

Learners in the Data Mist: Visualizations for Optimization of Learning Environments

Sigve Solvaag

Supervisor

Professor Barbara Wasson



Master thesis for the degree of Master in Information Science
at the Department of Information Science and Media Studies,
University of Bergen

2018

"From each according to his ability, to each according to his needs"

- **Karl Marx** (1818 - 1883), Critique of the Gotha Program

"Has Anyone Really Been Far Even as Decided to Use Even Go Want to do Look More Like?"

- **Anonymous** (2009)

Abstract

This master thesis presents an Learning Analytics artifact designed to support learning environment optimization. The Design Science research project developed an artifact through iterative design processes that were informed by both quantitative and qualitative evaluation methods. Through five iterations, ranging from a proof of concept artifact to a functional artifact that could serve visualization data through a REST API application. The case for this research project was a SPOC course managed by OsloMet, which offered their data as a case study to examine it with Learning Analytics processes.

Each iteration is detailed in its own chapters, with sections labeled according to Dresch, Lacerda and Antunes Jr's proposal for a Design Science research model. Each evaluation gathered data using a variety of evaluation methods to ensure multiple perspectives on the visualizations and the artifact's performance factors. Especially the Stakeholders of the SPOC course were involved in semi-structured interviews to obtain functional requirements and establish user needs.

The research project eventually produced a functional artifact, which we could use to test whether it was capable of supporting learning environment optimization. We found that the artifact had promising potential, but would be better evaluated through testing over several course runs and with the consumption of multiple data-sets.

Acknowledgements

I would like to thank my supervisor Barbara Wasson for her expertise and wisdom, not to mention incredible patience for putting up with someone as challenging as me.

I would also like to thank Ann-Karin Valle and Weiqin Chen for their help, insights and fruitful collaboration. Without them, this thesis would not have worked.

I am eternally grateful for the unconditional support both Ankica Babic and Marija Slavkovic provided throughout the time of my master's degree. I'm sure chocolate will find its way to their offices in due time.

I would like to thank my cohort of living spell-checkers; Bjørn Helge Sandblåst, Candida Sylta, Einar Søreide Johansen, Jonas Svendsen, Julia Beate Bådsvik, Ole Andreas Morvik, Sami Mobarak-Sadokierski and Simen Karlsen. You relieved a very stressed out student. I'm very grateful and I totally owe you all a favor or two. A big thanks to the participants of the research - your enthusiasm was inspiring.

A special thanks to Hedda Paulsen who made coffee every morning, shared laughs, beers and moments, big and small.

At last I would like to thank my wife Holly and my sons Victor and Theodore, I love you all. Without you, I would not have the love, inspiration, support or motivation to pull through this journey.

| | |
|--|-----------|
| Abstract | 2 |
| Acknowledgements | 2 |
| 1 Introduction | 7 |
| 1.1 Motivation | 7 |
| 1.2 Research Question | 8 |
| 2 Preliminaries | 9 |
| 2.1 Massive Open Online Courses | 9 |
| 2.2 Small Private Online Courses | 10 |
| 2.3 Learning Analytics | 11 |
| 2.4 MOOCs and Learning Dashboards | 13 |
| 2.5 Information Visualization | 14 |
| 3 Related Work | 16 |
| 3.1 Learning Analytics Artifacts | 16 |
| 3.2 Data-driven Improvement of Learning Environments | 17 |
| 4 Methodology | 19 |
| 4.1 Design Science Research | 19 |
| 4.1.1 Artifact | 22 |
| 4.1.2 Class of Problems | 22 |
| 4.1.3 Systemic Thinking | 23 |
| 4.1.4 System Usability Scale | 23 |
| 4.1.5 Use Case Evaluation | 24 |
| 4.1.5 Likert scale | 24 |
| 4.1.6 Semi-structured interviews | 24 |
| 5 First Iteration | 25 |
| 5.1 Identification of the Problem | 25 |
| 5.2 Awareness of the Problem | 26 |
| 5.3 Literature Review | 31 |
| 5.4 Identification of the artifacts and configuration of the classes of problems | 32 |
| 5.5 Proposition of artifacts | 33 |
| 5.6 Design & Development of the Artifact: | 33 |
| 5.7 Evaluation of the Artifact: | 34 |
| 6 Second Iteration | 35 |
| 6.1 Identification of the Problem | 35 |
| 6.2 Awareness of the Problem | 36 |
| 6.3 Proposition of Artifact: | 36 |
| 6.4 Design of the selected artifact | 37 |
| 6.5 Development of the Artifact: | 40 |
| 6.6 Evaluation of the Artifact | 41 |

| | |
|---|-----------|
| 6.6.1 Use Case Evaluations Results | 42 |
| 6.6.2 System Usability Scale Results | 45 |
| 6.6.3 System Performance tests: | 46 |
| 6.6.4 Heuristics Extraction: | 48 |
| 7 Third Iteration | 51 |
| 7.1 Identification of the Problem | 51 |
| 7.2 Awareness of the Problem | 52 |
| 7.2.1 Assessment of the Data | 53 |
| 7.3 Proposition of Artifact | 55 |
| 7.4 Design of the selected Artifact | 55 |
| 7.5 Development of the Artifact | 56 |
| 7.6 Evaluation of the Artifact | 59 |
| 7.6.1 System Usability Scale results | 59 |
| 7.6.2 System Performance Results | 61 |
| Feedback from Stakeholders | 62 |
| 7.6.3 Heuristics Extraction: | 62 |
| 8 Fourth Iteration | 64 |
| 8.1 Awareness of the Problem | 64 |
| 8.2 Proposition of Artifact | 65 |
| 8.3 Design of the selected Artifact | 65 |
| 8.4 Development of the Artifact | 66 |
| 8.5 Evaluation of the Artifact | 69 |
| 8.5.1 System Usability Scale results | 70 |
| 8.5.2 System Performance tests | 74 |
| 8.5.3 Generalization to Heuristics | 75 |
| 9 Fifth Iteration | 76 |
| 9.1 Awareness of the Problem | 76 |
| 9.2 Proposition of Artifact | 77 |
| 9.3 Design of the selected Artifact | 78 |
| 9.4 Development of the Artifact | 78 |
| 9.5 Evaluation of the Artifact | 85 |
| 9.5.1 System Usability Scale survey | 85 |
| 9.5.2 Likert survey | 86 |
| 9.5.3 Feedback from Stakeholders | 88 |
| 9.5.4 Heuristics Extraction | 88 |
| 10 Discussion | 90 |
| 10.1 Findings through application of Design Science | 90 |
| 10.1.1 Constraints of Policy | 90 |
| 10.1.2 Constraints of Infrastructure | 90 |
| 10.1.3 Constraints of Evaluation | 91 |

| | |
|---|------------|
| 10.1.4 Outcomes of Heuristics | 92 |
| 10.2 Response to the Research Question | 93 |
| 11 Conclusions | 96 |
| 11.1 Summary | 96 |
| 11.2 Future Work | 97 |
| 12 References | 98 |
| Appendix A - Likert Scale Survey | 101 |
| Appendix B - Semi-Structured Interview Guide | 103 |

1 Introduction

1.1 Motivation

Modern education in the Western world is increasingly moving into virtual learning environments, with the use of computers and advanced electronics being incorporated directly into the learning environment for learners of all ages. Many of these systems have been the norm for decades, while novelty educational technologies such as Khan Academy and Brilliant.org have appeared in recent years. While the focus of many such systems is to facilitate learning whether by organizing schedules, serving files or automated instruction, a secondary benefit to such systems is the capability to track user activity.

With the fidelity of web analytics, the emergence of Internet of Things and the proliferation of sensors in devices such as cellphones, being able to observe learners in their environment increasingly more possible. However, the wealth of data alone does not equal information nor wisdom. A challenge appears in the process of simply making sense of the data that we can produce and store.

A field which has emerged in the recent years is Learning Analytics, where the First International Learning Analytics and Knowledge conference was held in 2011 [22], and according to Ferguson[11], the field of Learning Analytics split from academic analytics in 2010. Learning Analytics is best understood as a cycle that begins with the collection of learner data, then the analysis of said data, which is then presented and acted upon by the teachers for the benefit of the learner and the environment in which they learn.

An organization that has a stated interest in Learning Analytics is the research centre for Science of Learning & Technology (abbr: SLATE), which is hosted by the Faculty of Psychology at the University of Bergen. Since its launch, SLATE has hosted several master students from the Department of Information and Media Science. This master's thesis is part of their research on a Small Private Online Course hosted by OsloMet that took place in Fall 2017.

1.2 Research Question

In this thesis we pose the following research question:

RQ1: *How can learner data be visualized to support teaching environment improvement?*

By answering the above question, we hope to contribute to the science of Learning Analytics to optimize the learning environment.

2 Preliminaries

2.1 Massive Open Online Courses

Massive Open Online Courses are courses that are distributed over the internet and sets no upper limit of participation. MOOCs build upon the teaching traditions of distance education that date back to the 1800s, as pioneered by Sir Isaac Pitman. [35] The practise of educating students over remote communications was then limited to postal correspondence, but with the emergence of newer technologies, the practise changed accordingly. [35]

Especially in the 1990s, the proliferation of free and open source movements began to influence thought in other parts of society. [38] With the opening of the MIT OpenCourseWare, the movement of Open Educational Resources started in the 2000s [33], later recognized by UNESCO in 2002 during the Forum on the Impact of Open Courseware for Higher Education in Developing Countries [20]. The stage was being set for MOOCs when syllabus became available in this way, along with research indicating a separation of learning outcomes and class sizes [3]. Later in the 2000s saw the first instances of MOOCs appear, pivotally with the release of the “Connectivism and Connective Knowledge” or the CCK08 course [12].

