terms of using the tool and understanding and interpreting the results. Therefore, different policies and mechanisms for providing help and documentation about the learning analytics indicators and their analysis and interpretation. Additionally, clear communication channels and support processes must be developed to provide a structured way of collecting feedback and requests for new analytics indicators. In general, users of a system can receive help and assistance by support staff or system documentation. The need for both types of assistance was also identified from the qualitative results from both case studies (section 7.2.2 and section 7.3.6) whereas an improvement to the existing prototype was mentioned that the users would benefit from the guidance of how to correctly and easily interpret the visualizations. Therefore, before an official release as a service comprehensive system documentation with tutorials and practical tips and tricks about interpreting the visualizations. Specific approach for this was presented by Lukarov (Lukarov, 2013) by proposing a structured indicator documentation (see Figure 59) for describing the indicator, with a sample snapshot and the indicator description and which use case it covers, what kind of data and analytic methods are used, with possible questions and ways of how to interpret the indicator, and its limitations.

Figure 59. Indicator documentation method (Lukarov, 2013)

The second type of offering help and assistance is through dedicated support staff from the department responsible for providing the learning analytics service. The development of new support processes with trained staff can be a long, tedious, and financially extensive process which can also include the development of communication channels, buying or developing ticketing infrastructure, and providing it back to the users. This investment can be and should be avoided with the decision to provide the learning analytics services within the existing learning platform, as was the case with this dissertation. The learning platform already provided the base for provision with its support staff and communication channels right from the beginning of each case study. Moreover, in the support process, newly filter and categorization was developed which diverted all of the support and new features requests to the staff that was trained to provide support and training for the learning platform and including the newly available “Insights” module in the course rooms (although it was a high-fidelity prototype). Furthermore, changes within the “Insights” interface were specifically built in to improve the integration with the learning platform in the sense of activation and deactivation of the module through the configuration of the course room, and also to provide diagnostics data as an attachment to the support request if they were generated through the "Contact us" support form of the learning platform. This integration simply re-used the existing communication and support infrastructure to provide a communication channel and assistance to the users.

The last part of the provision mechanisms and strategies for introducing learning analytics services is the stepwise introduction of distinct types of analyses and visualizations for the different stakeholder groups. When introducing analytics, one must be acutely aware of the level of data literacy, the preparedness to accept data-driven decisions and in some cases "uncomfortable" information through the visualizations, and the complexity of the visualizations within the indicators. This was also evident in the analysis and of the requirements from the literature and the ones collected from the users (see sections 5.2.3 and section 5.2.4). The visualizations and the indicators should be simple (preferably with one or two data dimensions) so that beginners can understand them without devoting cognitive effort. This is crucial for the initial success of learning analytics as a service because on short-term the users need to clearly see and understand the value of the indicators by themselves. This is required to cover most of the beginners. However, there are users who are literate in analyzing data and making their own conclusions (as the study feedback pointed out that there were participants who would like to export the data and analyze it by themselves). These users or experts would like to analyze the data with respect to correlation to student behavior, performance, and other data factors that affect student performance and student success. Therefore, to keep the beginners engaged and develop their analytics skills and literacy over time, and at the same time keep the experts also interested in the tool, based on the experiences from the requirements engineering process and the two case studies, I formulated a stepwise introduction of learning analytics services.

Namely, learning analytics in higher education should be introduced in steps with continuous longitudinal studies between each step, and with a different pace of provision of new indicators and data for the different stakeholders (specifically for the teaching staff and the students). The first step is to provide basic analytics and reports that implement easy-to-use visualizations and indicators which simply show the usage behaviors of the students without correlating any data. The second step is to provide analytics based on the correlation of teaching and learning activities for the teaching staff. This way they can observe the effects of their teaching activities on the students' online behavior in their course rooms (introduction of two data dimensions over time). The third step is the provision of analytics results and indicators that correlate teaching and learning activities with performance data from the formative and summative assessment results within the courses (the introduction of four data dimensions over time). The last step is the introduction of prediction, automatic intervention, and recommendation within the analytics services (the introduction of Artificial Intelligence and smart agents). This stepwise introduction process for learning analytics is a long-term commitment process which can take several years of continuous research, development, and evaluation to embark objective impact and success in changing and enhancing the teaching and learning experiences in a higher education institution.

8.4 Evaluation strategies for learning analytics

Evaluation of learning analytics tools covers the evaluation of the interface design, the usefulness, and utility of the tool, and most importantly effectiveness and impact. In the course of this dissertation, these evaluation objectives were attained by applying various evaluation techniques from the human-computer interaction (HCI) research field and the behavioral sciences. The purpose of the two evaluation objectives was to ensure that the developed learning analytics infrastructure and, especially, the prototypes were properly built and fulfilled the collected requirements, and identify and, if possible, measure the usefulness and the impact of the learning analytics infrastructure with its user interfaces (dashboards).

The first objective for evaluation was fulfilled in an iterative way through the applied formative evaluations during the incremental development of the learning analytics infrastructure, and the development of the user interfaces of the analytics prototypes. The evaluation processes were integrated within the rapid application prototyping and development strategy (section 6.1) and started with creating paper prototypes and sketches of the interface of the indicators and the visualizations and their quick and dirty evaluations with heuristics. After the first prototypes were developed, an evaluation round without users consisted of evaluation with experts by applying heuristics and conducting a cognitive walkthrough. For this evaluation round a collection of heuristics was used from previous research, and for the cognitive walkthrough, the personas were used for describing the use cases. The most notable results from this evaluation round were that the total number of visualizations on the dashboard should not be greater than ten indicators (ideally five to seven) and that the available interactions within the visualizations should be streamlined and with the least possible clicks and effort to interact with them. The second and third evaluation rounds were conducted with users from two stakeholder groups (students and teaching staff) by conducting user studies which followed the think-aloud approach. The conducted sessions and evaluations provided results and feedback concerning the functionality and utility of the tools and as well as, the usability of the user interfaces. The user tests provided valuable feedback and identified many practical problems with the system and the interface. Moreover, they showed that the average user had problems with understanding complex visualizations (hence, simpler ones should be used) and that they really appreciated the simple language and labeling of the data points and having guidance and help through the indicators. Overall, different evaluation methods with and without users were applied to evaluate and improve the prototypes so that they are easy to use and that provided pleasant end-user experience. The choice of colors and icons with a familiar design was appreciated by the users, as well as, the smoothness and fluidity of the interface. In the end, these formative evaluation methods helped in building better prototypes and established that the implemented features corresponded with the ones derived from the requirements analysis.

The second objective of the evaluation was to identify, and if possible, measure the usefulness and the impact of the learning analytics prototypes to the users. Identifying and measuring usefulness and impact is a very challenging task. Moreover, the experiences gathered in previous studies and by the results of this dissertation showed that this process took a long period of time and repetitions of the study. I chose to conduct longitudinal case studies because it allowed greater freedom for conducting them, afforded data triangulation and collection of multiple data sources for corroboration, and thus allowed for in-depth investigation of one or a small number of cases. The two case studies were designed with the aims to detect information whether the teaching staff would use the “Insights” prototype to become more aware of their students’ online behavior, initiate a reflection process towards their teaching activities, resources, or scenarios, and potentials for initializing activities based on these new findings. These aims were rendered and transformed into concrete steps and measurable statements presented in the introduction to chapter 7. These concrete steps were selected as a choice to conduct longitudinal case studies with one stakeholder group was the correct one. Each of the studies ran in a real-world scenario with many courses from different faculties for an entire semester and an ample amount of data and feedback was collected and analyzed after the end of the semester. The results and the findings of the two case studies showed that the “Insights” module did help them to be more aware of their students’ online behavior, and it clearly initiated a reflection process partly towards their teaching activities and learning resources. However, partly the tool initiated a reflection process partly towards the validity and trustworthiness of the presented results and the tool itself. The evaluation also showed that the “Insights” module inspired an initiation towards activities and interventions within the course structure or the learning resources (section 7.3.6). However, again, this finding was also hindered with the fact that the learning analytics prototype was simply too new and not available for a long period of time, although it was available for an entire semester. Consequently, one can be safe to assume that if learning analytics tools strive to have an impact to the users, they need to establish themselves and be present for prolonged periods of time (preferably years) to provide an actionable and measurable impact on the teaching and learning processes. An impact cannot simply be achieved by having access to a learning analytics service for short periods of time and several discrete events of usage. Case studies in real-world scenarios have to run over the course of a long period of times, and when coupled with multiple data sources, telemetry data, and qualitative feedback and provide conclusive and empirical evidence about the impact learning analytics tools and services have on the users.

8.5 Conclusion

The important principles for scaling up learning analytics in blended learning scenarios in higher education focus on five aspects: collecting the correct requirements, preparing the legal and technical foundations on an institutional level, continuously develop and improve the learning analytics services, and continuously evaluate the learning analytics services. The learning analytics services should be introduced in a stepwise manner so that the users and stakeholders are not overwhelmed by the visualizations and the indicators. The presented data and the indicators must cover the needs, goals, and requirements of the stakeholders, and they should always have a say in the end results. The dashboards that hold the indicators need to be simple, consistent, and pleasant to use. They need to engage the user, have her understand what is being shown, and guide her through the reflection processes with suitable help mechanisms. The institution needs to develop internal strategies concerning data privacy, foster development of knowledge base for learning analytics, provide support on different levels (leadership, financial planning), build the technical base and infrastructure to provide learning analytics, and provide favorable conditions to build development and support teams dedicated for the provision of learning analytics. The last aspect is to continuously evaluate and improve analytics services to maximize outreach and impact. Examples and marketing campaigns should be identified and developed to demonstrate benefits from learning analytics and continuously do pilot phases and studies to introduce new and improved indicators and analytics features.

9 CONCLUSIONS

This dissertation took upon the extensive advances in the research field and outlined a process for a scaling up of learning analytics as a service for wide adoption in a higher education institution. The work the dissertation covered collecting requirements relevant to learning analytics in connection with the different stakeholder groups by applying well-established requirements engineering approaches. As part of the requirements engineering process, a business innovation strategy from the field of market research was applied to identify the needs and development potentials not just for analytics, but in broader terms of e-learning services because learning analytics offers innovative approach to enhance teaching and learning experiences, and leverage data-driven decision making in higher education institutions. As part of the process, an institutional preparation was conducted for the introduction of learning analytics services in the learning processes covering the legal aspects by developing an institutional framework to regulate the usage of personal data within the e-learning services offered at the higher education institution. The technical aspects were implemented by the design and development of a sustainable learning analytics infrastructure that used learning data to provide analytics in every course room on the learning platform. Moreover, the infrastructure also had dashboards and interfaces to provide analytics results and actionable intelligence to the administration and provide support in the decision-making processes regarding the e-learning initiatives and activities in a higher education institution. The sustainable learning analytics infrastructure was deployed, fielded and evaluated in real-world learning scenarios and reached many users through two pilot phases. This helped in evaluating the analytics infrastructure and validating its integration and utility in the learning processes and within the existing e-learning services. This set-up provided invaluable feedback, knowledge, and conclusions to create appropriate design rules and guidelines for scaling up learning analytics on an institutional level.

As I wrote in chapter one, scaling up Learning Analytics implementation as a service in Germany was a big challenge. The first step toward creating the process was to try and find earlier practical experiences about scaling up learning analytics in higher education institutions. In chapter two, I presented past experiences about institutions who had supplied learning analytics at scale and their initiatives for supplying learning analytics tools to their students and teaching staff. However, these successful examples were from institutions from the United States of America, where the universities are bound by different federal and state rules and regulations. Moreover, these examples never included the complete process, or how the actual work should be carried out so that it can be taken at face-value and followed. Furthermore, I also looked for practical experiences about how a higher education institution can be prepared for scaling up learning analytics. I found ample research, theoretical work, and frameworks that have examined the institutional preparation and produced results about which capacities should be built and fostered for the provision of learning analytics services. These results included suggestions that the institutions should build learning analytics initiatives, policies, processes and practices, infrastructure with analytics tools, and to build dedicated teams with analytics skills and values. However, the focus was that these capacities should be there, but there was no practical approach of how to build them. In the second part of chapter two, the focus was on earlier experiences from researchers and practitioners about building learning analytics tools and services. These experiences were divided into research and experiences about collecting requirements for learning analytics, research, and experiences about prototyping and development of learning analytics tools, and the last part was about research and experiences for evaluating learning analytics tools.

Chapter three and chapter four provided the applied design-based research method and the scientific foundations and terminology used for this dissertation work. I introduced and collected concepts and results concerning blended learning, data privacy in higher education, analytics in education, the research field of learning analytics and its evaluation. The concept of blended learning provided the context in which this dissertation was to be accomplished and provided the perspectives of the teaching staff and students in these blended learning scenarios. This was important because the developed and deployed learning analytics tools can only inform on the students' online behavior, motivation and attitude and the learning analytics results should be interpreted as one part of the complete picture. The concept of privacy and the right for informational self-determination were presented in the context of blended learning technologies and scenarios. The concept of data-minimalism from the state and federal data privacy laws (also present in the EU's GDPR) was followed when designing the learning data collection methods and strategies. The second part of chapter four introduced the concept of analytics and its relevant building blocks for practical application in blended learning scenarios. These building blocks included the current definitions of learning analytics, its processes, the concepts of learning dashboards and indicators, and research about evaluation concepts and strategies.

Chapter five provided the three main components that were conducted as part of the preparation for scaling up learning analytics. The first component was the engineering of the requirements, the second component was the actual institutional preparation for scaling up learning analytics, and the third part was the development of the outline for the technical implementation of a sustainable learning analytics infrastructure. The requirements engineering practices and methods were borrowed from the software engineering field, the innovation management, and market research business field, and statistics and data science. The combination of the applied requirements engineering methods provided comprehensive results which captured the user, the functional, and non-functional requirements. Additionally, they provided feedback and knowledge about issues and features that are closely related to the general provision of e-learning solutions, services, and the connected goals, expectations and ideas for the improvements of the available and existing e-learning services. The elicited requirements were structured as personas with practical use cases and were related to specific indicators for development. The institutional preparation for scaling up learning analytics covered the development of rules and regulations which sanctioned the e-learning services on an institutional level. This solution included learning analytics implementations as another tier of the different available e-learning services in a given higher education institution. The entire preparation process took more than a year and led to the development of the so-called "eLearning Ordnung zum Schutz personenbezogener Daten bei multimedialer Nutzung von E-Learning-Verfahren an der Rheinisch-Westfälischen Technischen Hochschule Aachen", or translated Regulations for the protection of personal data in multimedia applications and use of e-learning methods at the RWTH Aachen University. This official institutional legal framework provided the legal foundation and the responsible bodies for the actual implementation and provision of the different e-learning services, including learning analytics. Afterward, the learning analytics services and tools could be developed as another e-learning service and are bound to the predefined procedures and operations for handling personal data (Verfahrensverzeichnis), and the description of the official documentation of the procedures and operations for handling personal data (Verfahrensbeschreibung). The last part of this chapter provided a proposal about data privacy conformant and sustainable learning analytics infrastructure. This included the identification of the data sources, a sketch for a data warehouse, and a learning analytics solution which used the data warehouse, analyzed the data with different analytics algorithms, and delivered the analytics results for visualization.

Chapter six presented the practical approaches that were taken for the implementation of the sustainable infrastructure, the analytics data management, and the user interface evolution through prototyping. The last part of the chapter outlined an implementation strategy for transferring the implementation towards other learning platforms. The entire implementation was iteratively developed following a rapid application prototyping approach to deliver working prototypes in a short period of time. The paradigm "separation of concern" was followed and each component of the infrastructure was independently developed and communicated with the other components via pre-defined RESTful APIs to ensure the logical and modular independence of the different analytics infrastructure components, thus making it easier to deploy and maintain the entire infrastructure. The raw data and the analytics engine were designed and optimized for system-wide adoption and worked on anonymous data to implement the identified indicators for the goals of the different personas presented in chapter 5. The implementation of the analytics engine and the user interface focused on providing learning analytic results to the normal users while considering their acknowledged necessities from the requirements engineering process. The user interface development resulted in iteratively developed prototypes which were continuously evaluated and improved. Overall 71 learning analytics indicators were designed and implemented through these prototypes. As part of the implementation work, a technical analysis showed that around half of the collected indicators can be implemented with anonymous data. The most completed prototype which went through many smaller iterations, and two major ones, was the "Insights" learning analytics prototype for the teaching staff and it was tested and evaluated in a real deployment scenario, to assess its effectiveness in the real-world application in blended learning scenarios.

Chapter seven provided the conducted evaluation and the outcomes of the two longitudinal case studies conducted with the teaching staff stakeholder group. The studies were conducted over two semesters with different courses and course types from different faculties. The results from both studies found that the learning platform is the place to provide learning analytics tools because they incorporated it in their day-to-day teaching activities. Additionally, the "Insights" module helped in increasing the teachers' awareness of the students' behavior and they were able to detect trends in online activities, reflect upon them and correlate them with their previous knowledge and experiences and devise some activities based on them. The results also showed that analytics results were meaningful to the users that and the users felt that there is an actual benefit for them, the tool had an impact on reflection and action. Overall, the conducted evaluation with the two case studies provided comprehensive and versatile descriptions of the interaction dynamics of a learning analytics module in blended learning scenarios, enabled awareness, reflection and inspired some forms of action, and proved to be widely suitable for deploying as a learning analytics tool at scale in blended learning scenarios in higher education.

Chapter eight provided summarized answers of the sub-research questions defined in chapter 1. The answers are organized as four bundles which summarized the generated knowledge and scientific discoveries achieved through the work of this dissertation. The summarized answers and outcomes focus on the following four topics:

- Requirements engineering and elicitation methods and guidelines including the requirements themselves;

- Privacy guidelines and approaches how to handle these issues in practical scenarios; provide strategies about provision and acceptance strategies for introducing learning analytics as an integral e-learning service for the stakeholders in blended learning scenarios in higher education, technical specifications and recommendations concerning learning data management, analysis algorithms, deployment and provision strategies for learning analytics and the technical implementation of the hardware and software infrastructure;
- Provide experiences, knowledge, and suggestions about practical implementations of the learning analytics components from the infrastructure;
- Provide experiences, knowledge, and suggestions about research approaches and methods for conductive formative and summative evaluation of the learning analytics prototypes and implementation.

The outcomes of this research can be taken and applied to similar projects in this research and problem domain, and as such can help future projects in the design, implementation, and provision of learning analytics services.

As a concluding observation, the main contribution of this dissertation is an evaluated practical process with concrete steps for scaling up learning analytics in blended learning scenarios in higher education.

Future work and challenges

The immediate future work for this dissertation is to bring all of the stakeholders on the same level as the teaching staff concerning the provision of learning analytics tools and services. This can be achieved by taking the comprehensive requirements developed as part of this dissertation and updating and improving the implemented prototypes for the rest of the stakeholder groups and conducting a longitudinal evaluation to validate and corroborate their added value for their intended users. However, the provision of learning analytics services to all stakeholders should not stop there. Learning analytics services is a long-term commitment that should enhance and improve the learning and teaching experiences and even when developed and provided in stages there are challenges to be resolved.

The first challenge is connecting the analytics results with educational theories and the applied learning scenarios. In theory, learning analytics should be well connected and completely aligned with the educational theories and sciences. According to Ferguson (2012a) to optimize and enhance learning and the learning experiences, one needs to first understand how knowledge develops over time and identify different ways and methods to support knowledge development. Researchers must find ways to connect cognition, meta-cognition, and pedagogy to improve the learning processes (Pea, 2014). The first two steps from section 8.1 are fairly straightforward and their value is in the reporting about the learners' behavior, usage patterns, and the influence of the teaching activities on the online learning platforms. However, it is still not clear how the particular learning analytics results from the following steps would be connected to the learning theories. Namely, the correlation of assessment results and just with a fraction of the complete student behavior (not just the online presence), or even making predictions and interventions based on predictions based on a fraction of the data about the student behavior can be dangerous. Additionally, how long-term analytics influences the course design and e-learning scenarios is still an open question, because so far, the courses, their assessment strategies, and resources are designed and implemented on a learning platform without the benefits and knowledge brought by learning analytics.