Several implementations of MOOCs appeared since then, but there are two clear paradigms within the MOOC world. The first is the cMOOC, which emphasizes an aggregation of syllabus and information, while the latter is the xMOOC, which follow a more traditional course structure [26]. Presently, most xMOOCs are openly available for enrollment on web-sites such as edX, FutureLearn and Udemy, usually featuring videos, discussion forums, quizzes and interactive modules as part of learning process.

Given the 'massive' part in its name, MOOCs are structured in such a way that the volume of students is scalable, but also manageable with a small staff of teachers. For this reason, there are several techniques in place to manage the volume of students, such as graded quizzes, automated assessment, peer-review and group collaboration.

2.2 Small Private Online Courses

Small Private Online Courses, or SPOCs first appeared in 2013, the term coined by Professor Armando Fox [13]. Although MOOCs are open for the public, the software they run on are also possible to adapt and gate for smaller scale teaching environments.

A common reason for creating SPOCs is to facilitate flipped classrooms Burge *et al* [4], where the "lecture" is presented as a video and made available on an SPOC module, while the classroom itself is used for increased personal interaction and tutoring. In this sense, SPOCs can be seen as courses that have been amplified with the toolkit of MOOCs, as opposed to MOOCs being used in a limited capacity.

Another reason for creating SPOCs is to facilitate blended learning programmes, which are to a degree more strict than flipped classrooms. While flipped classrooms do not necessarily require a student's presence, blended learning courses require a degree of student presence for certain parts of the course. [25]

2.3 Learning Analytics

Learning Analytics is a broad field of research which emerged from disciplines such as Educational Data Mining and uses methods found in Web Analytics. The promise of Learning Analytics is that it can create a feedback loop between a learner, teacher and the environment in which they act. Learning Analytics is defined as:

“the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.” [31]

This description is the one this research project works with.

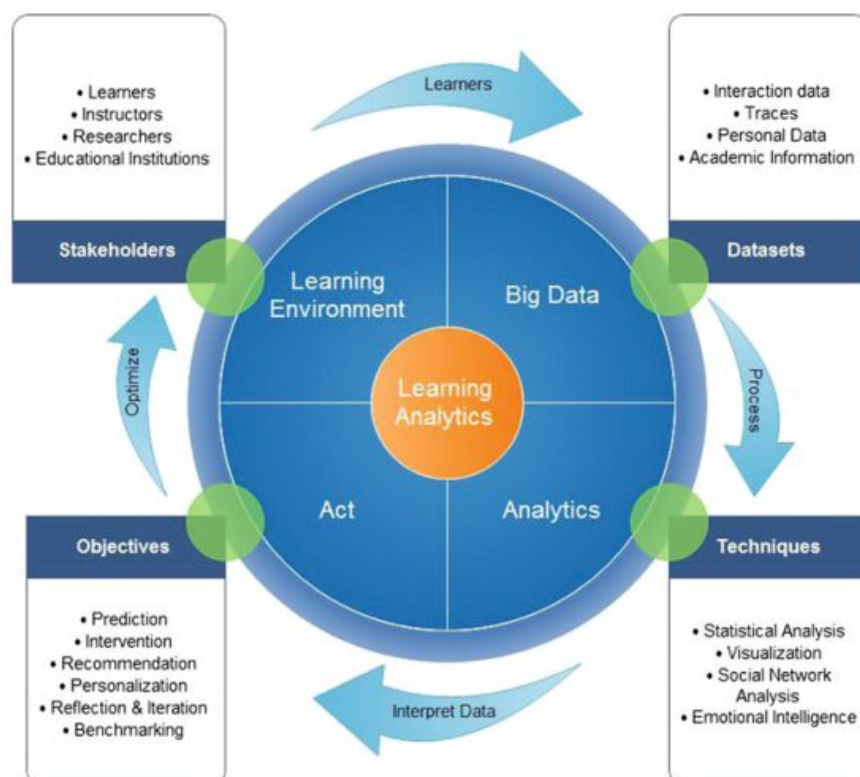


Figure 2.1 Khalil and Ebner's Learning Analytics life cycle [21].

Khalil and Ebner's Learning Analytics life cycle [21] describes a process in which the learners produce data, which is in turn processed with different techniques and then interpreted into actions. Finally these actions are used for optimization in the

learning environment. Another model for a Learning Analytics cycle was described by Clow [7] in figure 2.2:

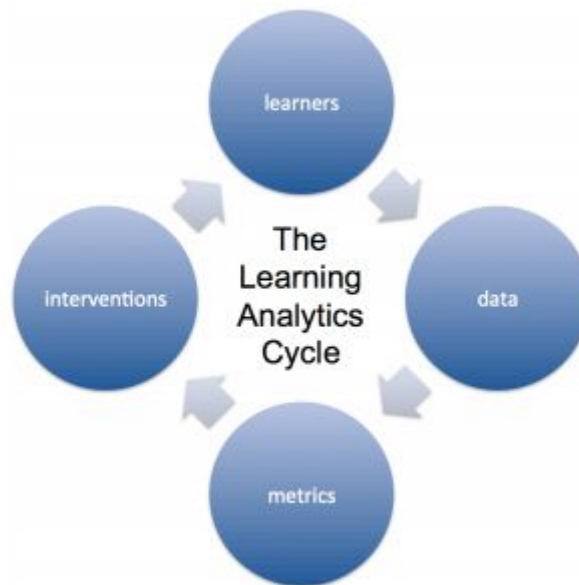


Figure 2.2 Clow's Learning Analytics Cycle.

The information produced by Learning Analytics is ostensibly made for the benefit of the learner [6], but it can be used to reflect on the learning environment as well. Learning Analytics employs methods such as statistics, web analytics, data mining, artificial intelligence, operational research, business intelligence, social network analysis and information visualization, with an increasing interest in technologies that can provide with more empirical data [8]. While Learning Analytics employ a broad array of methods, this also introduces a demand for data or information interoperability, given that the trace data being studied do not necessarily subscribe to the same model [11].

2.4 MOOCs and Learning Dashboards

Learning Dashboards were suggested by Schwendimann *et al* [29] to be defined as:

“A learning dashboard is a single display that aggregates different indicators about learner(s), learning process(es) and/or learning context(s) into one or multiple visualizations.”

This definition is applicable to both Learning Analytics and Educational Data Mining, and for the context of our research we use the definition for Learning Analytics. Schwendimann *et al* [29] found several synonymous terms for Learning Dashboards in their systematic literature review, but common for them all is they use information visualization to represent learner data.

Learning Dashboards can be integrated directly into the virtual learning environment, as with Insights and ANALYZE. [15] Since Schwendimann *et al* [29] studied over 55 works on Learning Dashboards, this mode of conveying Learning Analytics is an emerging trend for virtual learning environments such as MOOCs.

2.5 Information Visualization

Infographics has a history dating back to 1626 [5], but until the advent of computer graphics, much of the work in visualization was work intensive and very limited. Since 1987, the use of computer graphics and computing power greatly changed the scope of how information visualization could be done. Visualization technologies developed rapidly and were distributed across several media, ranging from television, newspapers and print materials. As vector graphics became available, desktop publishing and geographic information systems were also popularized.

Visual representation is the act of translating data into graphic objects with several visual attributes [30]. By applying rules that map data values to certain visual attributes, such as a numeric value to the width of a rectangle, a visualization allows a human reader to evaluate data with cognitive ease. As an example, consider the values in this list; 1300, 1840, 1460, 1180, 1600, 1740. Which of the values were highest? Lowest? Now consider the following visualization:

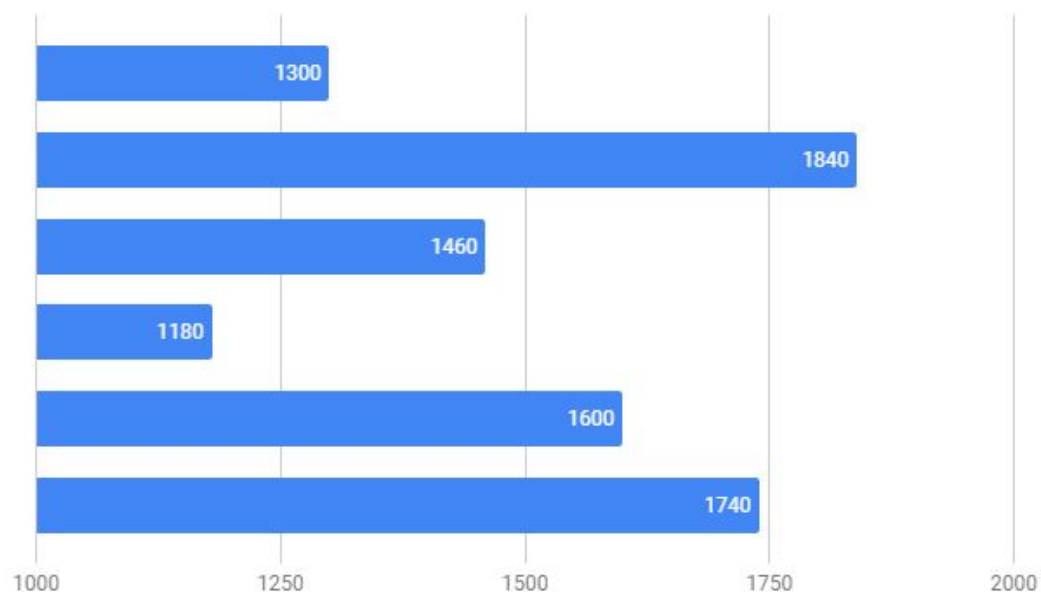


Figure 2.3 Line graph of the values 1300, 1840, 1460, 1180, 1600, 1740.