The second challenge is connected with the data privacy aspects and issues that will arise when more complex and composite learning analytics indicators which use a lot more personal data in comparison with pseudonymized usage traces. The legal framework was developed and provided the legal basis on which e-learning services which include learning analytics were provided to the stakeholders, but it does not solve all of the practical problems concerning data privacy and the ethics for using massive amounts of personal data. For example, it is still not possible to collect all kinds of personal data without clear evidence and added value of the composite indicators that use performance data in relation to other personal data because it breaks the basic approach of data minimalism when providing a service. Moreover, if there already is collected personal data from the students and the teaching staff, this data cannot be simply repurposed or reused to provide other services or re-analyze it just because it is already available. Hence, the development and advancement of e-learning services which heavily rely on personal data have to be tightly regulated and controlled and practical mechanisms have to be developed which take care of consent, data accuracy, interpretation, ownership, and data preservation. In combination with the first challenge, the institution must have empirical evidence that analytics results which are based on such personal data have a clear advantage and benefits to the educational processes.

The third challenge is data analysis and integrity in connection with the data privacy aspects and issues. Data collection and analysis are a challenge for learning analytics, although in the research community there is a mutual understanding that "infinite amounts" of data is readily available. As mentioned before, not all available data can be or should be collected. Furthermore, there are still unsolved practical questions and problems of how to aggregate and integrate raw data from different data sources, which come in various forms and sizes, and out of them produce a usable educational dataset for providing analytics (M. Chatti & Dyckhoff, 2012; Ferguson, 2012a). Another aspect is the data availability and validity in correlation with the outlined analytics goals. There is a real possibility of assimilating and organizing the data in an erroneous presentation and usage which can negatively influence the analytics results and cause many discrepancies and misinterpretations. Moreover, there with the emergence of newly available technologies and devices, there will be a need to shift the focus on data that may include mobile, biometric, and mood data. In parallel to the individual learning aspect of learning analytics, in blended learning scenarios there focus on collaboration, cooperation, and interactions of the learners within and outside the learning environment, which further complicates the problem of collecting data, analyzing it and validating its integrity (Avella et al., 2016).

The last challenge is how to evaluate the developed learning analytics tools and service to provide empirical evidence about the usability, usefulness, and objective impact on the stakeholders and the learning processes and experiences. Usability and user experience are relatively easy to evaluate and measure with methodologies from the HCI field, the challenge is to investigate how and in which ways learning analytics tools could impact the learning processes and how this impact or effect could be measured and evaluated. As the evaluation in this dissertation showed, the measuring the usefulness, utility and ultimately the effect of the learning analytics tools is a challenging task and the process needs long periods of time (over several iterations) and a lot of effort and active participation from researchers, developers, and participants. The good news is that there are several initiatives in the research community that try to solve this challenge and that they have received promising results. Overall, further research is necessary to investigate effective mixed-method evaluation approaches that focus on measuring learning analytics tools' impact on the learning processes.

In conclusion, learning analytics has a future and it is here to stay and will change education. The question that remains unanswered is: How? I personally think that it will be for the better.

11 BIBLIOGRAPHY

Abela, A. (2014). Advanced Presentations by Design. *Creating Communication That Drives Action*.

Adlin, T., & Pruitt, J. (2010). *The Essential Persona Lifecycle: Your Guide to Building and Using Personas*. *The Essential Persona Lifecycle: Your Guide to Building and Using Personas*. http://doi.org/10.1016/C2009-0-62475-2

Agresti, A. (2002). *Categorical Data Analysis: Second edition*. *John Wiley & Sons, INC*. http://doi.org/10.1037/023990

Aguilar, S., Lonn, S., & Teasley, S. D. (2014). Perceptions and use of an early warning system during a higher education transition program. *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 113–117. http://doi.org/10.1145/2567574.2567625

Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers and Education*. http://doi.org/10.1016/j.compedu.2011.08.030

Anderson, T., & Dron, J. (2011). Three generations of distance education pedagogy. *International Review of Research in Open and Distance Learning*. http://doi.org/10.19173/irrodl.v12i3.890

Arnold, K. E., Hall, Y., Street, S. G., Lafayette, W., Pistilli, M. D., Hall, Y., … Lafayette, W. (2012). Course Signals at Purdue: Using Learning Analytics to Increase Student Success. *Lak '12*, (May), 2–5. http://doi.org/10.1145/2330601.2330666

Arnold, K. E., Lynch, G., Huston, D., Wong, L., Jorn, L., & Olsen, C. W. (2014). Building institutional capacities and competencies for systemic learning analytics initiatives. *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 257–260. http://doi.org/10.1145/2567574.2567593

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*. http://doi.org/10.1145/2330601.2330666

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Journal of Asynchronous Learning Network*. http://doi.org/10.24059/olj.v20i2.790

Baker, B. M. (2007). *A conceptual framework for making knowledge actionable through capital formation*. *ProQuest Dissertations and Theses*.

Baker, R. S. J. D. R., & Yacef, K. (2009). The State of Educational Data Mining in 2009 : A Review and Future Visions. *Journal of Educational Data Mining*. http://doi.org/http://doi.ieeecomputersociety.org/10.1109/ASE.2003.1240314

Bakharia, A., Corrin, L., Barba, P. De, Kennedy, G., Gasevic, D., Mulder, R., … Lockyer, L. (2016). A Conceptual Framework linking Learning Design with Learning Analytics. In

Proceedings of the 6th International Conference on Learning Analytics & Knowledge. http://doi.org/10.1145/2883851.2883944

Bakharia, A., & Dawson, S. (2011). SNAPP: A Bird's-Eye View of Temporal Participant Interaction AneeshaBakharia. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, 168. http://doi.org/10.1145/2090116.2090144

Barab, S., & Squire, K. (2004). Design-Based Research: Putting a Stake in the Ground. *Journal of the Learning Sciences*. http://doi.org/10.1207/s15327809jls1301_1

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, *2*(2), 131–160. http://doi.org/10.1037/1082-989X.2.2.131

Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*. http://doi.org/10.1145/3027385.3027403

Bogarín, A., Romero, C., Cerezo, R., & Sánchez-Santillán, M. (2014). Clustering for improving educational process mining. *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 11–15. http://doi.org/10.1145/2567574.2567604

Boroujeni, M. S., & Dillenbourg, P. (2018). Discovery and temporal analysis of latent study patterns in MOOC interaction sequences. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*. http://doi.org/10.1145/3170358.3170388

Boroujeni, M. S., Hecking, T., Hoppe, H. U., & Dillenbourg, P. (2017). Dynamics of MOOC discussion forums. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*. http://doi.org/10.1145/3027385.3027391

Boroujeni, M. S., Sharma, K., Kidziński, Ł., Lucignano, L., & Dillenbourg, P. (2016). How to quantify student's regularity? In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-45153-4_21

Bos, N., & Brand-Gruwel, S. (2016). Student differences in regulation strategies and their use of learning resources. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. http://doi.org/10.1145/2883851.2883890

Bote-Lorenzo, M. L., & Gómez-Sánchez, E. (2017). Predicting the decrease of engagement indicators in a MOOC. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*. http://doi.org/10.1145/3027385.3027387

Bratitsis, T., & Dimitracopoulou, A. (2008). Interaction Analysis as a multi-support approach of social computing for learning, in the "collaborative Era": Lessons learned by using the DIAS system. In *Proceedings - The 8th IEEE International Conference on Advanced Learning Technologies, ICALT 2008*. http://doi.org/10.1109/ICALT.2008.260

Brooks, C., Panesar, R., & Greer, J. (2006). Awareness and Collaboration in the iHelp Courses Content Management System. In *Proc. of the 1st Europ. Conf. on Technology Enhanced Learning (EC-TEL 2006)*. http://doi.org/10.1007/11876663_5

Broos, T., Verbert, K., Langie, G., Van Soom, C., & De Laet, T. (2018). Multi-institutional positioning test feedback dashboard for aspiring students. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*. http://doi.org/10.1145/3170358.3170419

Brouwer, N., Bredeweg, B., Latour, S., Berg, A., & van der Huizen, G. (2016). Learning analytics pilot with coach2 - Searching for effective mirroring. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-45153-4_28

Brown, M. (2012). Learning Analytics: Moving from Concept to Practice. *EDUCAUSE Learning Initiative*. http://doi.org/10.1177/0022146513479002

Bundesdatenschutzgesetz. (1990). Bundesdatenschutzgesetz. Retrieved from https://www.gesetze-im-internet.de/bdsg_1990/

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic Analytics. *Educause Review*, *42*(October), 40–57.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Information visualization: using vision to think. In *Readings in information visualization: using vision to think*.

Carpendale, S. (2008). Evaluating information visualizations. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-540-70956-5_2

Charleer, S., Klerkx, J., & Duval, E. (2014). Learning Dashboards. *Journal of Learning Analytics*. http://doi.org/10.18608/jla.2014.13.22

Charleer, S., Klerkx, J., Duval, E., De Laet, T., & Verbert, K. (2016). *Creating Effective Learning Analytics Dashboards: Lessons Learnt*. *Adaptive and Adaptable Learning: 11th European Conference on Technology Enhanced Learning, EC-TEL 2016, Lyon, France, September 13-16, 2016, Proceedings*. http://doi.org/10.1007/978-3-319-45153-4_4

Charleer, S., Santos, J. L., Klerkx, J., & Duval, E. (2014). LARAe: Learning analytics reflection & awareness environment. In *CEUR Workshop Proceedings*.

Chatfield, C. (1985). The Initial Examination of Data. *Journal of the Royal Statistical Society. Series A (General)*, *148*(3), 214. http://doi.org/10.2307/2981969

Chatfield, C. (1986). Exploratory data analysis. *European Journal of Operational Research*. http://doi.org/10.1016/0377-2217(86)90209-2

Chatti, A. M., Lukarov, V., Thüs, H., Muslim, A., Yousef, F. A. M., Wahid, U., … Schroeder, U. (2014). Learning Analytics: Challenges and Future Research Directions. *Eleed*, *10*(1). Retrieved from http://nbn-resolving.de/urn:nbn:de:0009-5-40350

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5/6), 318. http://doi.org/10.1504/IJTEL.2012.051815

Chatti, M., & Dyckhoff, A. (2012). A reference model for learning analytics. *… Enhanced Learning*. http://doi.org/10.1504/IJTEL.2012.051815

Chen, Chiang, & Storey. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*. http://doi.org/10.2307/41703503

Christensen, C. M. (2010). Integrating Around the Job to Be Done. *Harvard Business School Module Note 611004*, 16.

Christensen, C. M., Hall, T., Dillon, K., & Duncan, D. S. (2016). *Competing Against Luck: The Story of Innovation and Customer Choice*. HarperBusiness.

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., & Lindstädt, S. (2018). Finding traces of self-regulated learning in activity streams. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* -

LAK '18. http://doi.org/10.1145/3170358.3170381

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6). http://doi.org/10.1080/13562517.2013.827653

Clow, D. (2014). Data Wranglers: Human interpreters to help close the feedback loop. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 49–53. http://doi.org/10.1145/2567574.2567603

Clow, D., & Hall, W. (2012). The Learning Analytics Cycle: Closing the loop effectively. In *LAK12*.

Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 83–92. http://doi.org/10.1145/2567574.2567586

Conole, G. (2016). MOOCs as disruptive technologies: strategies for enhancing the learner experience and quality of MOOCs. *Revista de Educación a Distancia (RED)*. http://doi.org/10.6018/red/50/2

Consortium, N. M., & Initiative, E. L. (2013). NMC Horizon Report 2013 Educause. *Www.Nmc.Org*.

Cooper, A. (2012). What is Analytics? Definition and Essential Characteristics. *CETIS Analytics Series*.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*. http://doi.org/10.1007/BF00988593

Corrigan, O., Smeaton, A. F., Glynn, M., & Smyth, S. (2015). Using educational analytics to improve test performance. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-24258-3_4

Corrin, L., & de Barba, P. (2015). How do students interpret feedback delivered via dashboards? *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 430–431. http://doi.org/10.1145/2723576.2723662

Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. http://doi.org/10.1145/2883851.2883931

Daniel, J. (2012). Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility. *Journal of Interactive Media in Education*. http://doi.org/10.5334/2012-18

Davis, D., Chen, G., van der Zee, T., Hauff, C., & Houben, G. J. (2016). Retrieval practice and study planning in MOOCs: Exploring classroom-based self-regulated learning strategies at scale. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-45153-4_5

Davis, D., & Hauff, C. (2018). The Half-Life of MOOC Knowledge A Randomized Trial Evaluating Knowledge Retention and Retrieval Practice in MOOCs. *Proceedings of the 8th International Conference on Learning Analytics & Knowledge - LAK '18*. http://doi.org/10.1145/3170358.3170383

Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., & Houben, G.-J. (2017). Follow the Successful Crowd: Raising MOOC Completion Rates through Social Comparison at Scale

*. In *The Seventh International Learning Analytics & Knowledge Conference*. http://doi.org/10.1145/3027385.3027411

Dawson, S., Bakharia, A., & Heathcote, E. (2010). SNAPP : Realising the affordances of real-time SNA within networked learning environments. *Proceedings of the 7th International Conference on Networked Learning*, 125–133. Retrieved from http://www.lancs.ac.uk/fss/organisations/netlc/past/nlc2010/abstracts/PDFs/Dawson.pdf

De Groot, R., Drachman, R., Hever, R. R., Schwarz, B. B., Hoppe, U., Harrer, A., … Baurens, B. (2007). Computer Supported Moderation of E-Discussions: the ARGUNAUT Approach. In *Proc. of the Conf. on Computer Supported Collaborative Learning 2007 (CSCL'07)*.

Dimitracopoulou, A. (2008). Computer based Interaction Analysis Supporting Self-regulation: Achievements and Prospects of an Emerging Research Direction. *Technology Instruction Cognition and Learning*.

Dix, Alan and Finlay, Janet E. and Abowd, Gregory D. and Beale, R. (2003). *Human-Computer Interaction (3rd Edition). Human-Computer Interaction*. http://doi.org/10.1207/S15327051HCI16234

Dix, A., Finlay, J., Abowd, G. D., & Russel, B. (2004). *Human-Computer Interaction*. Pearson.

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work - CSCW '92*. http://doi.org/10.1090/S0002-9904-1966-11639-7

Drachsler, H., & Greller, W. (2016). Privacy and Analytics – it's a DELICATE Issue A Checklist for Trusted Learning Analytics. In *6th Conference on Learning Analytics and Knowledge*. http://doi.org/10.1145/2883851.2883893

Drachsler, H., Greller, W., Greller Wolfgang, & Drachsler Hendrik. (2012). The Pulse of Learning Analytics - Understandings and Expectations from the Stakeholders. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, *15*(3), 42–57. http://doi.org/10.1145/2330601.2330634

Dunbar, R. L., Dingel, M. J., & Prat-Resina, X. (2014). Connecting Analytics and Curriculum Design: Process and Outcomes of Building a Tool to Browse Data Relevant to Course Designers. *Journal of Learning Analytics*. http://doi.org/10.18608/jla.2014.13.26

Duval, E. (2011a). Attention please!: Learning analytics for visualization and recommendation. *LAK '11 Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, 9–17. http://doi.org/10.1145/2090116.2090118

Duval, E. (2011b). Attention Please! Learning Analytics for Visualization and Recommendation Erik. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*. http://doi.org/10.1145/2090116.2090118

Dyckhoff, A. L. (2011). Implications for Learning Analytics Tools: A Meta-Analysis of Applied Research Questions. *International Journal of Computer Information Systems and Industrial Management Applications*.

Dyckhoff, A. L. (2014). Action Research and Learning Analytics in Higher Education.

Dyckhoff, A. L., Lukarov, V., Muslim, A., Chatti, M. a., & Schroeder, U. (2013). Supporting action research with learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*, 220. http://doi.org/10.1145/2460296.2460340

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and

Implementation of a Learning Analytics Toolkit for Teachers Design and Implementation of a Learning Analytics Toolkit for Teachers. *Source: Journal of Educational Technology & Society Educational Technology & Society*, *15*(153), 58–76. Retrieved from http://www.jstor.org/stable/jeductechsoci.15.3.58%5Cnhttp://www.jstor.org/stable/jeductechsoci.15.3.%5Cnhttp://about.jstor.org/terms

Dziuban, C., Moskal, P., & Hartman, J. (2005). HIGHER EDUCATION , BLENDED LEARNING AND THE GENERATIONS : KNOWLEDGE IS POWER - NO MORE. *Research Initiative for Teaching Effectiveness,*. http://doi.org/10.1080/09687761003657614

Elias, T. (2011). Learning Analytics : Definitions , Processes and Potential. *Online*. http://doi.org/10.1.1.456.7092

Ellis, G., & Dix, A. (2006). An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for information visualization - BELIV '06*. http://doi.org/10.1145/1168149.1168152

European Union. (2016). Regulation 2016/679 of the European parliament and the Council of the European Union. *Official Journal of the European Communities*, 1–88. http://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf

Fekete, J. D., Van Wijk, J. J., Stasko, J. T., & North, C. (2008). The value of information visualization. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-540-70956-5_1

Ferguson, R. (2012a). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304. http://doi.org/10.1504/IJTEL.2012.051816

Ferguson, R. (2012b). The State of Learning Analytics in 2012 : A Review and Future Challenges a review and future challenges. *Media*. http://doi.org/10.1504/IJTEL.2012.051816

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., … Vuorikari, R. (2016). *Research Evidence on the Use of Learning Analytics*. http://doi.org/10.2791/955210

Ferguson, R., Clow, D., Macfadyen, L., Essa, A., Dawson, S., & Alexander, S. (2014). Setting learning analytics in context. *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 251–253. http://doi.org/10.1145/2567574.2567592

Ferguson, R., Macfadyen, L. P., Clow, D., Tynan, B., Alexander, S., & Dawson, S. (2014). Setting Learning Analytics in Context: Overcoming the Barriers to Large-Scale Adoption. *Journal of Learning Analytics*. http://doi.org/10.18608/jla.2014.13.7

Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. *Distribution*. http://doi.org/10.1016/S0262-4079(06)60844-0

Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., & Berners-Lee, T. (1999). RFC2616 - Hypertext transfer protocol–HTTP/1.1. *Internet Engineering Task Force*, 1–114. http://doi.org/http://www.ietf.org/rfc/rfc2616.txt

Figl, K. (2009). ISONORM 9241/10 und Isometrics: Usability-Fragebögen im Vergleich. *Tagungsband Mensch & Computer*, 143–152. Retrieved from http://nm.wuwien.ac.at/research/publications/b796.pdf

Figueira, A. (2015). Predicting Results from Interaction Patterns During Online Group Work BT - Design for Teaching and Learning in a Networked World. In G. Conole, T. Klobučar, C. Rensing, J. Konert, & E. Lavoué (Eds.), (pp. 414–419). Cham: Springer International Publishing.

Friesen, N. (2012). Report : Defining Blended Learning. *Norm Friesen, August*.

Fritz, J. (2011). Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers. *Internet and Higher Education*. http://doi.org/10.1016/j.iheduc.2010.07.007

García-Saiz, D., & Zorrilla, M. (2011). E-learning Web Miner: A data mining application to help instructors involved in virtual courses. *Proceedings of the 4th International Conference on EDM*.

García-solórzano, D., Poblenou, R., Morán, J. A., Monzo, C., & Melenchón, J. (2012). Educational Monitoring Tool Based on Faceted Browsing and Data Portraits Categories and Subject Descriptors. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 170–178. http://doi.org/10.1145/2330601.2330645

Garrison, D. R., & Vaughan, N. D. (2012). *Blended Learning in Higher Education: Framework, Principles, and Guidelines*. *Blended Learning in Higher Education: Framework, Principles, and Guidelines*. http://doi.org/10.1002/9781118269558

Gaševi, D., Dawson, S., & Siemens, G. (2015). Let ’ s not forget : Learning Analytics are about Learning Course Signals : Lessons Learned. *TechTrends*, *59*(1), 64–71.

Glahn, C., Gruber, M. R., & Tartakovski, O. (2015). Beyond delivery modes and apps: A case study on mobile blended learning in higher education. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-24258-3_10

González Agulla, E., Alba Castro, J. L., Argones Rúa, E., & Anido Rifón, L. (2009). Realistic Measurement of Student Attendance in LMS Using Biometrics. In *Proc. of the Int. Symposium on Engineering Education and Educational Technologies (EEET’09)*.

Gospodinova, S. (2018). *Developing learning analytics prototypes for supporting e-learning scenarious in higher education*.

Govaerts, S., & Duval, E. (2012). The Student Activity Meter for Awareness and Self-reflection. *CHI’12 Extended Abstracts on Human Factors in Computing Systems*, 869–884.

Govaerts, S., Verbert, K., & Duval, E. (2011). Evaluating the student activity meter: Two case studies. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7048 LNCS*, 188–197. http://doi.org/10.1007/978-3-642-25813-8_20

Hecking, T., Ziebarth, S., & Hoppe, H. U. (2014). Analysis of dynamic resource access patterns in a blended learning course. *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK ’14*, *1*, 173–182. http://doi.org/10.1145/2567574.2567584

Hevner, A. R., & Chatterjee, S. (2010a). *Design Research in Information Systems: Theory and Practice*. *Springer*. http://doi.org/10.1007/978-1-4419-6108-2

Hevner, A. R., & Chatterjee, S. (2010b). Design Science Research in Information Systems. *Design Research in Information Systems*. http://doi.org/10.1007/978-1-4419-5653-8

Hillaire, G., Rappolt-Schlichtmann, G., & Ducharme, K. (2016). Prototyping Visual Learning Analytics Guided by an Educational Theory Informed Goal. *Journal of Learning Analytics*. http://doi.org/10.18608/jla.2016.33.7

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. In *the Seventh International Learning Analytics & Knowledge Conference*. http://doi.org/10.1145/3027385.3027449

Holman, C., Aguilar, S. J., Levick, A., Stern, J., Plummer, B., & Fishman, B. (2015). Planning for success: how students use a grade prediction tool to win their classes. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 260–264. http://doi.org/10.1145/2723576.2723632

Holman, C., Fishman, B., & Aguilar, S. (2013). GradeCraft : What Can We Learn From a Game-Inspired Learning Management System ? *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 260–264. http://doi.org/10.1145/2460296.2460350

Huberth, M., Chen, P., Tritz, J., & McKay, T. A. (2015). Computer-tailored student support in introductory physics. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0137001

IBM2011. (2011). *The 2011 IBM Tech Trends Report: The Clouds are Rolling In...Is Your Business Ready?*

Iliinsky, N., & Steele, J. (2011). *Designing Data Visualizations*. *O'Reilly Media, Inc.,*. http://doi.org/10.1073/pnas.0805135105

ISO. (2010). ISO 9241 Part 210: Human-centred design for interactive systems. In *ISO 9241 Ergonomics of humansystem interaction*. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075

Jadin, T., & Gaisch, M. (2014). Extending the MOOCversity A Multi-layered and Diversified Lens for MOOC Research. *Proceedings of the European MOOC Stakeholder Summit 2014*. http://doi.org/10.13140/RG.2.1.5105.8404

Jakob Nielsen. (1995). 10 Heuristics for User Interface Design: Article by Jakob Nielsen. http://doi.org/10.1016/B978-0-12-385241-0.00003-8

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers and Education*. http://doi.org/10.1016/j.compedu.2006.01.004

Jo, I.-H., Kim, D., & Yoon, M. (2014). Analyzing the log patterns of adult learners in LMS using learning analytics. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 183–187. http://doi.org/10.1145/2567574.2567616

Johnson, L., & Adams, S. (2011). *The Horizon Report 2011*. *Methodology*.

Jovanović, J., Gašević, D., Brooks, C., Devedžić, V., & Hatala, M. (2007). LOCO - Analyst : A Tool for Raising Teachers ' Awareness in Online Learning Environments. *LNCS*. http://doi.org/10.1007/978-3-540-75195-3_9

Khan, I., & Pardo, A. (2016). Data2U : Scalable real time student feedback in active learning environments. In *LAK '16 6th International Conference on Learning Analytics and Knowledge*. http://doi.org/10.1145/2883851.2883911

Kim, J., Jo, I. H., & Park, Y. (2016). Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review*. http://doi.org/10.1007/s12564-015-9403-8

Klerkx, J., Verbert, K., & Duval, E. (2017). Learning Analytics Dashboards. In *Handbook of*

Learning Analytics (pp. 143–150). http://doi.org/10.18608/hla17.012

Knight, D. B., Brozina, C., & Novoselich, B. (2016). An investigation of first-year engineering student and instructor perspective of learning analytics approaches. *Journal of Learning Analytics*. http://doi.org/10.18608/jla.2016.33.11

Kosba, E., Dimitrova, V., & Boyle, R. (2005). Using student and group models to support teachers in web-based distance education. *Proceedings of the 10th International Conference on User Modeling*, (Section 2), 124–133. http://doi.org/10.1007/11527886_17

Koulocheri, E., & Xenos, M. (2013). Considering formal assessment in learning analytics within a PLE: The HOU2LEARN case. In *ACM International Conference Proceeding Series*. http://doi.org/10.1145/2460296.2460304

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 184–193. http://doi.org/10.1145/2723576.2723623

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Dawson, S., & Gašević, D. (2017). Understanding the relationship between technology use and cognitive presence in MOOCs. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*. http://doi.org/10.1145/3027385.3029471

Krumm, A. E., Waddington, R. J., Teasley, S. D., & Lonn, S. (2014). A learning management system-based early warning system for academic advising in undergraduate engineering. In *Learning Analytics: From Research to Practice*. http://doi.org/10.1007/978-1-4614-3305-7_6

Kurzweil, M., & Stevens, M. (2015). *Building a Pathway to Student Success at Georgia State University*.

Lauría, E. J. M., Baron, J. D., Devireddy, M., Sundararaju, V., & Jayaprakash, S. M. (2012). Mining academic data to improve college student retention : An open source perspective. *Proceedings of the Second International Conference on Learning Analytics And Knowledge - LAK '12*, (May), 139–142. http://doi.org/10.1145/2330601.2330637

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). Research Methods in Human-Computer Interaction. *Evaluation*, 426. http://doi.org/10.1002/asi.21187

Lazar, J., Feng, J. H., & Hochheiser, H. (2017a). Chapter 1 – Introduction to HCI research. In *Research Methods in Human Computer Interaction*. http://doi.org/10.1016/B978-0-12-805390-4.00001-7

Lazar, J., Feng, J. H., & Hochheiser, H. (2017b). Chapter 10 – Usability testing. In *Research Methods in Human Computer Interaction*. http://doi.org/https://doi.org/10.1016/B978-0-12-805390-4.00010-8

Lazar, J., Feng, J., & Hochheiser, H. (2017c). Chapter 7 - Case Studies. In *Research Methods in Human Computer Interaction*. http://doi.org/10.1016/B978-0-12-805390-4.00006-6

Lazar, J., Heidi Feng, J., & Hochheiser, H. (2017). Chapter 5 - Surveys. In *Research Methods in Human Computer Interaction (Second Edition)*. http://doi.org/https://doi.org/10.1016/B978-0-12-805390-4.00005-4

Li, N., Kidziński, L., Jermann, P., & Dillenbourg, P. (2015). MOOC video interaction patterns: What do they tell us? In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-24258-3_15

Lockyer, L., & Dawson, S. (2011). Learning designs and learning analytics. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, 153. http://doi.org/10.1145/2090116.2090140

Long, P., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *Educause Review*, *46*(5), 31–40. Retrieved from http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume46/PenetratingtheFogAnalyticsinLe/235017

Lukarov, V. (2013). *Learning Analytics Tools for raising Awareness and Reflection*.

Lukarov, V., Chatti, M. A., & Schroeder, U. (2015). Learning analytics evaluation -beyond usability. In *CEUR Workshop Proceedings*.

Lukarov, V., Chatti, M. A., Thüs, H., Kia, F. S., Muslim, A., Greven, C., & Schroeder, U. (2014). Data Models in Learning Analytics., 88–95.

Lukarov, V., & Schroeder, U. (2017a). AiX-Analytics: Analytics Tool at {RWTH} Aachen University. In C. Igel, C. Ullrich, & W. Martin (Eds.), *Bildungsräume 2017: DeLFI 2017, Die 15. e-Learning Fachtagung Informatik, der Gesellschaft für Informatik e.V. (GI), 5. bis 8. September 2017, Chemnitz* (Vol. P-273, pp. 393–394). Gesellschaft für Informatik, Bonn. Retrieved from https://dl.gi.de/20.500.12116/4873

Lukarov, V., & Schroeder, U. (2017b). Analytics Prototype for Data Driven Decision Making for Blended Learning Strategies in HEI. *Bildungsräume 2017*.

Lukarov, V., & Schroeder, U. (2018). Learning analytics case study in blended learning scenarios. In *CEUR Workshop Proceedings*.

MacNeill, S., Campbell, L. M., & Hawksey, M. (2014). Analytics for education. In *Reusing Open Resources: Learning in Open Networks for Work, Life and Education*. http://doi.org/10.4324/9780203780190

MacNeill, S., & Powell, S. (2012). Institutional Readiness for Analytics. *CETIS Analytics Series*. http://doi.org/10.1109/5.726784

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. (2015). The LATUX workflow: designing and deploying awareness tools in technology-enabled learning settings. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 1–10. http://doi.org/10.1145/2723576.2723583

Martinez-Maldonaldo, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. (2015). LATUX: An iterative workflow for designing, validating, and deploying learning analytics visualizations. *Journal of Learning Analytics*. http://doi.org/10.18608/jla.2015.23.3

Mattingly, K. D., Rice, M. C., & Berge, Z. L. (2012). Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management and E-Learning*. http://doi.org/10.1145/2460296.2460308

May, M., George, S., & Prévôt, P. (2011). TrAVis to Enhance Students' Self-monitoring in Online Learning Supported by Computer-Mediated Communication Tools. *Ijcisim*.

Mazza, R., & Dimitrova, V. (2004). Visualising student tracking data to support instructors in web-based distance education. *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters - WWW Alt. '04*, 154. http://doi.org/10.1145/1013367.1013393

Mazza, R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-*

Computer Studies, *65*(2), 125–139. http://doi.org/10.1016/j.ijhcs.2006.08.008

Mazza, R., & Milani, C. (2004). GISMO: a Graphical Interactive Student Monitoring Tool for Course Management Systems. *Technology Enhanced Learning '04 International Conference*. http://doi.org/- http://dx.doi.org/10.1016/j.ejso.2013.02.002

McKay, T., Miller, K., & Tritz, J. (2012). What to do with actionable intelligence:E2Coach as an intervention engine. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 88–91. http://doi.org/10.1145/2330601.2330627

Mentiu, M. (2018). *Evaluating Learning Analytics Prototypes for E-Learning*.

Merceron, A., & Yacef, K. (2005). TADA-Ed for Educational Data Mining. *IME of CEL*.

Millecamp, M., Gutiérrez, F., Charleer, S., Verbert, K., & De Laet, T. (2018). A qualitative evaluation of a learning dashboard to support advisor-student dialogues. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*. http://doi.org/10.1145/3170358.3170417

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*. http://doi.org/10.1037/h0043158

Miller, G. A., & Miller GA. (1994). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*. http://doi.org/10.1037/h0043158

Mochizuki, T., Kato, H., Hisamatsu, S.-I., Yaegashi, K., Fujitani, S., Nagata, T., … Suzuki, M. (2005). Promotion of self-assessment for learners in online discussion using the visualization software. *Computer Supported Collaborative Learning 2005: The Next 10 Years - Proceedings of the International Conference on Computer Supported Collaborative Learning 2005, CSCL 2005*. http://doi.org/10.1111/j.1467-8535.2008.00890_11.x

Muñoz-Merino, P. J., Valiente, J. a R., & Kloos, C. D. (2013). Inferring higher level learning information from low level data for the Khan Academy platform. *LAK '13: Proceedings of the Third International Conference on Learning Analytics and Knowledge,* 112. http://doi.org/10.1145/2460296.2460318

Myatt, G. J. (2007). *Making Sense of Data A Practical Guide to Exploratory Data Analysis and Data Mining*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Nacu, D., Baltes, J., Hamid, T., Gemmell, J., & Pinkard, N. (2018). A framework for developing metrics of youth engagement in informal learning environments. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*. http://doi.org/10.1145/3170358.3170393

Nielsen, J. (1993). *Usability Engineering*. *Morgan Kaufmann Publishers Inc*. http://doi.org/10.1145/1508044.1508050

Nielsen, J. (1994). Usability inspection methods. In *Conference companion on Human factors in computing systems - CHI '94*. http://doi.org/10.1145/259963.260531

Nielsen, J. (1995). 10 Usability Heuristics for User Interface Design. http://doi.org/10.1145/191666.191729

Nielsen, J. (2006). The 90-9-1 rule for participation inequality in social media and online communities.

Niemann, K., Scheffel, M., & Wolpers, M. (2012). An overview of usage data formats for recommendations in TEL. In *CEUR Workshop Proceedings*.

http://doi.org/10.1253/circj.CJ-10-0301

NMC Horizon Report. (2016). *Horizon Report - 2016 Higher Education Edition*. *NMC Horizon Report*. http://doi.org/ISBN 978-0-9968527-5-3

Oliver, M., & Trigwell, K. (2005). Can “Blended Learning” Be Redeemed? *E-Learning*. http://doi.org/10.2304/elea.2005.2.1.2

Paechter, M., & Maier, B. (2010). Online or face-to-face? Students’ experiences and preferences in e-learning. *Internet and Higher Education*. http://doi.org/10.1016/j.iheduc.2010.09.004

Papamitsiou, Z., Terzis, V., & Economides, A. A. (2014). Temporal learning analytics for computer based testing. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge - LAK ’14*, 31–35. http://doi.org/10.1145/2567574.2567609

Pardo, A., Han, F., & Ellis, R. A. (2016). Exploring the relation between self-regulation, online activities, and academic performance. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK ’16*. http://doi.org/10.1145/2883851.2883883

Park, J., Denaro, K., Rodriguez, F., Smyth, P., & Warschauer, M. (2017). Detecting changes in student behavior from clickstream data. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK ’17*. http://doi.org/10.4319/lo.2001.46.2.0230

Pea, R. (2014). A report on building the field of learning analytics for personalized learning at scale. *The Learning Analytics Workgroup*. http://doi.org/10.1016/0016-3287(90)90080-2

Petropoulou, O., Lazakidou, G., Georgiakakis, P., & Retalis, S. (2012). Making adaptations of CSCL scripts by analyzing learners’ online behavior. In *Studies in Computational Intelligence*. http://doi.org/10.1007/978-3-642-28586-8_8

Pijeira Díaz, H. J., Ruiz, J. S., Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., & Delgado Kloos, C. (2015). Using video visualizations in open edX to understand learning interactions of students. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-24258-3_51

Piller, F., Brenk, S., & Nacken, H. (2017). OUTCOME DRIVEN INNOVATION IN THE CONTEXT OF BLENDED LEARNING – A NEED-BASED INNOVATION APPROACH FOR DIGITALIZATION OF HIGHER EDUCATION (pp. 2609–2623). http://doi.org/10.21125/edulearn.2017.1540

Pohl, K., & Rupp, C. (2015a). *Requirements Engineering Fundamentals: A Study Guide for the Certified Professional for Requirements Engineering Exam-Foundation Level – IREB compliant,. Rocky Nook*. http://doi.org/10.1017/CBO9781107415324.004

Pohl, K., & Rupp, C. (2015b). *Requirements Engineering Fundamentals*. *Rocky Nook*. http://doi.org/10.1017/CBO9781107415324.004

Poon, J. (2013). Blended Learning: An Institutional Approach for Enhancing Students’ Learning Experiences. *Journal of Online Learning & Teaching*. http://doi.org/10.1016/j.sbspro.2014.01.992

Preece, J., Sharp, H., & Rogers, Y. (2015). *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, Inc.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*. http://doi.org/10.1186/2047-2501-2-3

Rodríguez-Triana, M. J., Prieto, L. P., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2018). The teacher in the loop: customizing multimodal learning analytics for blended learning. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*. http://doi.org/10.1145/3170358.3170364

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., & Kloos, C. D. (2014). A demonstration of ALAS-KA: A learning analytics tool for the khan academy platform. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-11200-8_55

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Kloos, C. D., Niemann, K., & Scheffel, M. (2014). Do optional activities matter in virtual learning environments? In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-11200-8_25

Santos, J. L., Govaerts, S., Verbert, K., & Duval, E. (2012). Goal-oriented visualizations of activity tracking : a case study with engineering students. *Learning Analytics and Knowledge*, 143–152. http://doi.org/10.1145/2330601.2330639

Santos, J. L., Verbert, K., Govaerts, S., & Duval, E. (2013). Addressing Learner Issues with StepUp!: An Evaluation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*. http://doi.org/10.1145/2460296.2460301

Sanz-Martínez, L., Martínez-Monés, A., Bote-Lorenzo, M. L., Muñoz-Cristóbal, J. A., & Dimitriadis, Y. (2017). Automatic group formation in a MOOC based on students' activity criteria. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-66610-5_14

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality Indicators for Learning Analytics. *Journal of Educational Technology & Society*, *17*(4), 117–132. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=99574662&site=ehost-live

Scheuer, O., & Zinn, C. (2007). How did the e-learning session go? The Student Inspector. *Proc. of the Conf. on Artificial Intelligence in Education (AIED'07)*, 487–494. Retrieved from http://dl.acm.org/citation.cfm?id=1563601.1563678

Schmitz, H. C., Scheffel, M., Friedrich, M., Jahn, M., Niemann, K., & Wolpers, M. (2009). CAMera for PLE. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-642-04636-0_47

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., … Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*. http://doi.org/10.1109/TLT.2016.2599522

Sclater, N. (2015). Gathering requirements for a student app for learning analytics. Retrieved from http://sclater.com/blog/gathering-requirements-for-a-student-app-for-learning-analytics/

Shimada, A., Taniguchi, Y., Okubo, F., Konomi, S., & Ogata, H. (2018). Online change detection for monitoring individual student behavior via clickstream data on E-book system. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*. http://doi.org/10.1145/3170358.3170412

Siemens, G. (2010). What are Learning Analytics?

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*. http://doi.org/10.1177/0002764213498851

Siemens, G., & Baker, R. S. J. (2012). Learning Analytics and Educational Data Mining : Towards Communication and Collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254. http://doi.org/10.1145/2330601.2330661

Siemens, G., Dawson, S., & Lynch, G. (2013). *Improving the Quality and Productivity of the Higher Education Sector*.

Sommerville, I., & Sawyer, P. (1997). *Requirements Engineering: A Good Practice Guide*. Chichester: Wiley & Sons Ltd. Retrieved from https://www.wiley.com/en-us/Requirements+Engineering%3A+A+Good+Practice+Guide-p-9780471974444

Suthers, D., & Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, 64–74. http://doi.org/10.1145/2090116.2090124

Teasley, S. D. (2017). Student Facing Dashboards: One Size Fits All? *Technology, Knowledge and Learning*, *22*(3), 377–384. http://doi.org/10.1007/s10758-017-9314-3

Tempelaar, D., Rienties, B., & Nguyen, Q. (2018). Investigating learning strategies in a dispositional learning analytics context: the case of worked examples. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*. http://doi.org/10.1145/3170358.3170385

The European Parliament, & The European Council. (2016). General Data Protection Regulation. *Official Journal of the European Union*, *2014*(October 1995), 20–30. http://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf

Ulwick, A. W. (2005). *What customers want: Using outcome-driven innovation to create breakthrough products and services*. New York: McGraw-Hill/Irwin.

Ulwick, A. W. (2011). A New Perspective on Strategy. Retrieved from http://roschuni.org/needfinding/Outcome-DrivenInnovation.pdf

Ulwick, A. W. (2014). What is Outcome-Driven Innovation (ODI)? *Strategyn*.