Due to the visual representation in figure 2.3, it is significantly easier to determine the highest and lowest values using our perception rather than having to parse and

compare the values in the list. Information visualization, according to Shredroff [30], exists in a “continuum of understanding” between data and information. Information visualization as a field has provided with several methods to organize and visually represent data such that information can be produced from it. Card *et al* [5] describes this process as “the use of computer-supported, interactive, visual representations of data to amplify cognition”.

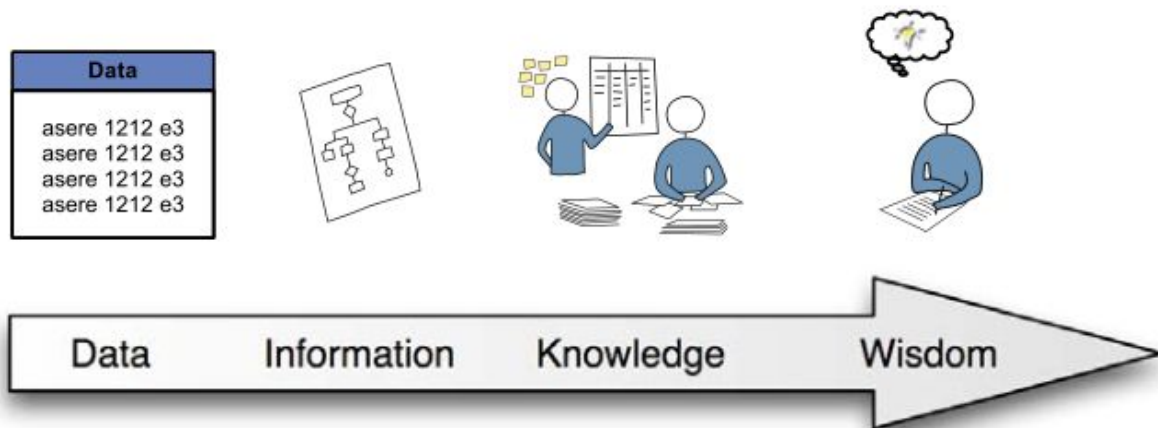


Figure 2.4 Shredroff's contium of understanding.

3 Related Work

3.1 Learning Analytics Artifacts

Visualization of data that learners produce in digital learning environments has been an ongoing field of study since the 1990s. The purposes of such visualizations vary from self-reflection to informing institutional stakeholders, as well as maintaining teacher awareness. In this section we have reviewed some of the artifacts that appeared in our literature review.

CourseVis [24] is one of the earliest systems to visualize student activity for the benefit of the teacher, such that they could identify situations in distance learning and intervene. This system appeared long before the term Learning Analytics came into use. CourseVis was developed to gather data from a Course Management System (CMS), particularly in three aspects; the Social, Cognitive and Behavioural. Mazza *et al* [24] argued that they, through the use of CourseVis, were able to identify and prevent problems from emerging in the distance education course. [24]

CAMera, Contextualized Attention Metadata, was a software tool developed by Schmitz *et al* to support self-monitoring and personal reflection. CAMera gathered metadata from the learner's personal computer, server data and data from a system called MACE (a software used for education in architecture) and visualized these metadata points. CAMera was developed for a personalized learning environment, and not a virtual learning environment such as MOOCs. [28]

The Self-reflection and Awareness or SAM was built for the purposes which its name implies, motivated to counteract a perceived information gap between students in distance education and students in classrooms. To facilitate self-reflection and teacher awareness for distance education students and stakeholders, which otherwise would be facilitated by continuous social contact, SAM gathered data from a variety of sources, including an instance of CAMera, and visualized them. [14]

Moodog [39] is a tool built for the Course Management System (CMS) called Moodle. Since Moodle in itself does not provide with analytics or tracking, Moodog was developed to

provide simple visualizations and statistics. Moodog allows students to check and compare their progress, while teachers can gain an overview of how the students are progressing. The developers of Moodog formulated a hypothesis that "*the availability of Moodog statistics will positively affect both how an instructor adapts the course and how students learn.*" [39]

ANALYZE [19] is a learning analytics tool built specifically for Open edX, which has been documented from inception to deployment. Its formative research highlights some of the challenges associated with the work of Learning Analytics visualization. [19] Many of the findings from the development of ANALYZE were also considered in this research thesis. ANALYZE, or "Add-on of the learNing AnaLYtics Support for open Edx", provides the user with a dashboard that visualizes different activities and provides new metrics that otherwise are not available in Open edX. [19][15]

These studies show that the design of these artifacts depend heavily on the systems they are supposed to complement. In many cases, the data that they visualize has not intentionally been gathered for visualization, but rather the researchers have tried to find viable information visualization techniques for the data as is.

3.2 Data-driven Improvement of Learning Environments

In terms of using data and visualizations to improve a learning environment, the research from Vulic *et al* [37] are of particular interest. They applied what they call 'Data Analytics' to conduct continuous improvement of their 'Through Engineer's Eyes' course, which were run for two iterations. By ordering learner activities into 'story-lines', they used impressions from the first course to inform the design of their second iteration, which they achieved by re-distributing their syllabus into less overwhelming segments. Vulic *et al* [37] argued that data produced by the learners could be used in a 'data-driven course development process' in order to produce the best possible learning experiences. [37] Although many of the terms Vulic *et al* [37] used were tangential to Learning Analytics vocabulary, the substance of their work was particularly relevant to our research. While FutureLearn and Open edX are very different in architecture and data products, this research served to show that course improvement through analytics is feasible.

Vigentini *et al* [36] have also worked with topics that overlapped with our research questions, albeit with different MOOC platforms. Vigentini *et al* [36] developed Learning Analytics dashboards for both Coursera and FutureLearn, aiming to build tools that allow for sensemaking in a system that is overwhelmingly rich in data. They showed how active engagement with stakeholders in early prototyping was a responsive process, such that the stakeholders' needs could be identified. They also emphasized that the development of these technologies are long-term and constantly evolving to the activity patterns in the users. Vigentini *et al* [36] concluded that "There is no doubt that dashboards offer great opportunities for understanding MOOC activity and the effectiveness of the pedagogies implemented in MOOCs; this [research] provides useful and practical observations in order to further the development of data-driven efforts to represent learning-in-action in online learning environments." [36]

Although there exists practises to use data as generated by users and systems in the context of MOOCs, there are also researchers that argue for adding more indicators to the MOOCs in order to facilitate action research. Dyckhoff *et al* [1] understands Learning Analytics as the "*development and exploration of methods and tools for visual analysis and pattern recognition in educational*" which should lead to reflection and in turn the "optimization of learning designs". [1] Such ideas were put into practise in the Imperial College London, where they studied the effects of students and staff engagement on student engagement. Their findings, based on Learning Analytics methods, allowed them to design improvements on their courses. [23] Several institutions have begun to see Learning Analytics not only as a tool for supporting Learner and Teacher awareness, but also as an integral method to evaluate the design of their courses and curriculum. [2][9][27]

These studies highlight that there exists not only a drive to make Learning Analytics aid and improve learner outcomes, but also to support the design and application of the learning environment that the learner and teachers interact with. Amongst many solutions, such changes can be the structure of the course content, the methods the teachers engage with the learners, selection of curriculum and the application of pedagogical methods in the course itself.

4 Methodology

4.1 Design Science Research

The methods of research associated with information systems have been evaluated at a rapid pace in the last few decades. Dresch, Lacerda and Antunes Jr [10] proposed a comprehensive Design Science method, arguing “researches concentrate on to describing, exploring, explaining and predicting phenomena, and little attention is devoted to prescribing solutions.” Their method is a synthesis of several proven methods ordered in twelve steps. The framework they propose suggests iteration as part of the research process, as well as multiple outputs. Even though an article or thesis is the expected result, Dresch *et al* [10] suggest also codifying heuristics at the “development of the artifact” and “evaluation of the artifact” steps, which can be set to use in other works.

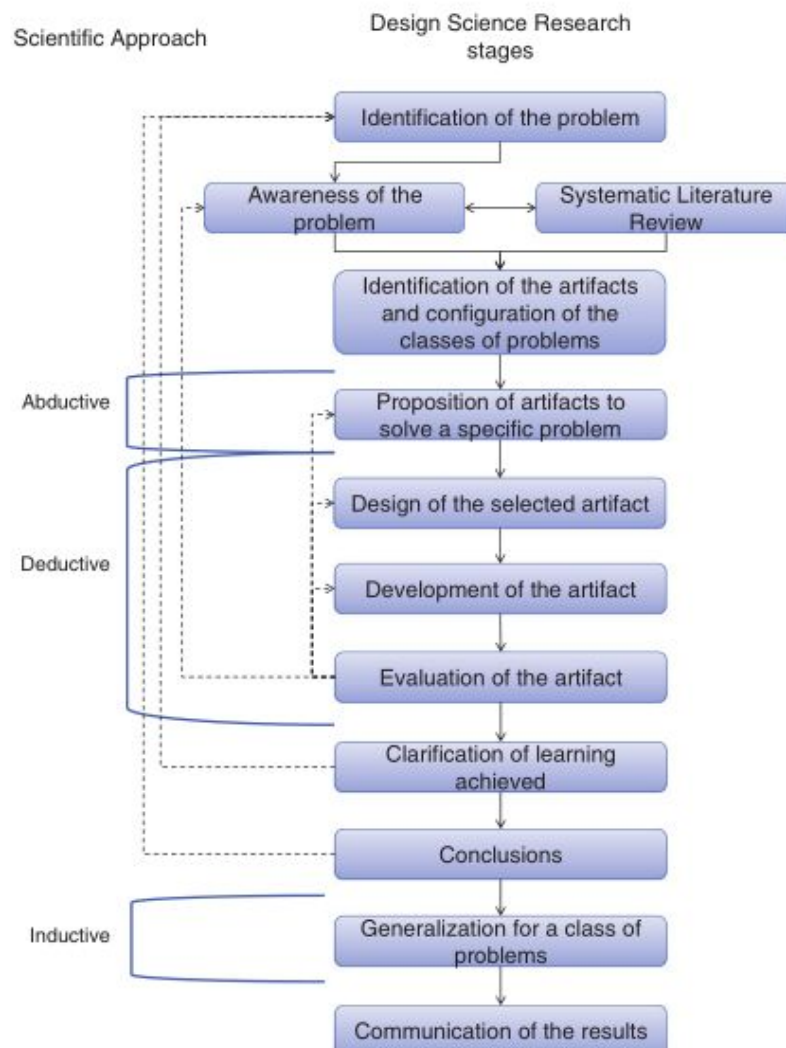


Figure 4.1 Dresch *et al*'s [10] diagram of their Design Science Research model.