Ulwick, A. W., & Bettencourt, L. A. (2008). Giving Customers a Fair Hearing. *MIT Sloan Management Review*, *49*(3), 62–68.

van Barneveld, A. Van, Arnold, K., & Campbell, J. (2012). Analytics in higher education: Establishing a common language. *Educause ELI Paper 1*. http://doi.org/10.1111/j.1468-2273.2009.00438.x

van Leeuwen, A. (2016). Learning analytics in a flipped university course. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. http://doi.org/10.1145/2883851.2883874

Vassiliou, A., McAleese, M., & Chair. (2014). *Modernization of higher education*. *European Commission*. http://doi.org/10.2766/81897

Vaughan, N. (2007). Perspectives on blended learning in higher education. *International Journal on E-Learning*.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning Analytics Dashboard Applications. *American Behavioral Scientist*, *57*(10), 1500–1509. http://doi.org/10.1177/0002764213479363

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2013). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, *18*(6), 1499–1514. http://doi.org/10.1007/s00779-013-0751-2

Vitiello, M., Walk, S., Chang, V., Hernandez, R., Helic, D., & Guetl, C. (2017). MOOC dropouts: A multi-system classifier. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-66610-5_22

Waisberg, D., & Kaushik, A. (2009). Web Analytics 2.0: empowering customer centricity. *The Original Search Engine Marketing …*.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*. http://doi.org/10.1007/BF02504682

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The cognitive walkthrough method: A practitioner’s guide. *Usability Inspection*. http://doi.org/10.1108/09685220910944731

Whittaker, S., Terveen, L., Hill, W., & Cherny, L. (1998). The dynamics of mass interaction. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work - CSCW ’98*. http://doi.org/10.1145/289444.289500

Wiegers, K., & Beatty, J. (2013). *Software Requirements, 3rd Edition*. *Microsoft Press*. http://doi.org/978-0-7356-7966-5

Yanosky, R., & Arroway, P. (2015). *The Analytics Landscape in Higher Education*. Louisville.

Yeomans, M., & Reich, J. (2017). Planning prompts increase and forecast course completion in massive open online courses. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK ’17*. http://doi.org/10.1145/3027385.3027416

Yousuf, B., & Conlan, O. (2015). VisEN: Motivating learner engagement through explorable visual narratives. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. http://doi.org/10.1007/978-3-319-24258-3_27

Yu, T., & Jo, I. (2014). Educational Technology Approach toward Learning Analytics : Relationship between Student Online Behavior and Learning Performance in Higher Education. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 269–270. http://doi.org/10.1145/2567574.2567594

Zhang, H., & Almeroth, K. (2010). Moodog: Tracking student activity in online course management systems. *Journal of Interactive Learning Research*.

Zorrilla, M. E., & Alvarez, E. (2008). MATEP: Monitoring and Analysis Tool for E-learning Platforms. In *Proceedings - The 8th IEEE International Conference on Advanced Learning Technologies, ICALT 2008*. http://doi.org/10.1109/ICALT.2008.33

APPENDIX

The following appendix contains the results and outcomes of the different requirements engineering techniques applied over the course of this dissertation. These requirements were used for the implementation and evaluation of the learning analytics infrastructure.

ODI Panel statements with importance, satisfaction and opportunity score of the students(Piller et al., 2017)

Nr	Statement	IMP	SAT	OPP
1	Minimize the likelihood that course content provided by third parties is unreliable.	56%	30%	8,2
2	Maximize the number of opportunities to work interactively and online in order to reduce the necessity of physical meetings.	26%	33%	2,6
3	Maximize the likelihood that literature is free, flexible and online available.	80%	23%	13,8
4	Increase the number of opportunities to state questions anonymously during lectures.	37%	33%	4,2
5	Increase the flexibility to study on the way.	59%	34%	8,5
6	Reduce the effort of communication in order to exchange views with other students.	42%	43%	4,2
7	Increase the flexibility to rework lectures individually (e.g. from home).	82%	30%	13,3
8	Maximize the number of opportunities to get direct feedback on e-learning exercises.	63%	21%	10,6
9	Maximize the number of achievements/success experiences during learning processes.	74%	26%	12,2
10	Increase the graphicness and clearness of course content to improve the understandability.	77%	24%	13,1
11	Reduce the likelihood to cheat oneself at learning.	47%	41%	5,4
12	Increase the number of e-learning exercises that demand to explain something in order to prove one's skills.	54%	21%	8,7
13	Reduce the effort to obtain additional information of a lecture (e.g. additional examples, further explanations, application tasks etc.).	70%	23%	11,7
14	Reduce the effort to plan the semester.	72%	28%	11,5
15	Reduce the amount of content that has to be memorized.	62%	26%	9,8
16	Reduce the likelihood that course content is not relevant for future jobs.	57%	26%	8,9
17	Minimize the likelihood that students do not engage in double loop learning processes.	55%	28%	8,2
18	Maximize the number of exercises with practical orientation.	70%	24%	11,7
19	Increase the possibility to adapt learning processes to individual learning behaviors.	43%	27%	6
20	Minimize the amount of missing learning material.	87%	42%	13,2
21	Reduce the insecurity of getting incorrect information.	78%	48%	10,8
22	Minimize the effort to exchange learning material (e.g. in study groups).	56%	44%	6,8
23	Increase the likelihood to effectively discuss course content in forums.	39%	27%	5,1
24	Reduce the restrictions of e-learning platforms in order to get access to course content of every course.	48%	34%	6,2
25	Reduce the effort to check one's personal learning progress.	56%	25%	8,8
26	Minimize the likelihood that my personal time management during learning does not work.	76%	29%	12,4
27	Reduce the effort to find certain information and content using e-learning (e.g. in study portals or platforms etc.).	72%	36%	10,8
28	Increase the fun factor of learning (e.g. in lectures).	63%	19%	10,7
29	Reduce the effort of communication between students or between students and lecturers.	43%	40%	4,6
30	Reduce the time effort to find the right contact person.	61%	38%	8,4
31	Reduce the risk of missing important statements/ questions of other students.	51%	27%	7,4
32	Increase the availability of e-learning devices in an offline environment.	62%	30%	9,3
33	Reduce the likelihood that introduced approaches to solve an exercise are not the most efficient ones.	56%	33%	7,9
34	Reduce the risk of not being able to solve an exercise on my own.	75%	26%	12,5
35	Increase the possibility to adapt learning processes to the individual learning progress.	49%	22%	7,7
36	Increase the possibility to actively participate in lectures.	38%	32%	4,5

ODI Panel statements with importance, satisfaction and opportunity score of teaching staff (Piller et al., 2017)

No.	Panel Statement	IMP	SAT	OPP
1	Reduce the effort to systematically distribute learning materials to students.	53,9%	20,9%	8,7
2	Minimize the risk of a decrease of participants in lectures due to e-learning formats.	40,6%	15,0%	6,6
3	Reduce the likelihood of technical errors when using e-learning formats.	57,5%	15,0%	10,0
4	Reduce the number of e-mails students send with questions they should be able to answer themselves.	49,2%	17,3%	8,1
5	Minimize the likelihood of misunderstandings between students and lecturers due to unclear communication.	59,4%	22,4%	9,6
6	Improve the readability of handwritings on white boards etc.	33,1%	29,5%	3,7
7	Minimize the likelihood that important courses cannot be offered due to unavailable professors.	28,3%	26,0%	3,1
8	Reduce the number of unauthorized access to course content on e-learning platforms.	36,2%	21,3%	5,1
9	Minimize the likelihood that technical devices are not compatible with e-learning formats.	48,8%	12,2%	8,5
10	Reduce the effort to adapt the course content to a new semester.	53,9%	19,7%	8,8
11	Reduce the time effort related to the development of new course content.	54,3%	11,0%	9,8
12	Reduce the time effort related to bureaucratic processes when setting up a new course.	77,2%	7,1%	14,7
13	Minimize the lack of technical support if problems with e-learning formats occur.	57,5%	15,0%	10,0
14	Reduce the time effort related to conduct exams.	52,8%	11,4%	9,4
15	Increase the likelihood that students engage in continuous learning processes.	77,6%	9,1%	14,6
16	Reduce the time of direct lectures in order to discuss the course content interactively.	36,6%	23,2%	5,0
17	Minimize the risk that students are overstained by too many e-learning formats.	48,4%	13,0%	8,4
18	Reduce the time effort to become familiar with new e-learning technologies and their application areas.	61,0%	8,7%	11,3
19	Minimize the risk of neglecting students' expectations which may results in misleading expectations.	35,8%	15,0%	5,7
20	Increase the possibility to implement creative teaching concepts and thus to foster the creativity of students.	49,2%	15,4%	8,3
21	Reduce the number of irrelevant course content in order to enable students to efficiently study for exams.	13,4%	24,0%	1,3
22	Minimize the effort to provide additional learning material in order to compensate knowledge deficiencies of students.	49,2%	15,7%	8,3
23	Increase the likelihood that students prepare and rework course content independently.	90,2%	11,4%	16,9
24	Minimize the likelihood that students snuggle with time management due to flexibility and freedom of e-learning formats.	32,3%	14,6%	5,0
25	Increase the options to work on assignments without any boundaries due to place and time.	32,3%	18,9%	4,6
26	Reduce the likelihood of infringements of copyrights when using e-learning formats.	66,9%	9,8%	12,4
27	Reduce the likelihood that students learn incorrect course content.	53,1%	22,0%	8,4
28	Increase the options to develop and adapt learning materials with other universities.	26,8%	9,4%	4,4
29	Reduce the number of students that do not participate in class.	48,8%	9,8%	8,8
30	Increase the possibility that students can state their opinion anonymously.	34,6%	18,5%	5,1
31	Reduce the likelihood that course content is too theoretical.	33,1%	24,4%	4,2
32	Reduce the number of students that missed important questions and answers in a lecture.	50,4%	25,6%	7,5
33	Minimize the opportunity to cheat in electronic exams.	59,4%	13,4%	10,6
34	Reduce the likelihood that courses do not meet the high-quality standards of the university.	59,4%	24,8%	9,4
35	Increase the number of opportunities to receive feedback during lectures.	46,5%	17,3%	7,6
36	Minimize the likelihood that that courses cannot be adapted to the individual requirements of lecturers (concerning time, content, methods gg)	52,4%	11,8%	9,3
37	Minimize the likelihood of gauging the skills of students incorrectly.	51,6%	15,0%	8,8
38	Increase the fun factor for students in lectures in order to foster learning motivation.	56,7%	18,5%	9,5
39	Reduce the likelihood that students are not intrinsically motivated to study for a lecture and rather learn to achieve a certain grade result (e.g. learn only to pass an exam etc.).	73,6%	6,3%	14,1
40	Increase the opportunity for students to learn how to write academic papers.	68,9%	10,6%	12,7
41	Increase the opportunity to evaluate groups more fairly.	47,6%	15,0%	8,0
42	Reduce the effort of grading (especially exams).	56,3%	13,8%	9,9

43	Reduce the effort of providing access to the exam after grading.	45,3%	16,1%	7,4

Questions that teaching staff has while implementing blended learning. (Anna L Dyckhoff, 2011)

Complex questions
How is the acceptance/preference of specific learning offerings differing according to user properties (e.g. previous knowledge)?
Will the access of specific learning offerings increase if lectures and exercises on the same topic are scheduled during the same week?
Are there differences in usage between specific groupings of learning offerings (e.g. between materials with or without exercises)?
Are students using specific learning materials (e.g. lecture recordings) in addition or alternatively to attendance?
How high/low is the number of the actual users in correlation to the potential target group?
How effective is the use of serious games in correlation to cognitive learning styles?
How do those low achieving students profit by continuous learning with e-test compared to those who have not yet used the e-tests?
Is the performance in e-tests somehow related to exam grades?
Do native speakers have less problems with the learning offering than non-native speakers?
Do students of all cognitive learning styles profit in equal measure?
By which properties can students be grouped?
Data consolidation
Which teaching activities increase learning activities (e.g. attendance in online discussions)?
How high/low is the usage of learning modules (materials or functions) compared to all the other offerings?
Which support offerings are accepted, due to students' reflection on their proficiency level?
How do learning offerings have to be provided and combined with support to increased usage?
Which learning offerings are preferably used to prepare or reinforce lecture-topics?
Which effects do specific learning offerings have on collaborative learning processes?
Which didactical activities facilitate continuous learning?
To which extent does the use of the learning offering ease the learning of a specific subject?
How much effort does this learning activity take compared to other learning activities?
How many (percent of the) learning modules are students viewing?
How does the use of the learning offering influence the students' motivations?
Log file analysis
Are students learning online?
Are students using specific learning offerings at home or mobile?
Are students printing learning materials?
When and how long are students learning?
When and how long are students accessing specific learning offerings (during a day)?
How often do students use a learning environment (during a week)?
How often do students attend lectures/class?
Are there specific learning offerings that are NOT used at all?
How intensely is the learning offering used for preparation of exams?
When do students use the help function?
Which features are important to the students?
Which tools do students use?

Qualitative evaluation A (B)
How do students learn with the learning offerings?
Are students learning in groups or all by themselves?
Do students like building groups?
Which interests do students have?
How did the students like the learning activity?
How difficult/easy is it to use the learning offering?
How do students like/rate/value specific learning offerings?
How satisfied are students with the learning offerings?
How do the students like the structure of the learning offering?
How do the students rate the personal gain in knowledge?
How informative was the learning activity for the students?
How useful and relieving are students perceiving specific learning offerings (in comparison)?
Are specific learning offerings suitable for learning?
How useful do students rate the learning offering?
Are students still motivated to use the learning offering for learning, after having used it?
Would they recommend the use of the learning offering in other courses?
What are students' intentions of using specific learning offerings?
Which strengths, weaknesses or possibilities for improvements do students detect?
Why do students appreciate the learning offering?

List of learning analytics indicators and their intended users

Indicator Name	Students	Teaching staff	Administration	IT Staff
Distribution of students posts in discussion forums (Ali et al., 2012; Jovanović, Gašević, Brooks, Devedžić, & Hatala, 2007)		X		
# Revisits per lesson/quiz (Ali et al., 2012; Davis et al., 2017; Jovanović et al., 2007; Kovanović et al., 2017; van Leeuwen, 2016)		X		
A group of (related) lessons? (Ali et al., 2012; Jovanović et al., 2007)		X		
Avg. cognitive load per lesson/topic? (Ali et al., 2012; Jovanović et al., 2007)		X		
Avg. number of incorrect answers per question in a quiz (Ali et al., 2012; Jovanović et al., 2007)	X	X		
Avg. number of revisits per lesson/topic (Ali et al., 2012; Cicchinelli et al., 2018; Jovanović et al., 2007)		X		
Avg. overall quiz score (Ali et al., 2012; Bakharia et al., 2016; Bos & Brand-Gruwel, 2016; Brouwer, Bredeweg, Latour, Berg, & van der Huizen, 2016; Jovanović et al., 2007)	X	X		
Avg. time spent per lesson/learning resource/quiz (Ali et al., 2012; Bos & Brand-Gruwel, 2016; Corrigan, Smeaton, Glynn, & Smyth, 2015; Crossley, Paquette, Dascalu, McNamara, & Baker, 2016; Davis et al., 2017; Davis & Hauff, 2018; Duval, 2011b; Jovanović et al., 2007; Nacu, Baltes, Hamid, Gemmell, & Pinkard, 2018)		X		
Avg. unfinished visits of a quiz (Ali et al., 2012; Jovanović et al., 2007)		X		
Standard deviation of avg. time spent on a quiz (Ali et al., 2012; Jovanović et al., 2007)	X	X		
Standard deviation of overall quiz score (Ali et al., 2012; Jovanović et al., 2007)	X	X		
Students' comprehension of the studied topics (based on his/her annotations) (Ali et al., 2012; Jovanović et al., 2007)		X		
Students' tags for the selected lesson group (Ali et al., 2012; Jovanović et al., 2007; Koulocheri & Xenos, 2013)	X	X		
The most difficult questions in a quiz (Ali et al., 2012; Bakharia et al., 2016; Jovanović et al., 2007)	X	X		
Time-depended distribution of students' posts in discussion forums (Ali et al., 2012; Jovanović et al., 2007)	X	X		
Total time spent per lesson (Ali et al., 2012; Bos & Brand-Gruwel, 2016; Duval, 2011a; Jovanović et al., 2007)		X		
All mistakes made by students (Merceron & Yacef, 2005; Scheuer & Zinn, 2007)				
Clusters of students who made a (specific) mistake (Merceron & Yacef, 2005; Scheuer & Zinn, 2007)		X		
Concepts that were involved in the mistakes (Merceron & Yacef, 2005)		X		
Frequency of mistakes made (Merceron & Yacef, 2005)		X		
Mistakes made by students in exercises that were never finished (Merceron & Yacef, 2005)		X		
Mistakes that often come together (a-priori): if students make mistake A, followed by mistake B, then later they make mistake C (Merceron & Yacef, 2005)		X		
Individual students' performance (grade) per quiz/exercise/topic/item (Aguilar et al., 2014; Bakharia et al., 2016; García-solórzano, Poblenou, Morán, Monzo, & Melenchón, 2012; Scheuer & Zinn, 2007; van Leeuwen, 2016; Zorrilla & Alvarez, 2008)	X	X		
Distribution of most frequent misconceptions/mistakes (ranking) (Scheuer & Zinn, 2007)	X	X		

Overall performance of the best X students (with more than Y exercises) in a specified time frame (Scheuer & Zinn, 2007)		**X**		
Overall performance of the best X students (with more than Y exercises) in a specified time frame including only students with specific scores (Scheuer & Zinn, 2007)		**X**		
Overall performance of the worst X students (with more than Y exercises) in a specified time frame (Scheuer & Zinn, 2007)		**X**		
Overall performance of the worst X students (with more than Y exercises) in a specified time frame including only students with specific scores (Scheuer & Zinn, 2007)		**X**		
# Messages post to forum per student (Bakharia et al., 2016; Bratitsis & Dimitracopoulou, 2008; Davis & Hauff, 2018; García-solórzano et al., 2012; Kovanović et al., 2015, 2017; Lauría, Baron, Devireddy, Sundararaju, & Jayaprakash, 2012; May et al., 2011; Riccardo Mazza & Dimitrova, 2004; Sanz-Martínez, Martínez-Monés, Bote-Lorenzo, Muñoz-Cristóbal, & Dimitriadis, 2017; Yu & Jo, 2014; Zhang & Almeroth, 2010; Zorrilla & Alvarez, 2008)		**X**		
# Messages read on forum per student (Bratitsis & Dimitracopoulou, 2008; García-solórzano et al., 2012; Kovanović et al., 2015, 2017; Lauría et al., 2012; May et al., 2011; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Zorrilla & Alvarez, 2008)		**X**		
# Messages replied to per student (Bratitsis & Dimitracopoulou, 2008; Crossley et al., 2016; García-solórzano et al., 2012; Kovanović et al., 2015; May et al., 2011; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Zorrilla & Alvarez, 2008)		**X**		
# Hits per student/user (weekly, daily, hourly) (Aguilar et al., 2014; Brouwer et al., 2016; Fritz, 2011; May et al., 2011; Sanz-Martínez et al., 2017; Zhang & Almeroth, 2010; Zorrilla & Alvarez, 2008)		**X**		**X**
# Session per student/user (weekly, daily, hourly) (Boroujeni, Sharma, Kidziński, Lucignano, & Dillenbourg, 2016; Davis et al., 2017; Fritz, 2011; García-Saiz & Zorrilla, 2011; Hlosta, Zdrahal, & Zendulka, 2017; Jo, Kim, & Yoon, 2014; Khan & Pardo, 2016; Lauría et al., 2012; Muñoz-Merino, Valiente, & Kloos, 2013; Yu & Jo, 2014; Zhang & Almeroth, 2010; Zorrilla & Alvarez, 2008)		**X**		**X**
Overall time spent per student (weekly, daily, hourly) (Bakharia et al., 2016; Boroujeni et al., 2016; Brouwer et al., 2016; García-Saiz & Zorrilla, 2011; Yu & Jo, 2014; Zhang & Almeroth, 2010; Zorrilla & Alvarez, 2008)		**X**		**X**
# Assignments submitted per student (May et al., 2011; Muñoz-Merino et al., 2013; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Sanz-Martínez et al., 2017; Zorrilla & Alvarez, 2008)		**X**		
Avg. number of sessions per week per student (Aguilar et al., 2014; Bakharia et al., 2016; Davis et al., 2017; Hlosta et al., 2017; May et al., 2011; Zorrilla & Alvarez, 2008)	**X**	**X**		
Avg. time per week per student (Boroujeni et al., 2016; García-Saiz & Zorrilla, 2011; Zorrilla & Alvarez, 2008)	**X**	**X**		**X**
Student age (Arnold & Pistilli, 2012; García-Saiz & Zorrilla, 2011; Hlosta et al., 2017; Zorrilla & Alvarez, 2008)		**X**		
Student gender (Arnold & Pistilli, 2012; García-Saiz & Zorrilla, 2011; Hlosta et al., 2017; Zorrilla & Alvarez, 2008)		**X**		
Learning paths analysis (Charleer, Klerkx, Duval, De Laet, & Verbert, 2016; Hecking, Ziebarth, & Hoppe, 2014; Schmitz et al., 2009; Zorrilla & Alvarez, 2008)		**X**		
# Assignments read per student (Kovanović et al., 2015; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Zorrilla & Alvarez, 2008)		**X**		
# Attempts in item per student (Davis et al., 2017; Muñoz-Merino et al., 2013; Tempelaar, Rienties, & Nguyen, 2018; Zorrilla & Alvarez, 2008)		**X**		
# Chat rooms entered per student (Zorrilla & Alvarez, 2008)		**X**		
# Content pages viewed per student (Khan & Pardo, 2016; Kovanović et al., 2017; Lauría et al., 2012; Pardo, Han, & Ellis, 2016; Zorrilla & Alvarez, 2008)		**X**		

# Distinct users (Glahn, Gruber, & Tartakovski, 2015; Zorrilla & Alvarez, 2008)		**X**	**X**	**X**
# Items in test per student (Zorrilla & Alvarez, 2008)		**X**		
# Messages read on mail per student (Zorrilla & Alvarez, 2008)		**X**		
# Quizzes done per student (Bakharia et al., 2016; Tempelaar et al., 2018; Zorrilla & Alvarez, 2008)		**X**		
# Quizzes failed per student (Bos & Brand-Gruwel, 2016; Tempelaar et al., 2018; Zorrilla & Alvarez, 2008)		**X**		
# Quizzes passed per student (Bos & Brand-Gruwel, 2016; Tempelaar et al., 2018; Zorrilla & Alvarez, 2008)		**X**		
# Wiki pages edited per student (Zorrilla & Alvarez, 2008)		**X**		
Avg. learning paths length/duration (Zorrilla & Alvarez, 2008)		**X**		
Avg. time spent online using the system (Davis et al., 2017; Zorrilla & Alvarez, 2008)		**X**		
Avg. visit duration (Cicchinelli et al., 2018; Zorrilla & Alvarez, 2008)		**X**		
Bar Chart: session Count, avg. pages/session, avg. time/session (Vitiello et al., 2017; Zorrilla & Alvarez, 2008)		**X**		
Delay among sessions per student (Zorrilla & Alvarez, 2008)		**X**		**X**
Frequency of usage of collaborative tools (Zorrilla & Alvarez, 2008)		**X**		**X**
Most frequent learning paths (Zorrilla & Alvarez, 2008)		**X**	**X**	
Student academic level (Zorrilla & Alvarez, 2008)		**X**		**X**
Table: Year, month, week, session count, avg. pages/session, avg. time/session (min) (Coffrin, Corrin, de Barba, & Kennedy, 2014; Vitiello et al., 2017; Zorrilla & Alvarez, 2008)				**X**
Time-depended distribution of distinct users (Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Zorrilla & Alvarez, 2008)		**X**		**X**
Top5 pages/resources (Zorrilla & Alvarez, 2008)		**X**		
Value of a lesson in relation to learning objectives (Zorrilla & Alvarez, 2008)		**X**		
# Threads started per student (Bakharia et al., 2016; May et al., 2011; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Zhang & Almeroth, 2010)	**X**	**X**		
# Quiz and assignment submissions per student (Bakharia et al., 2016; Cicchinelli et al., 2018; Davis & Hauff, 2018; Khan & Pardo, 2016; Kovanović et al., 2017; Lauría et al., 2012; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Tempelaar et al., 2018)		**X**		
Access to content pages by topic (Cicchinelli et al., 2018; Glahn et al., 2015; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Vitiello et al., 2017)		**X**		
Course access by student per date (Bakharia et al., 2016; Cicchinelli et al., 2018; Corrigan et al., 2015; Hlosta et al., 2017; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Park, Denaro, Rodriguez, Smyth, & Warschauer, 2017)		**X**		
Discussions related to course units and group activities (Koulocheri & Xenos, 2013; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)		**X**		
Global accesses to the course (Cicchinelli et al., 2018; Clow, 2014; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)		**X**		
Matrix on students' performance on quizzes related to domain (Student, concept and level of knowledge) (R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)		**X**		
Progress with the course schedule per student (R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)		**X**		

Time, discussion topic, student, number of follow-ups (Kovanović et al., 2015; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)		X		
Student risk group/status (Arnold et al., 2012)		X		
# Participants per group (Fritz, 2011; Rodríguez-Triana, Prieto, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2018)		X		X
Avg. hits per user in a group (Fritz, 2011; Rodríguez-Triana et al., 2018)		X		
Avg. session per users of each group (Fritz, 2011)		X		
# Hits per resource (weekly, daily, hourly) (Bakharia et al., 2016; Schmitz et al., 2009; Zhang & Almeroth, 2010)		X		
# Posts per student compared to group average (weekly, daily, hourly) (Zhang & Almeroth, 2010)	X	X		
# Unique users per resource (weekly, daily, hourly) (Zhang & Almeroth, 2010)		X		
Resources that have NOT been accessed (weekly, daily, hourly) (Zhang & Almeroth, 2010)		X		
Students who have NOT accessed a specific resource (weekly, daily, hourly) (Zhang & Almeroth, 2010)		X		
Viewed resources per student (weekly, daily, hourly) (Kovanović et al., 2015; Zhang & Almeroth, 2010)		X		
# Messages posted per group (Koulocheri & Xenos, 2013; May et al., 2011; Santos, Verbert, Govaerts, & Duval, 2013)		X		
# Messages read per group (May et al., 2011)		X		
# Messages replied to per group (Koulocheri & Xenos, 2013; May et al., 2011; Santos et al., 2013)		X		
# Assignments per group (May et al., 2011; Rodríguez-Triana et al., 2018)		X		
# Files downloaded per group (May et al., 2011; Rodríguez-Triana et al., 2018; van Leeuwen, 2016)		X		
# Files downloaded per student (Bos & Brand-Gruwel, 2016; Duval, 2011a; Figueira, 2015; Hlosta et al., 2017; May et al., 2011; Yu & Jo, 2014)		X		
# Files exchanged per group (May et al., 2011)		X		
# Files exchanged per student (May et al., 2011)		X		
# Files per group (May et al., 2011)		X		
# Files uploaded per group (Koulocheri & Xenos, 2013; May et al., 2011)		X		
# Files uploaded per student (Figueira, 2015; May et al., 2011)		X		
# Forums participated in per group (May et al., 2011)		X		
# Forums participated in per student (May et al., 2011)		X		
# Messages quoted per group (May et al., 2011)		X		
# Messages quoted per student (May et al., 2011)		X		
# Threads started per group (May et al., 2011)		X		
Connection frequency per group (May et al., 2011)		X		
Connection frequency per student (May et al., 2011)		X		
Message has been displayed, read partially or entirely? (May et al., 2011)	X	X		
Time gap between two different readings of a specific message		X		

(Bogarín, Romero, Cerezo, & Sánchez-Santillán, 2014; May et al., 2011)				
Time spent by each user on reading a message (Kovanović et al., 2015; May et al., 2011)		**X**		
Advice because of the student's progress / delay in time (Kosba, Dimitrova, & Boyle, 2005)		**X**		
Advice because student has (not) completely mastered the related prerequisite concepts (Kosba et al., 2005)	**X**	**X**		
Advice because student has (not) completely read and worked on the learning objects and assessment quizzes related to the concept (Kosba et al., 2005)	**X**	**X**		
Advice because student has (not) participated in the communication activities related to the concept (Kosba et al., 2005)	**X**	**X**		
Advice to groups concerning common problems (Kosba et al., 2005)	**X**	**X**		
Advice to groups concerning satisfactory/unsatisfactory learning levels (Kosba et al., 2005)	**X**	**X**		
Advice to groups concerning uncommunicative behavior (Kosba et al., 2005)	**X**	**X**		
Advice to groups related to the knowledge states of groups (Kosba et al., 2005)	**X**	**X**		
Advice to students who have not started working on the course (Kosba et al., 2005)	**X**	**X**		
Advice to students who have unsatisfactory learning levels because they are weak and normally communicative (Kosba et al., 2005)	**X**	**X**		
Advice to students who have unsatisfactory learning levels because they are weak and uncommunicative (Kosba et al., 2005)	**X**	**X**		
Advice to students who have unsatisfactory learning levels because they are weak but highly communicative (Kosba et al., 2005)		**X**		
Advice to the teacher concerning excellent and weak students relative to the whole class (Kosba et al., 2005)		**X**		
Advice to the teacher concerning least and most active students relative to the class (Kosba et al., 2005)		**X**		
Advice to the teacher concerning least and most communicative students relative to the class (Kosba et al., 2005)		**X**		
Advice to the teacher concerning teaching parts of the course causing problems to the majority of students (Kosba et al., 2005)		**X**		
Advice to the teacher concerning the most preferable types of learning objects visited by the students (Kosba et al., 2005)		**X**		
Advice to the teacher concerning the status and behavior of the whole class (Kosba et al., 2005)		**X**		
Advice to uncommunicative advanced students to help others (Kosba et al., 2005)	**X**			
Goal: prediction of final mark according to student's activity (García-Saiz & Zorrilla, 2011)	**X**	**X**		
Resources frequently used together (forum, mail, etc.) in each learning session (García-Saiz & Zorrilla, 2011)	**X**	**X**		
Patterns of interactions between specific participants (De Groot et al., 2007)		**X**		
Relation between contributions? (De Groot et al., 2007)		**X**		
Resources watched (read) in a session per student (via webcam) (González Agulla, Alba Castro, Argones Rúa, & Anido Rifón, 2009)	**X**	**X**		
Bookmarked learning resources (in a selected week) (Schmitz et al., 2009)		**X**		

Daily content history/log (time, file, action) (Schmitz et al., 2009)		X		
Downloaded learning resources (in a selected week) (Duval, 2011a; Schmitz et al., 2009)	X	X		
Keyword analysis derived from email exchange data per student (Schmitz et al., 2009)		X		
Social network derived from email exchange data per student (Schmitz et al., 2009)		X		
Students' tag cloud of learning resources (Schmitz et al., 2009)	X	X		
Students' tag cloud of learning resources in a selected past week (Schmitz et al., 2009)	X	X		
Topics of interest (Schmitz et al., 2009)	X	X		
Combination of (parallel coordinates) total time spent on the course, avg. time spent a resource, number of resources used, and median of the time of day a student works (Govaerts & Duval, 2012)	X	X		
Resources students spent time with (Govaerts & Duval, 2012)		X		
Students' distribution for total time spent and resources used (bar chart) (Govaerts & Duval, 2012)	X	X		
Trends in students activity (based on time spent) (Govaerts & Duval, 2012)	X	X		
# Answers to user by others per student (Bratitsis & Dimitracopoulou, 2008; Kovanović et al., 2017)		X		
# Follow-up contributions per group (Bratitsis & Dimitracopoulou, 2008)		X		
# Type t posts per group (Bratitsis & Dimitracopoulou, 2008)		X		
# Type t posts per student (Bratitsis & Dimitracopoulou, 2008)		X		
Activity indicator regarding # posts, # posts reads, # type per group (Bratitsis & Dimitracopoulou, 2008)		X		
Activity indicator regarding # posts, posts reads, type per student (Bratitsis & Dimitracopoulou, 2008; Santos, Govaerts, Verbert, & Duval, 2012)		X		
Avg. number of posts per group (Bratitsis & Dimitracopoulou, 2008)		X		
Avg. number of posts per student (Bratitsis & Dimitracopoulou, 2008)		X		
Avg. thread depths/weight (Bratitsis & Dimitracopoulou, 2008)		X		
Contribution indicator per group (Bratitsis & Dimitracopoulou, 2008)	X	X		
Contribution indicator per student (Bratitsis & Dimitracopoulou, 2008)		X		
Group interactivity indicator (Bratitsis & Dimitracopoulou, 2008)	X	X		
Relative activity regarding # posts, # types, # initiated threads per group (Bratitsis & Dimitracopoulou, 2008)		X		
Relative activity regarding # posts, # types, # initiated threads per student (Bratitsis & Dimitracopoulou, 2008)		X		
Relative posts read to number of posts per group (Bratitsis & Dimitracopoulou, 2008)	X	X		
Social Network Analysis (Bratitsis & Dimitracopoulou, 2008)	X	X		
Thread Depths/weight per group (Bratitsis & Dimitracopoulou, 2008)	X	X		
Thread Depths/weight per student (Bratitsis & Dimitracopoulou, 2008)	X	X		
Tree structure of the forum, highlighting the messages of a corresponding student (Bratitsis & Dimitracopoulou, 2008)		X		
User classification indicators: posts reads relative to # posts per student (Bratitsis & Dimitracopoulou, 2008)		X		
User type messages per week (question, clarification, argument, disagreement, etc.)		X		

(Bratitsis & Dimitracopoulou, 2008)				
# Messages per participant per course phase (Davis et al., 2017; Petropoulou, Lazakidou, Georgiakakis, & Retalis, 2012)		**X**		
Argumentation (message annotation) per course phase (Petropoulou et al., 2012)		**X**		
Avg. number of contributions per course phase (Petropoulou et al., 2012)		**X**		
Collaboration (interaction base message characterization) per course phase (Petropoulou et al., 2012)		**X**		
Participation count (number of posted messages) per course phase (Petropoulou et al., 2012)		**X**		
SNA: actor's degree centrality per course phase (Petropoulou et al., 2012)		**X**		
Work amount (message dimension per user) per course phase (Petropoulou et al., 2012)		**X**		
Learner isolation / students with limited connectivity (Bakharia & Dawson, 2011; Dawson, Bakharia, & Heathcote, 2010)	**X**	**X**		
Social graph of student community (Bakharia & Dawson, 2011; Dawson et al., 2010)		**X**		
Social graphs of forum interaction data (Bakharia & Dawson, 2011; Dawson et al., 2010)		**X**		
Sociogram of interaction between teacher and participant (Bakharia & Dawson, 2011; Dawson et al., 2010)		**X**		
Frequency a student used each keyword (Mochizuki et al., 2005)	**X**	**X**		
Level of participation per student (Mochizuki et al., 2005)	**X**	**X**		
Relation between keywords and students (Mochizuki et al., 2005)	**X**	**X**		
# Messages read by the user in relation to # messages available (Brooks, Panesar, & Greer, 2006)	**X**	**X**		
Content currently read by one or more students (Brooks et al., 2006)	**X**	**X**		
Sociogram of participants, lurkers, and delinquents (Brooks et al., 2006)		**X**		
Contribution of each student to his/her group's online communication (Janssen, Erkens, Kanselaar, & Jaspers, 2007)	**X**	**X**		
Final Grade (Score between 0 and 100) (Brouwer et al., 2016; Davis & Hauff, 2018; Kovanović et al., 2015; Pardo et al., 2016; van Leeuwen, 2016)	**X**	**X**		
Course Completion (pass fail, binary indicator) (Davis & Hauff, 2018)	**X**	**X**		
Video interactions (play, pause, fast forward, rewind, scrub) (Crossley et al., 2016; Davis & Hauff, 2018; Khan & Pardo, 2016; Kovanović et al., 2017; Li, Kidziński, Jermann, & Dillenbourg, 2015; Pardo et al., 2016; Pijeira Díaz, Ruiz, Ruipérez-Valiente, Muñoz-Merino, & Delgado Kloos, 2015; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Vitiello et al., 2017)		**X**	**X**	**X**
Physical attendance (Nacu et al., 2018)		**X**	**X**	**X**
Histogram Performance of peers for each key moment (Millecamp, Gutiérrez, Charleer, Verbert, & De Laet, 2018)	**X**	**X**		
Key moment with all courses of that moment (Millecamp et al., 2018)		**X**		
Peer performance for a course (Millecamp et al., 2018)	**X**	**X**		
Failed courses per student (Millecamp et al., 2018)		**X**		
Study trajectory of previous students with similar profile (Millecamp et al., 2018; Yousuf & Conlan, 2015)	**X**	**X**		
Study trajectories for different profiles (Millecamp et al., 2018)	**X**	**X**	**X**	
Individual student score and compare to scores of other students (Broos, Verbert, Langie, Van Soom, & De Laet, 2018; Davis, Chen, van der Zee, Hauff, & Houben, 2016)	**X**	**X**		

Compare individual student score to scores from last year students (Broos et al., 2018)	X	X	X	
Tips based on the student score for improvement (Broos et al., 2018)	X			
Self-regulation behavior: Planning activities, and regulation activities (Cicchinelli et al., 2018)	X			
Total clicks per session (Brouwer et al., 2016; Cicchinelli et al., 2018; Hlosta et al., 2017; Vitiello et al., 2017)	X	X		
Average time monitoring per session (Cicchinelli et al., 2018; Hlosta et al., 2017)	X	X		
Average regulation per session (Cicchinelli et al., 2018)	X	X		
Average time monitoring (Cicchinelli et al., 2018)	X	X		
Average time until a student returns to the platform after each class (Cicchinelli et al., 2018)		X		
Mastery, the proportion of all exercises successfully answered in relation to the total number of exercises/quizzes (percentage) (Tempelaar et al., 2018)	X	X		
Learner watched video before submitting assignment (Boroujeni & Dillenbourg, 2018)		X		
Learning submitted assignment without watching videos (Boroujeni & Dillenbourg, 2018)		X		
Learner watched video without submitting assignment (Boroujeni & Dillenbourg, 2018)		X		
Learner did not watch video and did not submit assignment (Bakharia et al., 2016; Boroujeni & Dillenbourg, 2018)		X		
Correlation between quiz scores and number of events (clicks) (Shimada, Taniguchi, Okubo, Konomi, & Ogata, 2018)	X	X		
Proportion of students who click each day in a course (Hlosta et al., 2017; Park et al., 2017)		X		
Average number of clicks per student and per course per day (Park et al., 2017)		X	X	X
Percentage of number of students who increased or decreased in a course over time (Glahn et al., 2015; Park et al., 2017)		X		
Number of withdrawn students in days relative to the start of the course (Hlosta et al., 2017)		X	X	
Flag indicating student submission (yes/no) (Hlosta et al., 2017)		X		
Number of forum messages per day (Bakharia et al., 2016; Boroujeni, Hecking, Hoppe, & Dillenbourg, 2017; Davis et al., 2017; Kovanović et al., 2017)		X		X
Number of forum messages after a video release in the learning resources (Boroujeni et al., 2017)		X		X
Number of forum messages before an assignment deadline (Boroujeni et al., 2017)		X		
Number of specific keywords in posts per user in a course (Boroujeni et al., 2017)		X		
Number of new discussions per day (Boroujeni et al., 2017)		X		X
Number of students/forum contributors per day in the discussions (Boroujeni et al., 2017)	X	X		
Mean time between successive forum writing events/messages (Boroujeni et al., 2017)		X		
Mean time between successive started threads (Boroujeni et al., 2017)		X		
Mean and standard deviation of threads count per day (Boroujeni et al., 2017)		X		
Mean and standard deviation of message counts per day (Boroujeni et al., 2017)		X		
Mean and standard deviation of forum contributors per day (Boroujeni et al., 2017)		X		
Percentage of lecture videos partially or totally watched in a course chapter (Bote-Lorenzo & Gómez-Sánchez, 2017; Crossley et al., 2016)		X		X