The method (hence referred to as DLA) proposed by Dresch *et al* [10] suggests the researcher identifies a problem to begin with, before building an awareness of the problem through the use of either systemic thinking or theory of constraints. This step should also involve a systematic literature review, alternating between consulting knowledge bases and using developing the researcher's awareness of the subject. Once this process has reached sufficient mass, the researcher should be able to move forward with identifying any existing artifacts and a configuration of the Class of Problems. Given that many fields of inquiry are not untouched domains, it is useful to integrate existing work in the field to the researcher's understanding of their given problem. If similar solutions exist, there may be generalizable heuristics to extract from those works into the researcher's own work.

Another highly iterative segment of the DLA is the abductive and deductive segment, which culminates in an evaluation of an artifact. Preceding this evaluation is the proposition, design and development of the same artifact. These steps are linked in such a way that the process can return either of the preceding steps in order to account for new advances or discoveries made during the processes leading up to the evaluation.

Further steps may also lead to a complete iteration, such as the conclusion of the project actually becoming a proposition or the basis for a proposition of a new artifact. Finally, DLA proposes extensive work to both generalize and communicate the body of work. If the work is not successfully communicated, the effort are essentially made in vain.

In this research, the following interpretations were made of the steps in figure 4.1:

- **Identification of the Problem:** In this step it is crucial to identify a problem for scientific inquiry. It must be seen as relevant in the state of the art or as a progression of earlier iterations, and it has to be objectively recognizable.
- **Awareness of the Problem:** In DLA, building an awareness of the problem is advised to be handled with systemic thinking. The objective is to understand the problem in a broader context and its position in causal relations. Dresch et al [10] suggest two approaches; either systemic thinking or using Theory of Constraints.

For this research, we used systemic thinking, as it is more intuitive to apply to the specific problem we dealt with.

- **Systematic Literature Review:** This step is the other pillar of information gathering. Consulting knowledge bases is crucial for developing an awareness of the problem as well, and Dresch *et al* [10] suggest that there should be a local iteration between this and the previously described step. The objective is to draw upon the accumulated knowledge of the traditional sciences and the design science discipline. These pillars should direct the researcher into transitioning to the next step of the method.
- **Identification of the artifacts and configuration of the classes of problems:** In this step, the researcher should attempt to identify any existing artifacts and attempt to frame the original problem in a class or classes of problems. The problem the researcher has identified may have been identified in other works, and there may exist artifacts that overlap in functionality or intent. Similarly, the problem may exist adjacent to other problems that are very similar in nature. Identifying these should inform the researcher about the best practises and valuable lessons learned by experts in the field.
- **Proposition of artifacts to solve a specific problem:** Based on the accumulated information from previous steps, the researcher should now be able to develop a proposition for an artifact.
- **Design of the selected artifact:** In this step, the design of the artifact itself takes place. Dresch *et al* [10] advises that that the researcher clearly expresses the procedures used in this step.
- **Development of the artifact:** In this step, any possible development of the artifact occurs. Two outputs are expected from this step. While the artifact is of primary interest, this step should also produce a set of construction heuristics. In software development methods, this could be considered a call to produce documentation. Other types of construction heuristics could be diagrams that reflect the relations both externally and internally in the artifact.
- **Evaluation of the artifact:** As the development of the artifact concludes, it is pertinent to perform an evaluation of it. The artifact should not qualify beyond this step unless it meets any requirements posed to it in earlier steps, especially when it comes to solving the identified problem. If the artifact has not met all requirements to a satisfactory level, the researcher should return it to previous steps and account for the deficiencies. In addition to producing an evaluated

artifact, this step should also produce contingency heuristics. Contingency heuristics should convey the performance of artifact, or reflect facts about the artifact on which further actions are contingent of.

- **Clarification of learning achieved:** At this step, the research question should have been answered to some degree, with the artifact's development having reached its final evaluation for the purposes of the research project.
- **Conclusions:** When reaching this point, the researcher should have results that are valid outcomes of the research process and the guidance of the DLA method.
- **Generalization for a class of problems:** In this step, the artifact and heuristics should be generalized as best as possible. If there is a significant Class of Problems, then the artifact and heuristics should be applicable to those as well.
- **Communication of the results:** Finally, a communication of the research project's result should be communicated to any parties that find the topic of research relevant to their studies.

4.1.1 Artifact

DLA refer to Simon [32], who described artifacts as designed artificial things, which are still subject to natural laws. In the environment the artifact exists, it can and should produce measurable results. For the purposes of this research, the artifact is a software system that creates the visualizations. Initially, the host project, which this research project is attached to, prescribed an extension of some manner to the Open edX instance on which the course is hosted.

4.1.2 Class of Problems

Dresch *et al* [10] define a Class of Problems as *"Thus, we define Class of Problems as the organization of a set of problems, either practical or theoretical, that contain useful artifacts for action in organizations."*, which is pertinent to the research direction of this project. Although we consider my research project to be in the domain of Learning Analytics, there are likely intersections in other fields that revolve around the same Class of Problems. Using visualizations for decision support is likely to exist in other fields of research as well.

4.1.3 Systemic Thinking

In the “Awareness of the Problem” step of the DLA method, we used Systemic Thinking.

Systemic Thinking [34], or systems thinking, is a method used to perceive systems and their behaviour. This method is often used to identify special relationships in a system, such as nonlinear behaviour and feedback loops. Systems thinking emphasizes understanding the subject of study in its relationship to others, instead of its essential characteristics and components. Systems thinking uses synthesis instead of analysis to achieve this understanding.

Since the artifact of this research is positioned in a context that involves several elements in separate domains, the use of systems thinking could provide with valuable knowledge. Systems thinking is frequently used in systems engineering, a field which overlaps highly with software engineering.

4.1.4 System Usability Scale

Another method from systems engineering that is useful for evaluation, System Usability Scale (SUS) questionnaires were used in the “Evaluation of the artifact” step in the DLA method. The questionnaire contained a set of questions with five possible responses, ranging from strongly disagree to strongly agree. The SUS questionnaire emphasized the responder’s ability to perceive the visualizations and make judgements on the information presented in them.

The following questions were asked in the survey:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the different visualizations to be consistent.
6. I think that many of the visualizations were unclear or noisy.

7. I think most of the visualizations are intuitive or meaningful.
8. I think data needs to be better presented.
9. I feel that the visualizations help me with my work.
10. I do not think I will use the visualizations to inform my opinion.

4.1.5 Use Case Evaluation

Use Case Evaluation was used for an early evaluation in the research project. The research used in particular the method described by Hornbæk et al [17].

4.1.5 Likert scale

Likert scales are used to operationalize beliefs, opinions, perceptions or attitudes. Invented by Rensis Likert, this scale was originally used as a psychometric scale, but has been broadly adopted in Sociology and other fields that can benefit from this type of metric. [16] In this research, a questionnaire with 26 questions was developed with a scale ranging from 1 to 5. This method was used in the final iteration to gauge participant motivation to improve learning environment conditions. The possible responses were:

1. Strongly disagree
2. Disagree
3. Neither
4. Agree
5. Strongly agree

4.1.6 Semi-structured interviews

Semi-structured interviews are hybrid interviews that follow a less rigid style of posing questions to the interview subject. [16] We elected to use these to pursue questions and ideas as they arose in conversations with the stakeholders. In the evaluations of each iteration, any mentionable output from such an interview was noted in the “Feedback from Stakeholder” sub-sections.

5 First Iteration

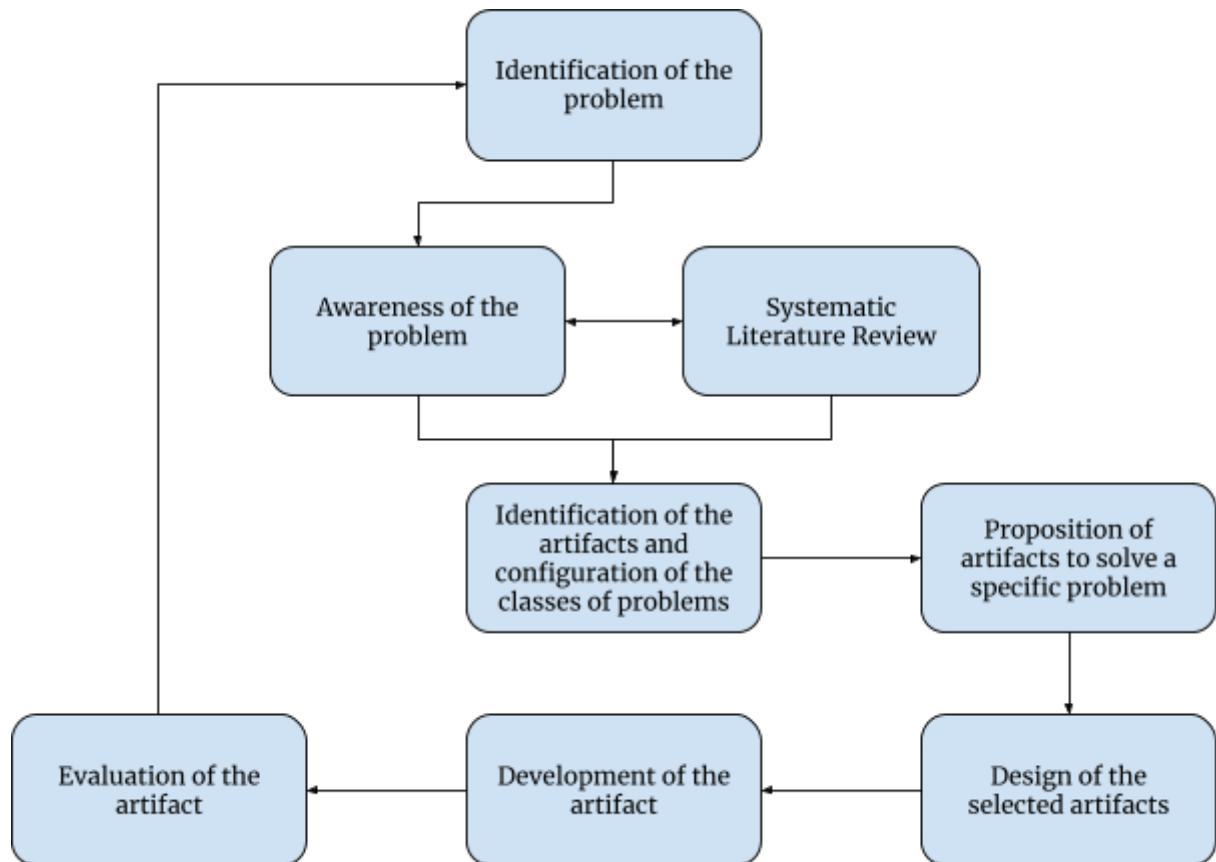


Figure 5.1 An overview of the first iteration cycle.

5.1 Identification of the Problem

This first iteration started on the 15th of August, which started with an effort to identify a problem for research. This aligned with a collaboration project between SLATE and Oslo Met (formerly University College of Oslo and Akershus), in which Oslo Met course developers requested Learning Analytics to be applied to their blended classroom course. Blended learning environments feature some challenges, especially that the teachers are not able to perceive what happens to the students' activities when they are active in the online environment. A base-level LMS such as edX has very simple feedback mechanisms in place, meaning the teachers have to rely on mostly manual methods to gain feedback on student activity.

A basis for the funding of the course was that Learning Analytics would be involved in the process of the course's lifetime. SLATE reached out to the managers of Akademix.no, which had experience in maintaining the infrastructure necessary for running edX courses.

Upon further examination, it was also found that a majority of the feedback that edX Insights provides is not necessarily relevant, such as demographic data. An identifiable problem appeared, where it would be necessary to provide with analysis that could give feedback in some manner would be more useful to the course stakeholders. The collaborators at SLATE were interested in developing a plugin for the Open edX framework that would be capable of rendering custom visualizations.

At this point, it was prudent to move on to both expand an awareness of the problem and conduct literature reviews.

5.2 Awareness of the Problem

At this stage, we attempted to apply systems thinking to the problem which we had earlier identified, and in addition, we had to build an awareness of the context which the problem exists in. Since at this stage there is only a want for an artifact, but no actual definition of it, the systems thinking process had to first include elements that surround the problem's position. In order to facilitate an awareness of the problem, we made lists of all the plausible elements to discuss in a systems perspective of the problem. Following that, we attempted to identify their relations.

There are several elements that are involved with the problem. Amongst these, there are several system elements, which are the following:

- A Server.
- A Client machine.

- The Open edX software.
- The courseware.
- The log database.
- The classroom
- The course

In addition to the system elements, there are also actor elements:

- The Learner (or Student)
- The Teacher

There are also organization elements at play:

- OsloMet
- Akademix
- SLATE

The results of this exercise led to several diagrams that show the relationships between the elements listed above.

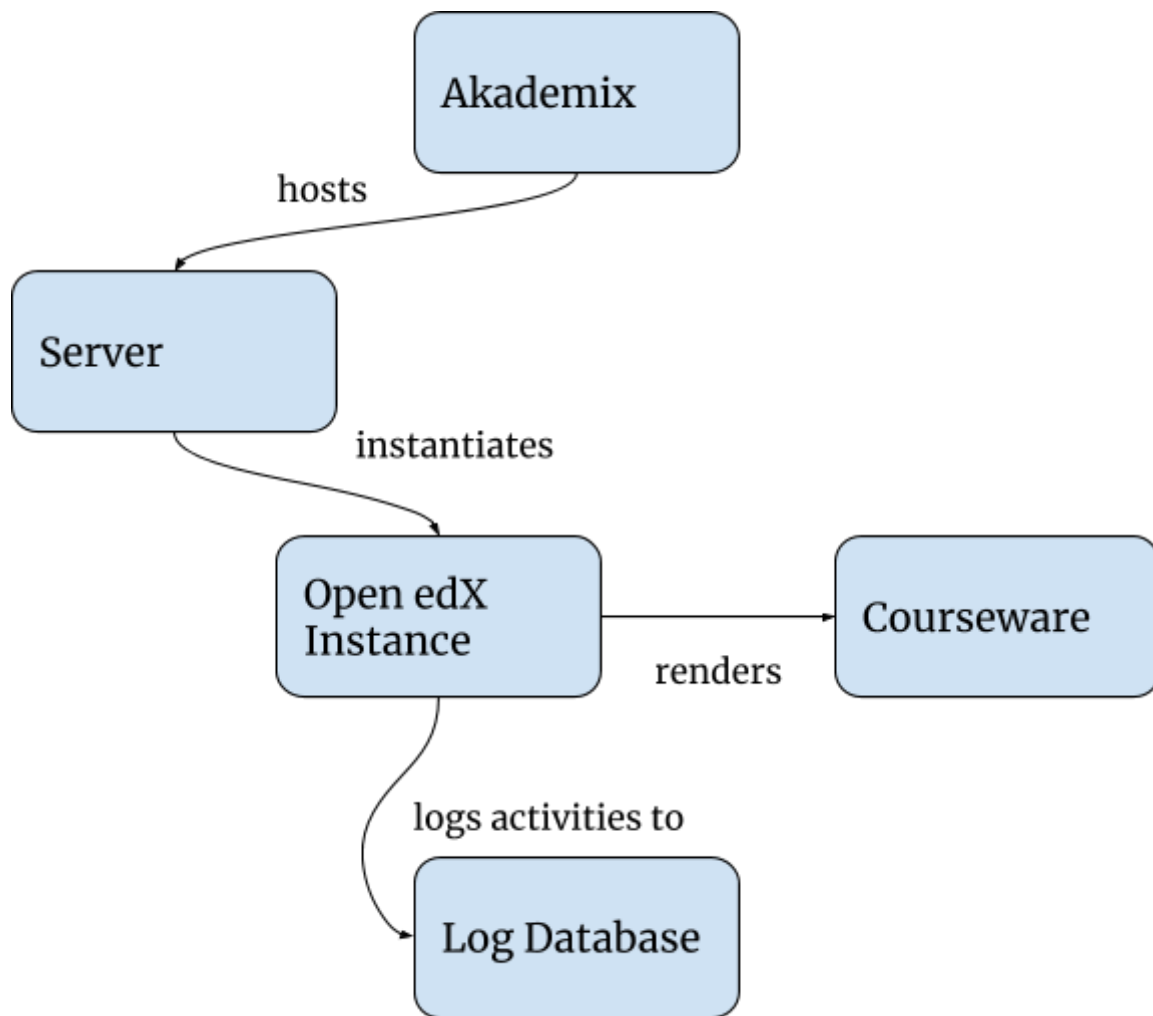


Figure 5.2 Illustrating the relationship between Akademix and the systems they administrate. In this figure there is no indication of user interaction with the courseware.