Percentage of finger exercises answered in a course chapter (Bote-Lorenzo & Gómez-Sánchez, 2017)		X		
Percentage of assignments submitted in a course chapter (Bote-Lorenzo & Gómez-Sánchez, 2017; Crossley et al., 2016)		X		
Normalized grade of finger exercises in a given chapter (Bote-Lorenzo & Gómez-Sánchez, 2017)		X		
Normalized grade of assignment in a given chapter (Bote-Lorenzo & Gómez-Sánchez, 2017)		X		
Value of video engagement indicator (Bote-Lorenzo & Gómez-Sánchez, 2017; Yousuf & Conlan, 2015)	X	X		
Value of exercise engagement indicator (Bote-Lorenzo & Gómez-Sánchez, 2017; Yousuf & Conlan, 2015)	X	X		
Value of assignment engagement indicator (Bote-Lorenzo & Gómez-Sánchez, 2017; Yousuf & Conlan, 2015)	X	X		
Distinctive words from selected planning topics that predict course completion (Yeomans & Reich, 2017)	X	X		
Quiz submission timelines in days (Davis et al., 2017)		X	X	X
Time spent by each user on the platform in hours (Bos & Brand-Gruwel, 2016; Davis et al., 2017; Jo et al., 2014; Muñoz-Merino et al., 2013; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Santos et al., 2013)		X	X	X
Time spent by each user in watching videos in hours (Bos & Brand-Gruwel, 2016; Brouwer et al., 2016; Davis et al., 2017; Pijeira Díaz et al., 2015)		X	X	X
Number of accessed videos per user (Davis et al., 2017; Muñoz-Merino et al., 2013; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014)		X		X
Number of revisited videos per user (Davis et al., 2017; Pijeira Díaz et al., 2015)		X		X
Number of forum visits (Crossley et al., 2016; Davis et al., 2017; Kovanović et al., 2015, 2017; Vitiello et al., 2017)		X		
Average length of a session (Davis et al., 2017; Khan & Pardo, 2016; Vitiello et al., 2017)		X		X
Mean time between two sessions (Davis et al., 2017)		X		X
Average Time between the first and last attempt for an assignment for a student (Crossley et al., 2016)		X		
Number of completed finger exercises/multiple choice questions by the video (Khan & Pardo, 2016; Pardo et al., 2016)		X		
Time spent on course assignments per user (Bogarín et al., 2014; Kovanović et al., 2015; Muñoz-Merino et al., 2013)		X		
Total time spent viewing course resources (Bogarín et al., 2014; Kovanović et al., 2015)		X		
Total time spent viewing course discussions (Kovanović et al., 2015)		X		
Total time spent posting in course discussions (Bogarín et al., 2014; Kovanović et al., 2015)		X		
Total time spent updating discussion messages (Bogarín et al., 2014; Kovanović et al., 2015)		X		
Total time spent on a question in a quiz giving a correct answer per user (Papamitsiou, Terzis, & Economides, 2014)		X		
Total time spent on a question in a quiz giving a wrong answer per user (Papamitsiou et al., 2014)		X		
How many times the student views a question in a quiz? (Papamitsiou et al., 2014)		X		
How many times the student changes the answer of a question in a quiz? (Papamitsiou et al., 2014)		X		
How much time the student spends on viewing a question in a quiz? (Papamitsiou et al., 2014)		X		
How much time the students spends on saving an answer to a question in a quiz? (Papamitsiou et al., 2014)		X		
How much time a student spends on quizzes? (Bogarín et al., 2014; Muñoz-Merino et al., 2013)		X		

How long the student waits in days to check new learning content counting from the time the learning content was available? (Bogarín et al., 2014)	X	X		
How long the student waits in days to check new quizzes or assignments counting from the time the learning content was available? (Bogarín et al., 2014)	X	X		
How many words have the students their forum posts? (Bogarín et al., 2014)		X		
How many sentences are there in the forum posts of the students? (Bogarín et al., 2014)		X		
Effective correct progress on the platform (Muñoz-Merino et al., 2013; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014)	X	X		
Efficient correct progress on the learning platform (Muñoz-Merino et al., 2013)	X	X		
Total working schedule that shows at which time of the day the learner watches the lecture videos (Muñoz-Merino et al., 2013)		X		X
Use of optional activities in a course? (Ruipérez-Valiente, Muñoz-Merino, Kloos, Niemann, & Scheffel, 2014)		X		
Ratio of accesses/clicks on weekdays vs weekend (Corrigan et al., 2015)		X		X
Ratio of on campus to off campus accesses/clicks (Corrigan et al., 2015)		X		X
Average quiz responses during the lecture and in the exam period (Glahn et al., 2015)		X		
Student engagement comparison per activity (Yousuf & Conlan, 2015)		X		
Number of students who have achieved a badge (including the user student) (Charleer et al., 2016)	X	X		
How many times a student has been awarded a specific badge? (Charleer et al., 2016)	X	X		
Number of badges awarded on each day (Charleer et al., 2016)	X	X		
Number of awards given per badge (Charleer et al., 2016)	X	X		
Individual student grade and compare to grades of other students (Davis et al., 2016)	X	X		
Studying on certain hours of the day per user (Boroujeni et al., 2016)		X		X
Studying on certain days of the week per user (Boroujeni et al., 2016)		X		X
Studying on similar weekdays, over weeks of the course (Boroujeni et al., 2016)		X		X
Time in seconds/minutes/hours between the submission time and the assignment deadline (Brouwer et al., 2016)		X		
Total clicks per day/week/month (Vitiello et al., 2017)		X	X	X

List of learning analytics indicators and what kind of data they need

Indicator Name	**Anonymous Data**	**Pseudomized Data**	**Personal Private Data**	**Course/Campus Data**
Distribution of students posts in discussion forums (Ali et al., 2012; Jovanović et al., 2007)	**X**	**X**		
# Revisits per lesson/quiz (Ali et al., 2012; Davis et al., 2017; Jovanović et al., 2007; Kovanović et al., 2017; van Leeuwen, 2016)	**X**			
A group of (related) lessons? (Ali et al., 2012; Jovanović et al., 2007)	**X**	**X**		
Avg. cognitive load per lesson/topic? (Ali et al., 2012; Jovanović et al., 2007)	**X**			
Avg. number of incorrect answers per question in a quiz (Ali et al., 2012; Jovanović et al., 2007)	**X**			
Avg. number of revisits per lesson/topic (Ali et al., 2012; Cicchinelli et al., 2018; Jovanović et al., 2007)	**X**			
Avg. overall quiz score (Ali et al., 2012; Bakharia et al., 2016; Bos & Brand-Gruwel, 2016; Brouwer et al., 2016; Jovanović et al., 2007)	**X**			
Avg. time spent per lesson/learning resource/quiz (Ali et al., 2012; Bos & Brand-Gruwel, 2016; Corrigan et al., 2015; Crossley et al., 2016; Davis et al., 2017; Davis & Hauff, 2018; Duval, 2011b; Jovanović et al., 2007; Nacu et al., 2018)	**X**			
Avg. unfinished visits of a quiz (Ali et al., 2012; Jovanović et al., 2007)	**X**			
Standard deviation of avg. time spent on a quiz (Ali et al., 2012; Jovanović et al., 2007)	**X**			
Standard deviation of overall quiz score (Ali et al., 2012; Jovanović et al., 2007)	**X**			
Students' comprehension of the studied topics (based on his/her annotations) (Ali et al., 2012; Jovanović et al., 2007)			**X**	
Students' tags for the selected lesson group (Ali et al., 2012; Jovanović et al., 2007; Koulocheri & Xenos, 2013)		**X**		
The most difficult questions in a quiz (Ali et al., 2012; Bakharia et al., 2016; Jovanović et al., 2007)	**X**			
Time-depended distribution of students' posts in discussion forums (Ali et al., 2012; Jovanović et al., 2007)	**X**			
Total time spent per lesson (Ali et al., 2012; Bos & Brand-Gruwel, 2016; Duval, 2011a; Jovanović et al., 2007)	**X**	**X**		
All mistakes made by students (Merceron & Yacef, 2005; Scheuer & Zinn, 2007)	**X**	**X**		
Clusters of students who made a (specific) mistake (Merceron & Yacef, 2005; Scheuer & Zinn, 2007)		**X**		
Concepts that were involved in the mistakes (Merceron & Yacef, 2005)	**X**			
Frequency of mistakes made (Merceron & Yacef, 2005)	**X**			
Mistakes made by students in exercises that were never finished (Merceron & Yacef, 2005)		**X**		
Mistakes that often come together (a-priori): if students make mistake A, followed by mistake B, then later they make mistake C (Merceron & Yacef, 2005)		**X**	**X**	
Individual students' performance (grade) per quiz/exercise/topic/item (Aguilar et al., 2014; Bakharia et al., 2016; García-solórzano et al., 2012; Scheuer & Zinn, 2007; van Leeuwen, 2016; Zorrilla & Alvarez, 2008)			**X**	

Distribution of most frequent misconceptions/mistakes (ranking) (Scheuer & Zinn, 2007)	X	X		
Overall performance of the best X students (with more than Y exercises) in a specified time frame (Scheuer & Zinn, 2007)		X		
Overall performance of the best X students (with more than Y exercises) in a specified time frame including only students with specific scores (Scheuer & Zinn, 2007)			X	
Overall performance of the worst X students (with more than Y exercises) in a specified time frame (Scheuer & Zinn, 2007)		X		
Overall performance of the worst X students (with more than Y exercises) in a specified time frame including only students with specific scores (Scheuer & Zinn, 2007)			X	
# Messages post to forum per student (Bakharia et al., 2016; Bratitsis & Dimitracopoulou, 2008; Davis & Hauff, 2018; García-solórzano et al., 2012; Kovanović et al., 2015, 2017; Lauría et al., 2012; May et al., 2011; Riccardo Mazza & Dimitrova, 2004; Sanz-Martínez et al., 2017; Yu & Jo, 2014; Zhang & Almeroth, 2010; Zorrilla & Alvarez, 2008)		X		
# Messages read on forum per student (Bratitsis & Dimitracopoulou, 2008; García-solórzano et al., 2012; Kovanović et al., 2015, 2017; Lauría et al., 2012; May et al., 2011; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Zorrilla & Alvarez, 2008)		X		
# Messages replied to per student (Bratitsis & Dimitracopoulou, 2008; Crossley et al., 2016; García-solórzano et al., 2012; Kovanović et al., 2015; May et al., 2011; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Zorrilla & Alvarez, 2008)		X	X	
# Hits per student/user (weekly, daily, hourly) (Aguilar et al., 2014; Brouwer et al., 2016; Fritz, 2011; May et al., 2011; Sanz-Martínez et al., 2017; Zhang & Almeroth, 2010; Zorrilla & Alvarez, 2008)	X	X		
# Session per student/user (weekly, daily, hourly) (Boroujeni et al., 2016; Davis et al., 2017; Fritz, 2011; García-Saiz & Zorrilla, 2011; Hlosta et al., 2017; Jo et al., 2014; Khan & Pardo, 2016; Lauría et al., 2012; Muñoz-Merino et al., 2013; Yu & Jo, 2014; Zhang & Almeroth, 2010; Zorrilla & Alvarez, 2008)	X	X		
Overall time spent per student (weekly, daily, hourly) (Bakharia et al., 2016; Boroujeni et al., 2016; Brouwer et al., 2016; García-Saiz & Zorrilla, 2011; Yu & Jo, 2014; Zhang & Almeroth, 2010; Zorrilla & Alvarez, 2008)	X	X		
# Assignments submitted per student (May et al., 2011; Muñoz-Merino et al., 2013; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Sanz-Martínez et al., 2017; Zorrilla & Alvarez, 2008)	X	X		
Avg. number of sessions per week per student (Aguilar et al., 2014; Bakharia et al., 2016; Davis et al., 2017; Hlosta et al., 2017; May et al., 2011; Zorrilla & Alvarez, 2008)	X	X		
Avg. time per week per student (Boroujeni et al., 2016; García-Saiz & Zorrilla, 2011; Zorrilla & Alvarez, 2008)	X	X		
Student age (Arnold & Pistilli, 2012; García-Saiz & Zorrilla, 2011; Hlosta et al., 2017; Zorrilla & Alvarez, 2008)			X	
Student gender (Arnold & Pistilli, 2012; García-Saiz & Zorrilla, 2011; Hlosta et al., 2017; Zorrilla & Alvarez, 2008)			X	
Learning paths analysis (Charleer et al., 2016; Hecking et al., 2014; Schmitz et al., 2009; Zorrilla & Alvarez, 2008)		X		
# Assignments read per student (Kovanović et al., 2015; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Zorrilla & Alvarez, 2008)	X	X		
# Attempts in item per student (Davis et al., 2017; Muñoz-Merino et al., 2013; Tempelaar et al., 2018; Zorrilla & Alvarez, 2008)	X	X		
# Chat rooms entered per student (Zorrilla & Alvarez, 2008)	X	X		
# Content pages viewed per student (Khan & Pardo, 2016; Kovanović et al., 2017; Lauría et al., 2012; Pardo et al., 2016; Zorrilla & Alvarez, 2008)	X	X		

# Distinct users (Glahn et al., 2015; Zorrilla & Alvarez, 2008)	X	X		
# Items in test per student (Zorrilla & Alvarez, 2008)		X		
# Messages read on mail per student (Zorrilla & Alvarez, 2008)		X		
# Quizzes done per student (Bakharia et al., 2016; Tempelaar et al., 2018; Zorrilla & Alvarez, 2008)		X		
# Quizzes failed per student (Bos & Brand-Gruwel, 2016; Tempelaar et al., 2018; Zorrilla & Alvarez, 2008)		X		
# Quizzes passed per student (Bos & Brand-Gruwel, 2016; Tempelaar et al., 2018; Zorrilla & Alvarez, 2008)		X		
# Wiki pages edited per student (Zorrilla & Alvarez, 2008)		X		
Avg. learning paths length/duration (Zorrilla & Alvarez, 2008)	X			
Avg. time spent online using the system (Davis et al., 2017; Zorrilla & Alvarez, 2008)	X			
Avg. visit duration (Cicchinelli et al., 2018; Zorrilla & Alvarez, 2008)	X			
Bar Chart: session Count, avg. pages/session, avg. time/session (Vitiello et al., 2017; Zorrilla & Alvarez, 2008)	X			
Delay among sessions per student (Zorrilla & Alvarez, 2008)	X	X		
Frequency of usage of collaborative tools (Zorrilla & Alvarez, 2008)	X			
Most frequent learning paths (Zorrilla & Alvarez, 2008)	X	X		
Student academic level (Zorrilla & Alvarez, 2008)			X	
Table: Year, month, week, session count, avg. pages/session, avg. time/session (min) (Coffrin et al., 2014; Vitiello et al., 2017; Zorrilla & Alvarez, 2008)	X			
Time-depended distribution of distinct users (Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Zorrilla & Alvarez, 2008)	X			
Top5 pages/resources (Zorrilla & Alvarez, 2008)	X			
Value of a lesson in relation to learning objectives (Zorrilla & Alvarez, 2008)	X			
# Threads started per student (Bakharia et al., 2016; May et al., 2011; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Zhang & Almeroth, 2010)		X		
# Quiz and assignment submissions per student (Bakharia et al., 2016; Cicchinelli et al., 2018; Davis & Hauff, 2018; Khan & Pardo, 2016; Kovanović et al., 2017; Lauría et al., 2012; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Tempelaar et al., 2018)		X		
Access to content pages by topic (Cicchinelli et al., 2018; Glahn et al., 2015; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Vitiello et al., 2017)	X			
Course access by student per date (Bakharia et al., 2016; Cicchinelli et al., 2018; Corrigan et al., 2015; Hlosta et al., 2017; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007; Park et al., 2017)	X			
Discussions related to course units and group activities (Koulocheri & Xenos, 2013; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)		X		
Global accesses to the course (Cicchinelli et al., 2018; Clow, 2014; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)	X			X
Matrix on students' performance on quizzes related to domain (Student, concept and level of knowledge) (R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)		X	X	
Progress with the course schedule per student (R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)			X	
Time, discussion topic, student, number of follow-ups (Kovanović et al., 2015; R. Mazza & Milani, 2004; Riccardo Mazza & Dimitrova, 2007)			X	
Student risk group/status			X	

(Arnold et al., 2012)				
# Participants per group (Fritz, 2011; Rodríguez-Triana et al., 2018)	**X**			
Avg. hits per user in a group (Fritz, 2011; Rodríguez-Triana et al., 2018)		**X**		
Avg. session per users of each group (Fritz, 2011)		**X**		
# Hits per resource (weekly, daily, hourly) (Bakharia et al., 2016; Schmitz et al., 2009; Zhang & Almeroth, 2010)	**X**	**X**		
# Posts per student compared to group average (weekly, daily, hourly) (Zhang & Almeroth, 2010)		**X**		
# Unique users per resource (weekly, daily, hourly) (Zhang & Almeroth, 2010)	**X**			
Resources that have NOT been accessed (weekly, daily, hourly) (Zhang & Almeroth, 2010)	**X**			
Students who have NOT accessed a specific resource (weekly, daily, hourly) (Zhang & Almeroth, 2010)		**X**		
Viewed resources per student (weekly, daily, hourly) (Kovanović et al., 2015; Zhang & Almeroth, 2010)		**X**		
# Messages posted per group (Koulocheri & Xenos, 2013; May et al., 2011; Santos et al., 2013)		**X**		
# Messages read per group (May et al., 2011)		**X**		
# Messages replied to per group (Koulocheri & Xenos, 2013; May et al., 2011; Santos et al., 2013)		**X**		
# Assignments per group (May et al., 2011; Rodríguez-Triana et al., 2018)		**X**		
# Files downloaded per group (May et al., 2011; Rodríguez-Triana et al., 2018; van Leeuwen, 2016)		**X**		
# Files downloaded per student (Bos & Brand-Gruwel, 2016; Duval, 2011a; Figueira, 2015; Hlosta et al., 2017; May et al., 2011; Yu & Jo, 2014)		**X**		
# Files exchanged per group (May et al., 2011)		**X**		
# Files exchanged per student (May et al., 2011)		**X**	**X**	
# Files per group (May et al., 2011)		**X**		
# Files uploaded per group (Koulocheri & Xenos, 2013; May et al., 2011)		**X**	**X**	
# Files uploaded per student (Figueira, 2015; May et al., 2011)		**X**	**X**	
# Forums participated in per group (May et al., 2011)	**X**		**X**	
# Forums participated in per student (May et al., 2011)		**X**	**X**	
# Messages quoted per group (May et al., 2011)		**X**		
# Messages quoted per student (May et al., 2011)		**X**		
# Threads started per group (May et al., 2011)		**X**	**X**	
Connection frequency per group (May et al., 2011)		**X**	**X**	
Connection frequency per student (May et al., 2011)		**X**	**X**	
Message has been displayed, read partially or entirely? (May et al., 2011)			**X**	
Time gap between two different readings of a specific message (Bogarín et al., 2014; May et al., 2011)	**X**			
Time spent by each user on reading a message (Kovanović et al., 2015; May et al., 2011)			**X**	
Advice because of the student's progress / delay in time (Kosba et al., 2005)			**X**	

Advice because student has (not) completely mastered the related prerequisite concepts (Kosba et al., 2005)			X	
Advice because student has (not) completely read and worked on the learning objects and assessment quizzes related to the concept (Kosba et al., 2005)			X	
Advice because student has (not) participated in the communication activities related to the concept (Kosba et al., 2005)			X	
Advice to groups concerning common problems (Kosba et al., 2005)			X	
Advice to groups concerning satisfactory/unsatisfactory learning levels (Kosba et al., 2005)			X	
Advice to groups concerning uncommunicative behavior (Kosba et al., 2005)			X	
Advice to groups related to the knowledge states of groups (Kosba et al., 2005)			X	
Advice to students who have not started working on the course (Kosba et al., 2005)			X	
Advice to students who have unsatisfactory learning levels because they are weak and normally communicative (Kosba et al., 2005)			X	
Advice to students who have unsatisfactory learning levels because they are weak and uncommunicative (Kosba et al., 2005)			X	
Advice to students who have unsatisfactory learning levels because they are weak but highly communicative (Kosba et al., 2005)			X	
Advice to the teacher concerning excellent and weak students relative to the whole class (Kosba et al., 2005)	X	X		
Advice to the teacher concerning least and most active students relative to the class (Kosba et al., 2005)	X	X		
Advice to the teacher concerning least and most communicative students relative to the class (Kosba et al., 2005)	X	X		
Advice to the teacher concerning teaching parts of the course causing problems to the majority of students (Kosba et al., 2005)	X			
Advice to the teacher concerning the most preferable types of learning objects visited by the students (Kosba et al., 2005)	X			
Advice to the teacher concerning the status and behavior of the whole class (Kosba et al., 2005)	X			
Advice to uncommunicative advanced students to help others (Kosba et al., 2005)	X			
Goal: prediction of final mark according to student's activity (García-Saiz & Zorrilla, 2011)			X	
Resources frequently used together (forum, mail, etc.) in each learning session (García-Saiz & Zorrilla, 2011)	X			
Patterns of interactions between specific participants (De Groot et al., 2007)		X		
Relation between contributions? (De Groot et al., 2007)		X		
Resources watched (read) in a session per student (via webcam) (González Agulla et al., 2009)			X	
Bookmarked learning resources (in a selected week) (Schmitz et al., 2009)	X			
Daily content history/log (time, file, action) (Schmitz et al., 2009)	X			
Downloaded learning resources (in a selected week)	X			