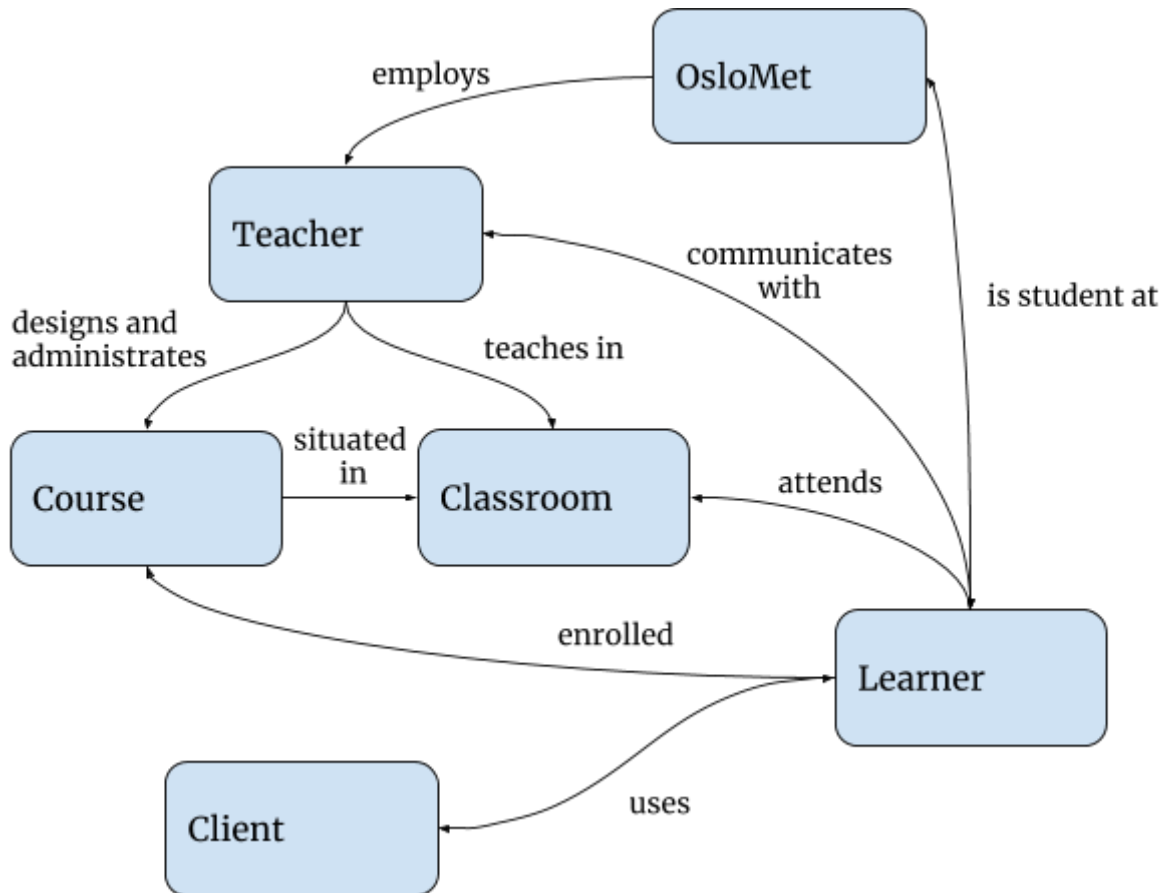


Figure 5.3 Illustration of the various elements of the OsloMet domain.

The domains illustrated in figures 5.2 and 5.3 interact as described in the following figure:

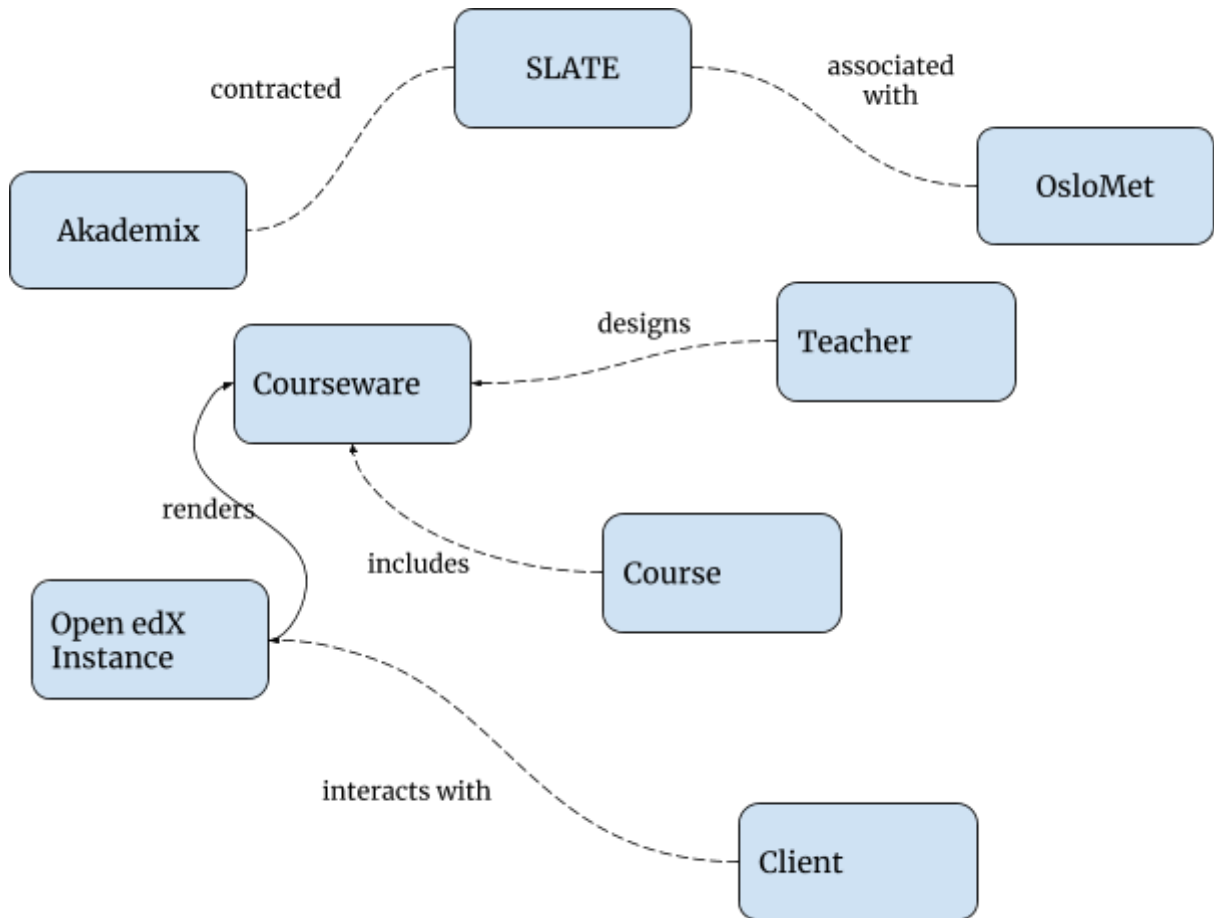


Figure 5.4 Illustration of how the elements from figure 5.2 and figure 5.3 interact.

When we joined these relations, it produced the following figure:

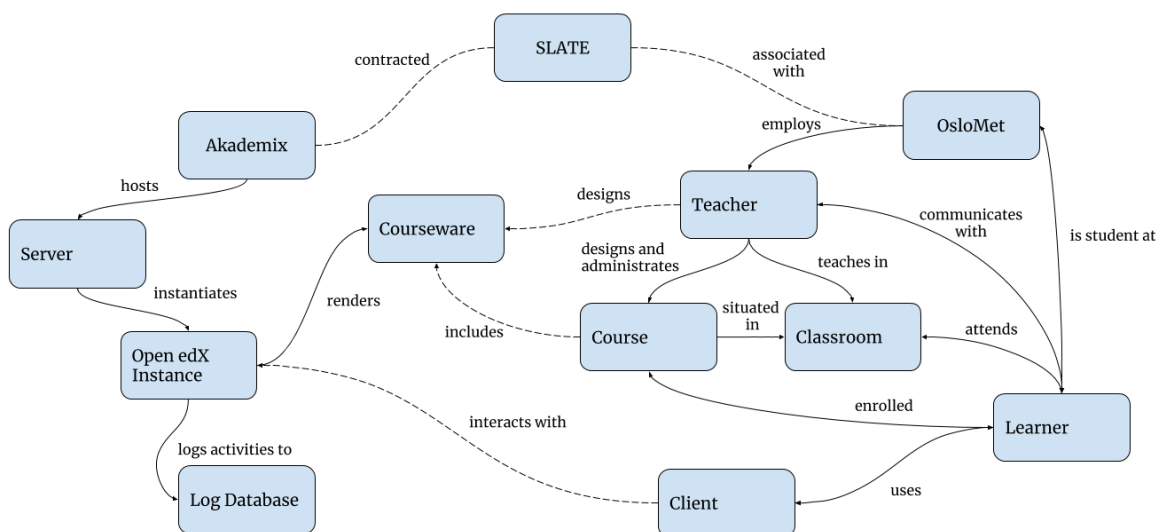


Figure 5.5 A comprehensive interaction diagram of the systems' interactions.

By utilizing systems thinking, we established that there is a close relationship between the Teacher and the Learner through the classroom. However, interaction between the Learner and the Open edX Instance is somewhat distanced from the Teacher. The Learner is free to communicate their experiences with the courseware to the Teacher, but the Learner's interactions on the Open edX Instance are not visible in any way to the Teacher. The Log Database, while technically accessible to Akademix staff, is not at all visible to the Teacher.

This means that much of the interactions between the Learner and the Courseware does not recur any feedback to the Teacher, unless either actor decided to communicate this between each other. Ideally, the position of an artifact could be between the Log Database, Teacher and Learner.

5.3 Literature Review

At this stage, we gathered publications that deal with the topic of Learning Analytics. Learning Analytics is a relatively new field, dating back to 2010, and has hundreds of publications that pertain to the field. In addition to this venue, we also compiled a list of documentation for Open edX, including the actual source code itself as a means of attaining understanding of the problem. For example, the documentation for edX Insights provided with a comprehensive overview of the learning analytics it provides for the edX platform, including screenshots and user instructions.

An important factor in identifying useful literature was to cross reference Learning Analytics and MOOC as search terms. One particular article stood out, the "Towards the Development of a Learning Analytics Extension in Open edX", which not only provided an analysis of the learning analytics in the platform, but also provided a review of the technical and infrastructural aspects. In addition, the paper identified related works which provided a longitudinal insight of other efforts to build Learning Analytics applications for similar environments.