(Duval, 2011a; Schmitz et al., 2009)				
Keyword analysis derived from email exchange data per student (Schmitz et al., 2009)			X	
Social network derived from email exchange data per student (Schmitz et al., 2009)			X	
Students' tag cloud of learning resources (Schmitz et al., 2009)			X	
Students' tag cloud of learning resources in a selected past week (Schmitz et al., 2009)			X	
Topics of interest (Schmitz et al., 2009)			X	
Combination of (parallel coordinates) total time spent on the course, avg. time spent a resource, number of resources used, and median of the time of day a student works (Govaerts & Duval, 2012)			X	
Resources students spent time with (Govaerts & Duval, 2012)	X	X		
Students' distribution for total time spent and resources used (bar chart) (Govaerts & Duval, 2012)	X	X		
Trends in students' activity (based on time spent) (Govaerts & Duval, 2012)	X			
# Answers to user by others per student (Bratitsis & Dimitracopoulou, 2008; Kovanović et al., 2017)		X	X	
# Follow-up contributions per group (Bratitsis & Dimitracopoulou, 2008)		X	X	
# Type t posts per group (Bratitsis & Dimitracopoulou, 2008)			X	
# Type t posts per student (Bratitsis & Dimitracopoulou, 2008)			X	
Activity indicator regarding # posts, # posts reads, # type per group (Bratitsis & Dimitracopoulou, 2008)		X	X	
Activity indicator regarding # posts, posts reads, type per student (Bratitsis & Dimitracopoulou, 2008; Santos et al., 2012)			X	
Avg. number of posts per group (Bratitsis & Dimitracopoulou, 2008)	X			
Avg. number of posts per student (Bratitsis & Dimitracopoulou, 2008)		X		
Avg. thread depths/weight (Bratitsis & Dimitracopoulou, 2008)	X			
Contribution indicator per group (Bratitsis & Dimitracopoulou, 2008)		X		
Contribution indicator per student (Bratitsis & Dimitracopoulou, 2008)		X		
Group interactivity indicator (Bratitsis & Dimitracopoulou, 2008)		X		
Relative activity regarding # posts, # types, # initiated threads per group (Bratitsis & Dimitracopoulou, 2008)		X	X	
Relative activity regarding # posts, # types, # initiated threads per student (Bratitsis & Dimitracopoulou, 2008)		X	X	
Relative posts read to number of posts per group (Bratitsis & Dimitracopoulou, 2008)		X	X	
Social Network Analysis (Bratitsis & Dimitracopoulou, 2008)			X	
Thread Depths/weight per group (Bratitsis & Dimitracopoulou, 2008)			X	
Thread Depths/weight per student (Bratitsis & Dimitracopoulou, 2008)			X	
Tree structure of the forum, highlighting the messages of a corresponding student (Bratitsis & Dimitracopoulou, 2008)			X	
User classification indicators: posts reads relative to # posts per student (Bratitsis & Dimitracopoulou, 2008)			X	
User type messages per week (question, clarification, argument, disagreement, etc.) (Bratitsis & Dimitracopoulou, 2008)			X	

# Messages per participant per course phase (Davis et al., 2017; Petropoulou et al., 2012)			X	
Argumentation (message annotation) per course phase (Petropoulou et al., 2012)			X	
Avg. number of contributions per course phase (Petropoulou et al., 2012)	X	X		
Collaboration (interaction base message characterization) per course phase (Petropoulou et al., 2012)	X			
Participation count (number of posted messages) per course phase (Petropoulou et al., 2012)	X			
SNA: actor's degree centrality per course phase (Petropoulou et al., 2012)			X	
Work amount (message dimension per user) per course phase (Petropoulou et al., 2012)			X	
Learner isolation / students with limited connectivity (Bakharia & Dawson, 2011; Dawson et al., 2010)			X	
Social graph of student community (Bakharia & Dawson, 2011; Dawson et al., 2010)		X		
Social graphs of forum interaction data (Bakharia & Dawson, 2011; Dawson et al., 2010)		X		
Sociogram of interaction between teacher and participant (Bakharia & Dawson, 2011; Dawson et al., 2010)			X	
Frequency a student used each keyword (Mochizuki et al., 2005)			X	
Level of participation per student (Mochizuki et al., 2005)			X	
Relation between keywords and students (Mochizuki et al., 2005)		X		
# Messages read by the user in relation to # messages available (Brooks et al., 2006)			X	
Content currently read by one or more students (Brooks et al., 2006)		X		
Sociogram of participants, lurkers, and delinquents (Brooks et al., 2006)		X		
Contribution of each student to his/her group's online communication (Janssen et al., 2007)			X	
Final Grade (Score between 0 and 100) (Brouwer et al., 2016; Davis & Hauff, 2018; Kovanović et al., 2015; Pardo et al., 2016; van Leeuwen, 2016)			X	
Course Completion (pass fail, binary indicator) (Davis & Hauff, 2018)	X	X	X	
Video interactions (play, pause, fast forward, rewind, scrub) (Crossley et al., 2016; Davis & Hauff, 2018; Khan & Pardo, 2016; Kovanović et al., 2017; Li et al., 2015; Pardo et al., 2016; Pijeira Díaz et al., 2015; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Vitiello et al., 2017)	X	X		
Physical attendance (Nacu et al., 2018)			X	
Histogram Performance of peers for each key moment (Millecamp et al., 2018)			X	
Key moment with all courses of that moment (Millecamp et al., 2018)			X	
Peer performance for a course (Millecamp et al., 2018)		X		
Failed courses per student (Millecamp et al., 2018)			X	
Study trajectory of previous students with similar profile (Millecamp et al., 2018; Yousuf & Conlan, 2015)		X		
Study trajectories for different profiles (Millecamp et al., 2018)	X	X		
Individual student score and compare to scores of other students (Broos et al., 2018; Davis et al., 2016)		X		
Compare individual student score to scores from last year students (Broos et al., 2018)		X		
Tips based on the student score for improvement (Broos et al., 2018)			X	

Self-regulation behavior: Planning activities, and regulation activities (Cicchinelli et al., 2018)			X	
Total clicks per session (Brouwer et al., 2016; Cicchinelli et al., 2018; Hlosta et al., 2017; Vitiello et al., 2017)			X	
Average time monitoring per session (Cicchinelli et al., 2018; Hlosta et al., 2017)			X	
Average regulation per session (Cicchinelli et al., 2018)			X	
Average time monitoring (Cicchinelli et al., 2018)			X	
Average time until a student returns to the platform after each class (Cicchinelli et al., 2018)			X	
Mastery, the proportion of all exercises successfully answered in relation to the total number of exercises/quizzes (percentage) (Tempelaar et al., 2018)			X	
Learner watched video before submitting assignment (Boroujeni & Dillenbourg, 2018)		X	X	
Learning submitted assignment without watching videos (Boroujeni & Dillenbourg, 2018)		X	X	
Learner watched video without submitting assignment (Boroujeni & Dillenbourg, 2018)		X	X	
Learner did not watch video and did not submit assignment (Bakharia et al., 2016; Boroujeni & Dillenbourg, 2018)		X	X	
Correlation between quiz scores and number of events (clicks) (Shimada et al., 2018)		X	X	
Proportion of students who click each day in a course (Hlosta et al., 2017; Park et al., 2017)	X	X		
Average number of clicks per student and per course per day (Park et al., 2017)	X	X		
Percentage of number of students who increased or decreased in a course over time (Glahn et al., 2015; Park et al., 2017)	X	X		
Number of withdrawn students in days relative to the start of the course (Hlosta et al., 2017)	X	X		
Flag indicating student submission (yes/no) (Hlosta et al., 2017)			X	
Number of forum messages per day (Bakharia et al., 2016; Boroujeni et al., 2017; Davis et al., 2017; Kovanović et al., 2017)	X			
Number of forum messages after a video release in the learning resources (Boroujeni et al., 2017)	X			
Number of forum messages before an assignment deadline (Boroujeni et al., 2017)	X			
Number of specific keywords in posts per user in a course (Boroujeni et al., 2017)			X	
Number of new discussions per day (Boroujeni et al., 2017)	X			
Number of students/forum contributors per day in the discussions (Boroujeni et al., 2017)	X	X		
Mean time between successive forum writing events/messages (Boroujeni et al., 2017)	X			
Mean time between successive started threads (Boroujeni et al., 2017)	X	X		
Mean and standard deviation of threads count per day (Boroujeni et al., 2017)	X	X		
Mean and standard deviation of message counts per day (Boroujeni et al., 2017)	X	X		
Mean and standard deviation of forum contributors per day (Boroujeni et al., 2017)	X	X		
Percentage of lecture videos partially or totally watched in a course chapter (Bote-Lorenzo & Gómez-Sánchez, 2017; Crossley et al., 2016)	X			
Percentage of finger exercises answered in a course chapter (Bote-Lorenzo & Gómez-Sánchez, 2017)	X			
Percentage of assignments submitted in a course chapter (Bote-Lorenzo & Gómez-Sánchez, 2017; Crossley et al., 2016)	X			
Normalized grade of finger exercises in a given chapter (Bote-Lorenzo & Gómez-Sánchez, 2017)	X			

Normalized grade of assignment in a given chapter (Bote-Lorenzo & Gómez-Sánchez, 2017)	X			
Value of video engagement indicator (Bote-Lorenzo & Gómez-Sánchez, 2017; Yousuf & Conlan, 2015)		X		
Value of exercise engagement indicator (Bote-Lorenzo & Gómez-Sánchez, 2017; Yousuf & Conlan, 2015)		X		
Value of assignment engagement indicator (Bote-Lorenzo & Gómez-Sánchez, 2017; Yousuf & Conlan, 2015)		X		
Distinctive words from selected planning topics that predict course completion (Yeomans & Reich, 2017)			X	
Quiz submission timelines in days (Davis et al., 2017)	X			
Time spent by each user on the platform in hours (Bos & Brand-Gruwel, 2016; Davis et al., 2017; Jo et al., 2014; Muñoz-Merino et al., 2013; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014; Santos et al., 2013)		X	X	
Time spent by each user in watching videos in hours (Bos & Brand-Gruwel, 2016; Brouwer et al., 2016; Davis et al., 2017; Pijeira Díaz et al., 2015)		X	X	
Number of accessed videos per user (Davis et al., 2017; Muñoz-Merino et al., 2013; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014)		X	X	
Number of revisited videos per user (Davis et al., 2017; Pijeira Díaz et al., 2015)		X	X	
Number of forum visits (Crossley et al., 2016; Davis et al., 2017; Kovanović et al., 2015, 2017; Vitiello et al., 2017)		X	X	
Average length of a session (Davis et al., 2017; Khan & Pardo, 2016; Vitiello et al., 2017)	X			
Mean time between two sessions (Davis et al., 2017)	X	X		
Average Time between the first and last attempt for an assignment for a student (Crossley et al., 2016)		X	X	
Number of completed finger exercises/multiple choice questions by the video (Khan & Pardo, 2016; Pardo et al., 2016)		X	X	
Time spent on course assignments per user (Bogarín et al., 2014; Kovanović et al., 2015; Muñoz-Merino et al., 2013)		X		
Total time spent viewing course resources (Bogarín et al., 2014; Kovanović et al., 2015)		X		
Total time spent viewing course discussions (Kovanović et al., 2015)		X		
Total time spent posting in course discussions (Bogarín et al., 2014; Kovanović et al., 2015)		X	X	
Total time spent updating discussion messages (Bogarín et al., 2014; Kovanović et al., 2015)		X		
Total time spent on a question in a quiz giving a correct answer per user (Papamitsiou et al., 2014)		X	X	
Total time spent on a question in a quiz giving a wrong answer per user (Papamitsiou et al., 2014)		X	X	
How many times the student views a question in a quiz? (Papamitsiou et al., 2014)		X		
How many times the student changes the answer of a question in a quiz? (Papamitsiou et al., 2014)			X	
How much time the student spends on viewing a question in a quiz? (Papamitsiou et al., 2014)		X	X	
How much time the students spends on saving an answer to a question in a quiz? (Papamitsiou et al., 2014)			X	
How much time a student spends on quizzes? (Bogarín et al., 2014; Muñoz-Merino et al., 2013)		X	X	
How long the student waits in days to check new learning content counting from the time the learning content was available? (Bogarín et al., 2014)		X	X	
How long the student waits in days to check new quizzes or assignments counting from the time the learning content was available? (Bogarín et al., 2014)		X	X	
How many words have the students their forum posts?	X			

(Bogarín et al., 2014)				
How many sentences are there in the forum posts of the students? (Bogarín et al., 2014)	**X**			
Effective correct progress on the platform (Muñoz-Merino et al., 2013; Ruipérez-Valiente, Muñoz-Merino, & Kloos, 2014)		**X**		
Efficient correct progress on the learning platform (Muñoz-Merino et al., 2013)		**X**		
Total working schedule that shows at which time of the day the learner watches the lecture videos (Muñoz-Merino et al., 2013)	**X**	**X**		
Use of optional activities in a course? (Ruipérez-Valiente, Muñoz-Merino, Kloos, et al., 2014)	**X**			**X**
Ratio of accesses/clicks on weekdays vs weekend (Corrigan et al., 2015)	**X**			**X**
Ratio of on campus to off campus accesses/clicks (Corrigan et al., 2015)	**X**			**X**
Average quiz responses during the lecture and in the exam period (Glahn et al., 2015)	**X**			**X**
Student engagement comparison per activity (Yousuf & Conlan, 2015)		**X**	**X**	
Number of students who have achieved a badge (including the user student) (Charleer et al., 2016)			**X**	
How many times a student has been awarded a specific badge? (Charleer et al., 2016)		**X**	**X**	
Number of badges awarded on each day (Charleer et al., 2016)	**X**			
Number of awards given per badge (Charleer et al., 2016)	**X**			
Individual student grade and compare to grades of other students (Davis et al., 2016)		**X**	**X**	
Studying on certain hours of the day per user (Boroujeni et al., 2016)			**X**	
Studying on certain days of the week per user (Boroujeni et al., 2016)			**X**	
Studying on similar weekdays, over weeks of the course (Boroujeni et al., 2016)	**X**			
Time in seconds/minutes/hours between the submission time and the assignment deadline (Brouwer et al., 2016)	**X**			
Total clicks per day/week/month (Vitiello et al., 2017)	**X**			**X**

Implemented learning analytics indicators during the dissertation

Indicators from the list of identified indicators

Indicators from the list of identified indicators	Used in Evaluation
# Hits per student/user (weekly, daily, hourly)	X
# Session per student/user (weekly, daily, hourly)	X
Avg. number of sessions per week per student	
Avg. time per week per student	
# Distinct users	X
# Content pages viewed per student	
Avg. learning paths length/duration	
Avg. time spent online using the system	
Bar Chart: session Count, avg. pages/session, avg. time/session	X
Most frequent learning paths	
Table: Year, month, week, session count, avg. pages/session, avg. time/session (min)	Partially
Access to content pages by topic	
Course access by student per date	
Global accesses to the course	X
# Hits per resource (weekly, daily, hourly)	X
# Unique users per resource (weekly, daily, hourly)	X
Resources frequently used together (forum, mail, etc.) in each learning session	X
Daily content history/log (time, file, action)	
Downloaded learning resources (in a selected week)	
Resources students spent time with	X
Time-depended distribution of distinct users	X
Students' distribution for total time spent and resources used (bar chart)	X
Trends in students' activity (based on time spent)	
Collaboration (interaction base message characterization) per course phase	X
Participation count (number of posted messages) per course phase	X
Study trajectories for different profiles	
Number of forum messages per day	
Number of forum messages after a video release in the learning resources	X
Number of new discussions per day	
Percentage of lecture videos partially or totally watched in a course chapter	X
Total working schedule that shows at which time of the day the learner watches the lecture videos	
Ratio of accesses/clicks on weekdays vs weekend	X
Studying on similar weekdays, over weeks of the course	X
Total clicks per day/week/month	X
Top used resources	X

Implemented Indicators (self-defined)

Implemented Indicators NOT in existing indicators	Used in Evaluation
Platform Based	
Desktop vs. Mobile learning devices	X
Desktop vs. Mobile learning hits/accesses	X
Different Desktop devices for learning	X
Different mobile devices for learning	X
Different platform parts/modules involved in learning	X
Learning resources distribution and usage among the faculties	X
Learning resources distribution and usage among the faculties over time	X
Comparison of learning materials and literature among the faculties	X
Comparison of learning materials and hyperlinks among the faculties	X
Comparison of learning materials and use of videos and media among the faculties	X
Assessments modules distribution and usage among the faculties	X
Assessments modules distribution and usage among the faculties over time	X
Assignments use among the faculties	X
Gradebook use among the faculties	X
Exam results among the faculties	X
Collaboration modules distribution and usage among the faculties	X
Collaboration modules distribution and usage among the faculties over time	X
Shared Docs usage over time among faculties	X
Forums usage over time among faculties	X
Groups usage among faculties	X
Wikis usage among faculties	X
Information distribution modules distribution and usage among the faculties	X
Information distribution modules distribution and usage among the faculties over time	X
Comparison of Information distribution channels (announcements and emails) among the faculties	X
Desktop/Mobile/API learning over time and overall for a single faculty and semester	X
Most active courses for a single faculty and semester	X
Learning resources over time and overall for a single faculty and semester	X
Assessment use over time and overall for a single faculty and semester	X
Information distribution use over time and overall for a single faculty and semester	X
Collaboration modules use over time and overall for a single faculty and semester	X
Course Based	
Teaching activities and their influence on student activity in learning resources	X
Course distribution of learning resources used by the students	X
Information dissemination in relation to information access by the students	X
Video based learning in courses with lecture videos	X

Most watched/repeated videos in a course	X
Used videos in relation to formative assessment and assignment solutions	X
Assignments submission over time	X
Assignment submissions in relation to learning resources activities over time	X
Electronic tests activities (Moodle and Dynexite) in relation to learning resources activities over time	X

Indicators that can be implemented

With updated/extended analytics engine with the existing data set

Total time spent per lesson
Overall time spent per student (weekly, daily, hourly)
Learning paths analysis
Avg. visit duration
Delay among sessions per student
Frequency of usage of collaborative tools
Resources that have NOT been accessed (weekly, daily, hourly)
Time gap between two different readings of a specific message
Advice to the teacher concerning the most preferable types of learning objects visited by the students
Advice to the teacher concerning the status and behavior of the whole class
Students’ distribution for total time spent and resources used (bar chart)
Avg. number of contributions per course phase
Proportion of students who click each day in a course
Average number of clicks per student and per course per day
Percentage of number of students who increased or decreased in a course over time
Number of forum messages after a video release in the learning resources
Number of students/forum contributors per day in the discussions
Mean time between successive forum writing events/messages
Mean time between successive started threads
Mean and standard deviation of threads count per day
Mean and standard deviation of message counts per day
Mean and standard deviation of forum contributors per day
Average length of a session
Mean time between two sessions

With updated/extended analytics engine with updated anonymous dataset

Distribution of students posts in discussion forums
Revisits per lesson/quiz
A group of (related) lessons?
Avg. cognitive load per lesson/topic?
Avg. number of incorrect answers per question in a quiz
Avg. number of revisits per lesson/topic
Avg. overall quiz score

Avg. time spent per lesson/learning resource/quiz
Avg. unfinished visits of a quiz
Standard deviation of avg. time spent on a quiz
Standard deviation of overall quiz score
The most difficult questions in a quiz
Time-depended distribution of students' posts in discussion forums
All mistakes made by students
Concepts that were involved in the mistakes
Frequency of mistakes made
Distribution of most frequent misconceptions/mistakes (ranking)
Assignments submitted per student
Assignments read per student
Attempts in item per student
Chat rooms entered per student
Value of a lesson in relation to learning objectives?
Participants per group
Forums participated in per group
Advice to the teacher concerning teaching parts of the course causing problems to the majority of students
Bookmarked learning resources (in a selected week)
Avg. number of posts per group
Avg. number of posts per student
Avg. thread depths/weight
Video interactions (play, pause, fast forward, rewind, scrub)
Number of withdrawn students in days relative to the start of the course
Number of forum messages before an assignment deadline
Percentage of finger exercises answered in a course chapter
Percentage of assignments submitted in a course chapter
normalized grade of finger exercises in a given chapter
normalized grade of assignment in a given chapter
Quiz submission timelines in days
How many words have the students their forum posts?
How many sentences are there in the forum posts of the students?
Use of optional activities in a course?
Ratio of on campus to off campus accesses/clicks
Average quiz responses during the lecture and in the exam period
Number of badges awarded on each day
Number of awards given per badge
Time in seconds/minutes/hours between the submission time and the assignment deadline