Many of the findings from this stage are reflected in the Preliminaries and Related Works chapters, found earlier in this document.

5.4 Identification of the artifacts and configuration of the classes of problems

After conducting the literature study, we identified several attempts to implement learning analytics and the different approaches they have taken. Some of the big MOOC platforms, including edX, are Coursera, Udacity and Khan Academy. Some similar platforms exist as well, such as Canvas, Moodle and It's Learning, which also incorporate learning analytics.

Although efforts have been made to apply Learning Analytics in these systems, it becomes increasingly clear that many such implementations are constrained by several factors. A major constraint that all these artefacts share is that their primary input is traces of student activity, and to a lesser degree teacher activity. In Virtual Learning Environments that do not have forums or messaging functions, it is even less likely to find traces of teacher activity. In many cases, students are the sole focus for Learning Analytics implementations, which to a degree matches the pattern of self-teaching which most MOOC designs. The principal traces that students leave behind are interactions with the system, primarily in the form of system logs and inputs through interaction with courseware components such as quizzes, videos and site widgets.

Several artefacts have been created in attempts to solve the identified problem, but with variances based on platform and infrastructure specifics. Since the systems discussed are similar, but not identical, the theory or competence questions that drove the development of these artefacts could be considered for the project we worked with. For instance, the Khan Academy module "ALAS-KA" emphasized visualizations of student efforts, albeit based on Khan Academy's courseware.

More adjacent artefacts exist as well, such as Students Activity Monitor and the CAMera artifact. CAMera collected metadata about the learner's activities, such as email exchange and interactions with certain desktop applications, and visualizes to the learner in order to allow for personal reflection. Students Activity Monitor offered self-monitoring and reporting to increase teacher awareness for a personal learning environment (PLE).

The closest and most usable artifact we could find that had already been in production was edX's own Insights. Despite having been in production, the edX Insights application is still undergoing development.

5.5 Proposition of artifacts

Based on information gathered in earlier steps, we determined that although there exists several artifacts that solve problems in the proximity of our identified problem, few of them are actually compatible with the systems involved. The best, but not ideal, candidate of existing artifacts would be the edX Insights extension, which is edX's own Learning Analytics solution. In spite of SLATE's wishes to use develop custom analyses, running Insights was also an alternative they were prepared to pursue.

Thus, with an eligible artifact as a solution to the problem, we proceeded to use a design that involved edX Insights.

5.6 Design & Development of the Artifact:

Following the selection of edX Insights as the the artifact for this iteration, these steps of the DLA method were nominal. We drafted a plan of action to install the edX Insights module and reached out to Akademix to facilitate the install of the module. Upon having communicated with Akademix staff, it appeared that it was

not possible to install edX Insights. The reason appeared to be related to policy, meaning that the artifact itself was simply ruled out for the purposes of this research project.

5.7 Evaluation of the Artifact:

Considering that the Akademix related systems did not allow for this artifact to function, this step was also shortened. The artifact did not satisfy the identified problem as expected, so we had to return to the "Identification of the Problem" step in order to internalize the new information this iteration had presented.

The first contingency heuristic from this step (henceforth CH1.1) was that the Akademix related systems by policy did not permit any modifications or customized installs. Uncovering this information incurred a significant course correction on the project.

The second contingency heuristic (henceforth CH1.2) was that Akademix could transfer data to us with certain intervals. The data they could transfer was user activity logs.

This also produced a third contingency heuristic (henceforth CH1.3), which is that any further development actions or potential artifacts cannot be directly deployed or tested with the Akademix systems.

6 Second Iteration

The second iteration began on the 20th of September 2017, shortly after a meeting with Akademix staff.

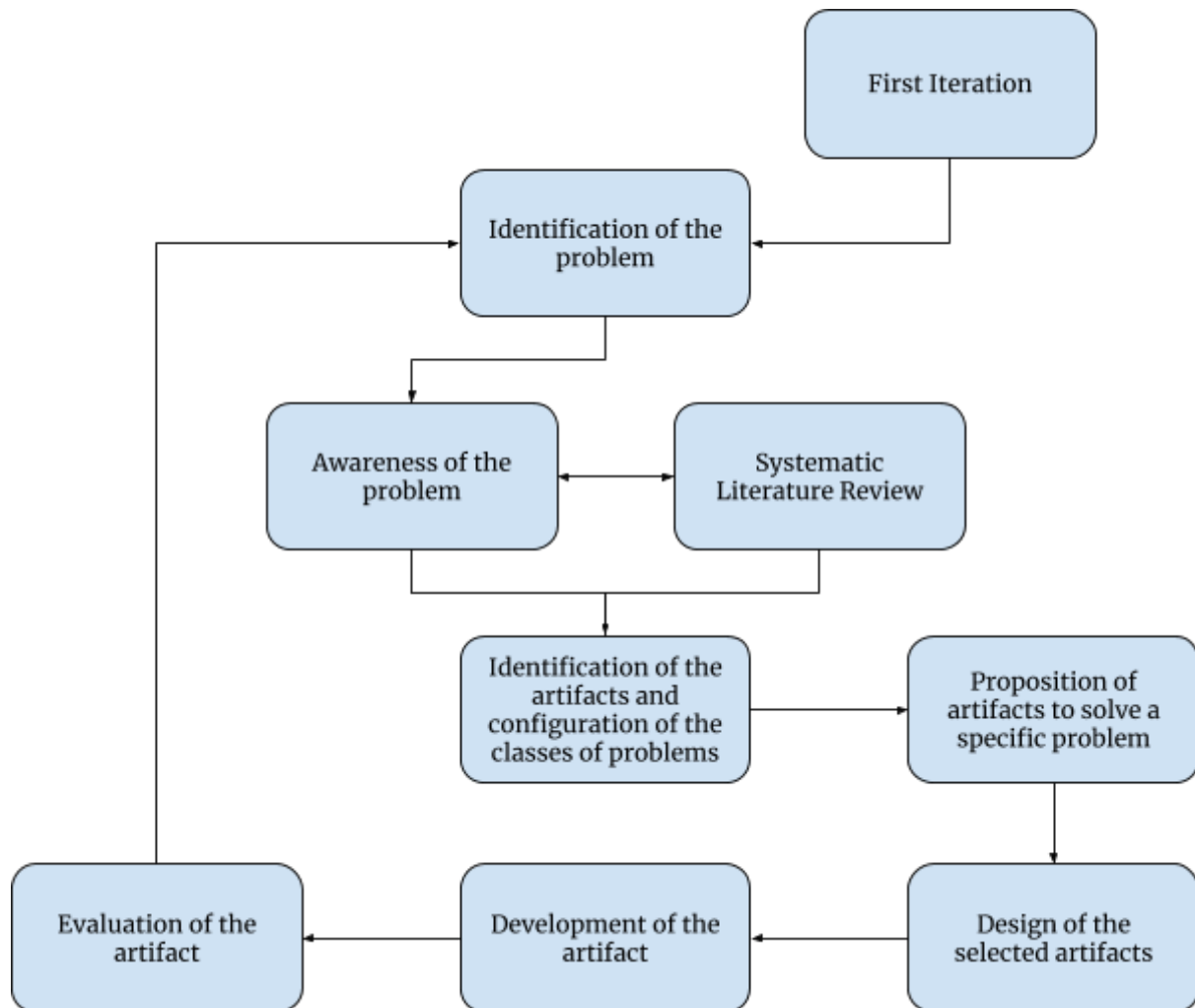


Figure 6.1 An overview of the second iteration cycle.

6.1 Identification of the Problem

Based on CH1, CH2 and CH3 from the first iteration, the problem's nature had changed. Insights was no longer a viable option, and conditions about the systems had changed in light of these new contingency heuristics. The problem for the research study became more narrowly defined; the stakeholders were not receiving any Learning Analytics on their course, and due to CH2, the only way of

producing Learning Analytics for them would be to analyse data sent to SLATE from Akademix.

6.2 Awareness of the Problem

Based on the Contingency Heuristics from the previous iteration, it had become clear that we could not depend on the infrastructure that Akademix provided. Instead we would have to build our own type of artifact which could function in complement to the situation at hand. Based on Contingency Heuristic 1.2, we could expect to receive data, but with little control over its format or rate of transfer.

6.3 Proposition of Artifact:

Based on information gathered from previous steps and the first iteration, several requirements could be established based on the configuration of systems and the environment we are acting in. At this point, the stakeholders were not yet involved, so user stories were defined by the researchers with perspectives on what could be done with the systemic conditions surrounding the problem.

The artifact would at least require the following characteristics:

- The artifact should be able to ingest log data.
- The artifact should be able to create visualizations.
- The artifact should be able to store results in a database.

Since communications with the stakeholders had not yet allowed for a planning game to be executed, we developed a persona-based stakeholder as a stand-in to justify building simple infrastructural user stories:

- A. As a user, I want to view the results of individual students.
- B. As a user, I want to view the results of all students.
- C. As a user, I want to view which parts of the site the student has visited.
- D. As a user, I want to see which dates the student is active.

E. As a user, I want to specify which data file is to be read.

Note, these stories are henceforth referred to by their letter on the above list.

With some user stories and characteristics to go by, the research could proceed to begin design of the artifact.

6.4 Design of the selected artifact

With some user stories in place, we started an Iteration Planning sequence as per the Extreme Programming methodology. We started with the Exploration Phase by resorting to systems thinking, with the intent of mapping the stories to concrete tasks. In figure 6.2 we show an output of such systems thinking, where we order the stories by theme.