Indicators that need pseudomized/personal data and different analytics engine

Students’ comprehension of the studied topics (based on his/her annotations)
Students’ tags for the selected lesson group
Mistakes made by students in exercises that were never finished
Mistakes that often come together (a-priori): if students make mistake A, followed by mistake B, then later they make mistake C
Individual students’ performance (grade) per quiz/exercise/topic/item
Overall performance of the best X students (with more than Y exercises) in a specified time frame
Overall performance of the best X students (with more than Y exercises) in a specified time frame including only students with specific scores
Overall performance of the worst X students (with more than Y exercises) in a specified time frame
Overall performance of the worst X students (with more than Y exercises) in a specified time frame including only students with specific scores
Messages post to forum per student
Messages read on forum per student
Messages replied to per student
Student age
Student gender
Items in test per student
Messages read on mail per student
Quizzes done per student
Quizzes failed per student
Quizzes passed per student
Wiki pages edited per student
Student academic level
Threads started per student
Quiz and assignment submissions per student
Matrix on students’ performance on quizzes related to domain (Student, concept and level of knowledge)
Progress with the course schedule per student
Time, discussion topic, student, number of follow-ups
Student risk group/status
Avg. hits per user in a group
Avg. session per users of each group
Posts per student compared to group average (weekly, daily, hourly)
Students who have NOT accessed a specific resource (weekly, daily, hourly)
Viewed resources per student (weekly, daily, hourly)
Messages posted per group
Messages read per group
Messages replied to per group
Assignments per group
Files downloaded per group
Files downloaded per student

Files exchanged per group
Files exchanged per student
Files per group
Files uploaded per group
Files uploaded per student
Forums participated in per student
Messages quoted per group
Messages quoted per student
Threads started per group
Connection frequency per group
Connection frequency per student
Message has been displayed, read partially or entirely?
Time spent by each user on reading a message
Advice because of the student’s progress / delay in time
Advice because student has (not) completely mastered the related prerequisite concepts
Advice because student has (not) completely read and worked on the learning objects and assessment quizzes related to the concept
Advice because student has (not) participated in the communication activities related to the concept
Advice to groups concerning common problems
Advice to groups concerning satisfactory/unsatisfactory learning levels
Advice to groups concerning uncommunicative behavior
Advice to groups related to the knowledge states of groups
Advice to students who have not started working on the course
Advice to students who have unsatisfactory learning levels because they are weak and normally communicative
Advice to students who have unsatisfactory learning levels because they are weak and uncommunicative
Advice to students who have unsatisfactory learning levels because they are weak but highly communicative
Advice to the teacher concerning excellent and weak students relative to the whole class
Advice to the teacher concerning least and most active students relative to the class
Advice to the teacher concerning least and most communicative students relative to the class
Advice to uncommunicative advanced students to help others
Goal: prediction of final mark according to student’s activity
Patterns of interactions between specific participants
Relation between contributions?
Resources watched (read) in a session per student (via webcam)
Keyword analysis derived from email exchange data per student
Social network derived from email exchange data per student
Students’ tag cloud of learning resources
Students’ tag cloud of learning resources in a selected past week
Topics of interest

Combination of (parallel coordinates) total time spent on the course, avg. time spent on a resource, number of resources used, and median of the time of day a student works
Answers to user by others per student
Follow-up contributions per group
Type t posts per group
Type t posts per student
Activity indicator regarding # posts, # posts reads, # type per group
Activity indicator regarding # posts, posts reads, type per student
Contribution indicator per group
Contribution indicator per student
Group interactivity indicator
Relative activity regarding # posts, # types, # initiated threads per group
Relative activity regarding # posts, # types, # initiated threads per student
Relative posts read to number of posts per group
Social Network Analysis
Thread Depths/weight per group
Thread Depths/weight per student
Tree structure of the forum, highlighting the messages of a corresponding student
User classification indicators: posts reads relative to # posts per student
User type messages per week (question, clarification, argument, disagreement, etc.)
Messages per participant per course phase
Argumentation (message annotation) per course phase
SNA: actor’s degree centrality per course phase
Work amount (message dimension per user) per course phase
Learner isolation / students with limited connectivity
Social graph of student community
Social graphs of forum interaction data
Sociogram of interaction between teacher and participant
Frequency a student used each keyword
Level of participation per student
Relation between keywords and students
Messages read by the user in relation to # messages available
Content currently read by one or more students
Sociogram of participants, lurkers, and delinquents
Contribution of each student to his/her group’s online communication
Final Grade (Score between 0 and 100)
Course Completion (pass fail, binary indicator)
Physical attendance
Histogram Performance of peers for each key moment
Key moment with all courses of that moment
Peer performance for a course
Failed courses per student
Study trajectory of previous students with similar profile
Individual student score and compare to scores of other students
Compare individual student score to scores from last year students

Tips based on the student score for improvement
Self-regulation behavior: Planning activities, and regulation activities
Total clicks per session
Average time monitoring per session
Average regulation per session
Average time monitoring
Average time until a student returns to the platform after each class
Mastery, the proportion of all exercises successfully answered in relation to the total number of exercises/quizzes (percentage)
Learner watched video before submitting assignment
Learning submitted assignment without watching videos
Learner watched video without submitting assignment
Learner did not watch video and did not submit assignment
Correlation between quiz scores and number of events (clicks)
Flag indicating student submission (yes/no)
Number of specific keywords in posts per user in a course
Value of video engagement indicator
Value of exercise engagement indicator
Value of assignment engagement indicator
Distinctive words from selected planning topics that predict course completion
Time spent by each user on the platform in hours
Time spent by each user in watching videos in hours
Number of accessed videos per user
Number of revisited videos per user
Average Time between the first and last attempt for an assignment for a student
Number of completed finger exercises/multiple choice questions by the video
Time spent on course assignments per user
Total time spent viewing course resources per user
Total time spent viewing course discussions per user
Total time spent posting in course discussions per user
Total time spent updating discussion messages per user
Total time spent on a question in a quiz giving a correct answer per user
Total time spent on a question in a quiz giving a wrong answer per user
How many times the student views a question in a quiz?
How many times the student changes the answer of a question in a quiz?
How much time the student spends on viewing a question in a quiz?
How much time the students spends on saving an answer to a question in a quiz?
How much time a student spends on quizzes?
How long the student waits in days to check new learning content counting from the time the learning content was available?
How long the student waits in days to check new quizzes or assignments counting from the time the learning content was available?
Effective correct progress on the platform
Efficient correct progress on the learning platform
Student engagement comparison per activity

Number of students who have achieved a badge (including the user student)
How many times a student has been awarded a specific badge?
Individual student grade and compare to grades of other students
Studying on certain hours of the day per user
Studying on certain days of the week per user

Personas (Gospodinova, 2018; Mentiu, 2018)

NAME Elisabeth van der Rate

DESCRIPTOR Professor, active, confident, focused, carrying

QUOTE

"I am experienced, successful and active university administrator and teaching professor. I like simple, well-understood, trustful, business-oriented data, which can be used easily and efficiently. I take my job concerns personally, since I am a mother of two students and care about their success."

WHO IS IT?

- 47 years, female, Dutch, married
- Professor of Economics and Finances in Maastricht University
- Part of Committee of Resources Allocation for Improvement and Innovation in Education
- Parent of two students

WHAT GOALS?

Observe and improve:

- Success rate of students
- Usage and costs of E-learning and its influence
- Learning process of my own children

WHAT ATTITUDE?

Tool should:

- Present good statistical overview on all major studying information
- Be easy to understand
- Provide presentative graphical data
- Enable easy export of the data to different types

WHICH BEHAVIOUR?

Elisabeth is a professor in three full lectures and is in cooperation with other professors from RWTH Aachen University in two more lectures.

She participates in monthly committee sessions, but she also needs to preform administrational tasks weekly.

She is frustrated by 'bad' graphical design and by complicated GUI.

She wants variety of simple and representative graphic types.

It is very important for Elisabeth that data is very accurate, since based on that she does administrational decisions, which impact whole university.

She does not like dealing with errors and their frequent appearance will make her lose trust in the tool and stop her from using it.

For Elisabeth, emotional and expressive benefits are important in the tool.

She is slow decision maker and makes them based on both facts and emotions.

NAME John Datov

DESCRIPTOR PhD, serious, hard-working, goal-oriented, innovative

QUOTE

"I am a busy, hard-working person, always trying to improve, find new solutions and discover new ideas. I like to use tools helping me in fast and accurate manner, to track and monitor the teaching processes I facilitate"

WHO IS IT?

- 33 years, male, German, single
- Teaching assistant in Mechanical Engineering classes, tutor, lecturer
- Lives in Aachen

WHAT GOALS?

- Scientific/working growth
- Improvement of learning processes
- Accurate, structured data about materials usage
- "Easy-to-grasp" summarized information about students and their progress
- Engagement of more students and enhancement of their progress

WHAT ATTITUDE?

Tool should:

- Reduce over-head by pointing out errors and weak points in my job with materials
- Highlight correlation between my input and effect on students
- Be easy to operate, responsive and accurate

WHICH BEHAVIOUR?

John is a teaching assistant in three mechanical engineering courses for bachelors.

He teaches only lectures in one course and in two others he is also responsible for practical exercises. His classes are very big – he does not have an opportunity to track personal progresses of students.

He is responsible for many materials of various types: books, slides, links, videos etc.

Managing learning materials takes big part of his time, thus, he wants to see how they are being used: by whom, when, for how long etc.

John has lack of time and does not want to spend much time in setting up the tool or waiting for its responses. He wants up-to-date information from the tool, which is constantly present.

He is always searching for new technologies and holds his hand on the pulse of his field.

He likes to experiment with new teaching approaches and he needs feedback on that from the tool.

For John functional benefits are most important, but expressiveness is relevant too.

NAME Ravi Infra Stracturi

DESCRIPTOR Introvert, IT specialist, gamer, GUI enthusiast

QUOTE

"I am a technical geek, I like attractive interfaces and employing gaming parts into learning tools. I develop from different platforms. I like to engage user in developing process and to follow Agile processes."

WHO IS IT?

- 28 years, male, Indian, married, no kids
- Developer in E-learning field
- Active gamer

WHAT GOALS?

- Develop usable, attractive, interesting, modern and cross-platform system
- Gain good feedback from users
- Establish low error rate

WHAT ATTITUDE?

Tool should:

- Be flexible and maintainable
- Have simple architecture
- Be responsive and engaging
- Have a user-friendly GUI

WHICH BEHAVIOUR?

Ravi works on already designed system by modifying it in following ways:

- Writing new functionalities
- Debugging errors
- Responding to user requests for changes

To do that he needs access to following information:

- Which platforms and data types users use
- Error rates and problem-prone parts of the system
- E-learning deployment status by departments of the university to form priorities for responding to new requests

Ravi is interested in functional benefits from the tool.

He is fast decision maker and makes them based on facts.

NAME Johannes Chillermann

DESCRIPTOR Student, communicative, struggling learner

QUOTE

"Hey guys, I am a relaxed easy-going person, for whom lots of studying is hard and boring. I like interactive activities and to hang out with people. Biology is in my interests – I want to invent something interesting to help people and that is why I need to become a better student"

WHO IS IT?

- 19 years, male, German, single
- Bachelor of Biology in 1st year
- Lives in Aachen
- Skater

WHAT GOALS?

- Improve my learning
- See:
  - where I am and where are others and what should I do to perform better
  - Success rates of courses to pick a course I can handle
  - Time load of courses so I can see how much I need to work to pass
  - The minimal amount of materials to pass

WHAT ATTITUDE?

Tool should:

- Have innovative and interactive look
- Tips and tricks on how to perform better
- Reminders when I am not working enough
- Engagement of others in E-learning
- My timely and progress feedback
- Features to promote collaboration and teamwork
- Information about support: whom can I ask to help

WHICH BEHAVIOUR?

Johannes has finished his first semester of studies and it was not very successful because he had problems with motivation, finding out classes structures and requirements.

Now, in the beginning of his second semester he decides to use the tool to improve.

He seeks different levels of motivation, support and collaboration, since he saw that other people did better than him.

He loses his attention fast, thus, he needs various engaging tasks and learning materials to keep studying.

He would appreciate some gaming features with scoring levels and boards.

He needs his profile to be simple, easy to understand with fresh and modern GUI.

Johannes wants to see how much work needed for which level of marks and needs constant encouragement.

For him, expressive and emotional benefits are important in the tool.

He is slow decision maker and makes them based on emotions.

NAME Lili Tuto

DESCRIPTOR Student, dedicated, sporty, passionate, friendly

QUOTE

"I am an advanced student, who studies a lot to excel in my field. I like to help people to improve their performance and to work in teams. I like Big Data and not afraid of complicated statistics."

WHO IS IT?

- 26 years, male, Chinese, single
- Master student of Informatics
- Assignment tutor for mathematics and programming
- Chess player

WHAT GOALS?

- Maintain my good level of learning
- See:
  - My progress and rates
  - Ranking of materials importance
  - Attending rates and success of others
- Gain feedback on my tutoring: effect on students
- Problems of students with E-learning

WHAT ATTITUDE?

Tool should:

- Provide detailed and thorough statistics from my goals
- Enable to configure advanced settings, so I can se up my profile
- Provide both course and group overview for my tutoring
- Have modern look and feel

WHICH BEHAVIOUR?

Lili is an active student: has good GPA, studies a lot, tutors mathematics and basics of programming.

She needs detailed information about her success and progress to maintain her excellent studying.

She needs a lot of statistics about other people and courses that have math and programming, so she can see where her tutoring help can be needed.

She wants to see what question people ask, how they attend the classes, what materials and when they are accessing.

Lili needs to know timings of people using the tool, so she knows when to attend to it in order to have optimal coverage on questions and improve her tutoring.

She wants to have feedback on her tutoring: how success and participation rate of her students changes.

She knows how a "good" system looks and feels like, so lots of bugs and inconsistencies will frustrate her and probably cause to stop using the tool.

She expects both functional and expressive benefits from the tool.

She is a fast decision maker, and makes them based on both fact and emotions.

Remaining use cases for the personas:

Elisabeth van der Rate use case 2
Problem: A part of Elisabeth job is to participate in resource allocation for E-learning development in the university. On next budget meeting the committee needs to decide to which department to give new video equipment. For this purposes Elisabeth needs to make a presentation comparing the video-enhanced learning deployed by different departments.

Solution: In the tool Elisabeth can easily select the department she wants to compare and find the resources usage indicators, where she can select all video types' files and generate comprehensible comparison visualization.

Result: On the meeting using the output from the tool, the committee can decide fast and accurate to which department to allocate the new equipment.

John Datov - Use Case 2
Problem: John has invited a guest lecturer from another university to talk about a "state-of-the-art" topic. He has posted the announcement in discussion boards of few courses. He wants to be sure that there are going to be enough people at the lecture.

Solution: John opens the tool and can see how many people have seen the announcement and what discussions it has led to. If he sees that there are too few people, he can send additional e-mails and remind students at the next course activities.

Result: On the day of the guest lecture the auditorium is full and students have heard a lot of interesting and motivating information.

Ravi Infra Stracturi – Use Case 1
Problem: Ravi received two "bug" reports from two different places in the system; he now wants to know what the priorities to fix them are.

Solution: Ravi opens the tool and can find out following information about the "bugs":

a. Number of users using the affected system part

b. Dependencies of the affected system part

And he can compare the values for the two reported "bugs".
Result: Ravi has discovered that one bug affects a big course with more than 1000 students, and another bug affects only 100 lab students. The priority is decided, and Ravi knows what to start working on.

Ravi Infra Stracturi – Use Case 2
Problem: Ravi needs to present to his project manager the current state of the e-learning system and its usage.

Solution: Ravi opens the tool and can find following information:

a. Platforms in use

b. Devices types

c. Types of files used in the system

d. Operating Systems

e. Browsers

f. Error rates correlated to a.-e.

g. Most active timings of system being accessed

Tool is fast and responsive when producing the information, since Ravi makes the live demo of the current search and representation of the indicators.
Result: The project manager is updated about current state and impressed by the usability and responsiveness of the tool.

Johannes Chillermann - Use Case 1

Problem: Johannes is sad since last semester he did not pass few subjects and those he passed did not have a high score. Now it is a beginning of new semester and he wants to select subjects that he can handle. For that he needs additional information.

Solution: Johannes opens the tool and can easily find a list of subjects from his major field and find following information:

a) Subject's load (per week/ month/ semester).

b) Average marks and passing rates.

He can easily compare subjects by that information. He can also filter the subjects by:

a) Minimum/maximum average mark.

b) Threshold for load.

Result: Using the provided information, Johannes has decided on the subject to pick for the next semester and is more confident about his upcoming performance.

Lili Tuto - Use Case 1

Problem: Lili is a responsible tutor and she wants to know how her students are performing and if there is anything she needs to improve in her tutoring activities. She also wants to know what the most usable timings of the day are by her students, so she knows when it is better to post important materials and messages to get the most attention.

Solution: Lili opens the tool and can easily find following:

a) Overall performance of her group compared to another groups/whole course.

b) Problems and, mistakes in last e-test done by her group compared to another groups/whole course.

c) Timing of her group using the system.

d) Posts of her group in correlation to her tutoring activities.

e) List of processed materials by her group.

Result: Lili has seen that her group is doing well, but that she needs to remind them to read new lecture and the best time to post it is from 2 till 4pm.

Use Case for Johannes Chillermann and Lili Tuto

Problem: A student is in the middle of his/her semester and wants to know how he/she is performing by now.

Solution: The student can open the tool and find following information in the personal view:

a) Numbers of assignments (points, percent) left in the course.

b) My current state in the course, compared to other students and to average passing paths.

c) List of most popular materials for the course and approximate time needed to process them.

Result: The student now knows how much more work is needed in the course and the current learning progress. The new interesting learning materials have been discovered as well.

List of publications

Lukarov, V. and Schroeder, U. 2018. Learning analytics case study in blended learning scenarios. CEUR Workshop Proceedings (2018).

Mendez, C., Lukarov, V., Greven, C., Valdez, A.C., Dietze, F., Schroeder, U. and Ziefle, M. 2017. User groups and different levels of control in recommender systems. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2017).

Lukarov, V. and Schroeder, U. 2017. Analytics Prototype for Data Driven Decision Making for Blended Learning Strategies in HEI. Bildungsräume 2017. (2017).

Lukarov, V. and Schroeder, U. 2017. AiX-Analytics: Analytics Tool at {RWTH} Aachen University. Bildungsräume 2017: DeLFI 2017, Die 15. e-Learning Fachtagung Informatik, der Gesellschaft für Informatik e.V. (GI), 5. bis 8. September 2017, Chemnitz (2017), 393–394.

Lukarov, V. and Schroeder, U. 2017. AiX-Analytics: Analytics Tool at RWTH Aachen University. (2017), 393–394.

Lukarov, V., Chatti, M.A. and Schroeder, U. 2015. Learning analytics evaluation -beyond usability. CEUR Workshop Proceedings (2015).

Lukarov, V., Chatti, M.A., Thüs, H., Kia, F.S., Muslim, A., Greven, C. and Schroeder, U. 2014. Data Models in Learning Analytics. (2014), 88–95.

Chatti, M.A., Lukarov, V., Thüs, H., Muslim, A., Mohamed, A., Yousef, F.A.M., Wahid, U., Greven, C., Chakrabarti, A. and Schroeder, U. 2014. Learning Analytics: Challenges and Future Research Directions. E-learning and Education (Eleed). (2014).

Dyckhoff, A.L., Lukarov, V., Chatti, M.A. and Schroeder, U. 2013. Welchen Einfluss hat ein Learning Analytics Tool auf das Verhalten der Lehrenden? DeLFI 2013. (2013).

Dyckhoff, A.L., Lukarov, V., Muslim, A., Chatti, M. a. and Schroeder, U. 2013. Supporting action research with learning analytics. Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK ’13. (2013), 220.

Lukarov, V. 2013. Learning Analytics Tools for raising Awareness and Reflection.