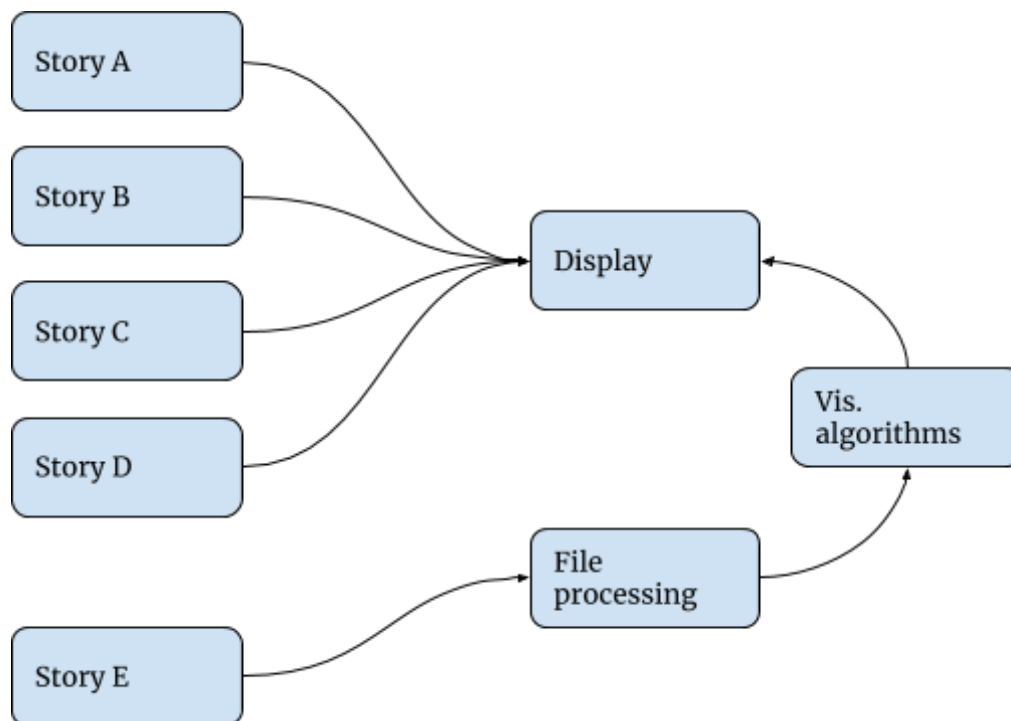


Figure 6.2 A diagram of how the user stories relate to functions in the artifact.

In figure 6.2 we discovered how little the user stories describe what needs to be done in terms of taking the information found in the data files to producing a

finished visualization. For this reason, we continued the systems thinking in an attempt to prescribe a process based on the elements established in the user stories and in figure 6.2.

This second effort established an something akin to a construction heuristic, as shown in figure 6.3. The product was not too dissimilar to an UML diagram. With this figure in mind, we developed tasks that were related to each of the steps described.

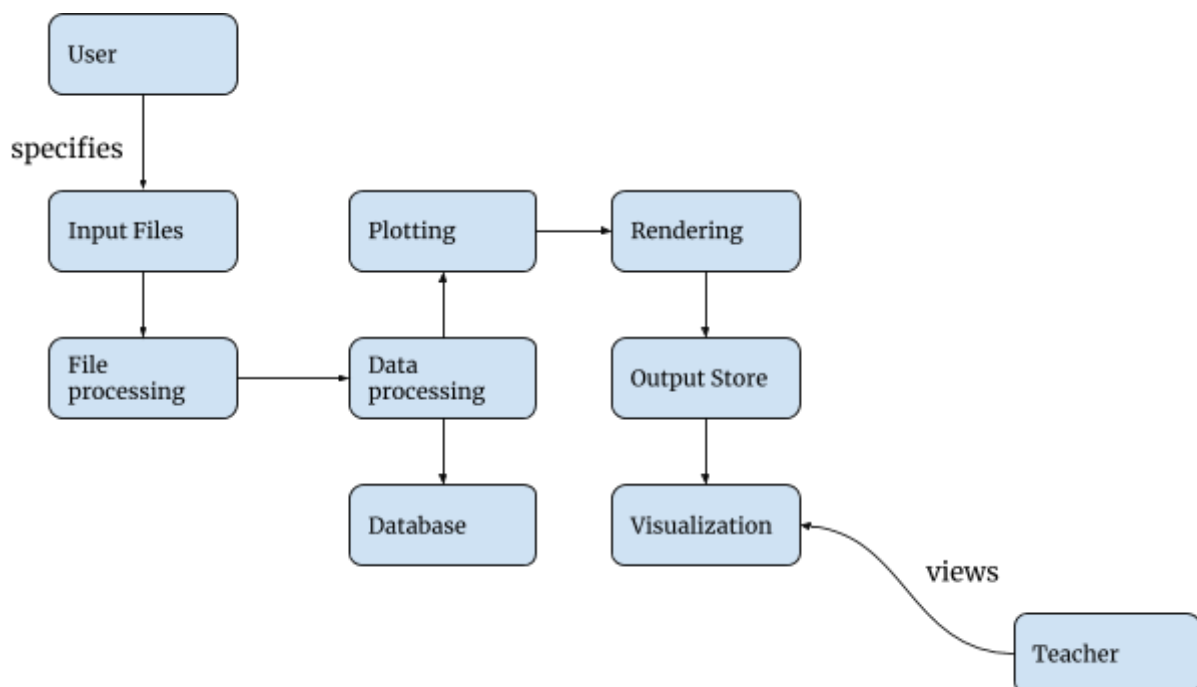


Figure 6.3 Systems thinking diagram of how the artifact takes input data to the final stage of visualization.

Exactly how these visualizations were to be presented was still up for debate at this stage. For this reason, we decided to store the results of the data processing in two formats. A general paradigm in modern web development is to provide data through API (Application Programming Interface) endpoints, which can then be consumed by almost any type of presentation layer on a website. We decided that taking this path could be a pragmatic step, and also separate a major part of the

artifact’s development into two distinct and separable releases. Thus the actual presentation and interaction part were ruled out as tasks for this iteration.

By considering the *Shared Understanding* principles, we decided that the code should be written in much the same style as Open edX, and be mostly compatible with its environment. For this reason, the selection of technologies were mostly foregone conclusions. The artifact would be written in Python, and it practically become a Django Representational State Transfer (REST) API. This decision was informed on the consideration that maybe the Akademix policy might be changed, or that OsloMet might consider running the courseware on another kind of infrastructure.

Stories A through D did not produce a lot of tasks, and Story E produced the most tasks. After generating tasks, we had a Kanban board that looked as on figure 6.4.

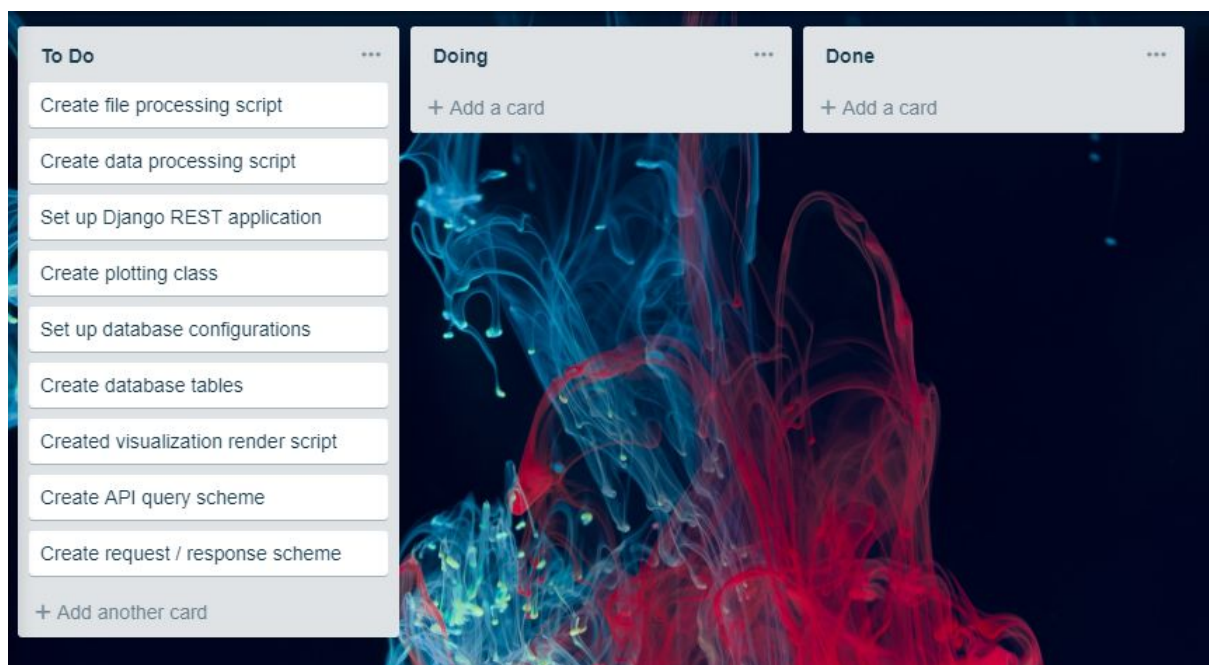


Figure 6.4 Screenshot from the project’s Kanban board on Trello.com.

The artifact’s elements were ordered in a relatively linear manner. We decided to prioritize tasks associated with the “topmost” processes first. This also meant that we did not adhere to the Planning Game’s assignment of value and vector to the

tasks. Ordinarily, the Exploration Phase calls for dependent tasks to be ordered according to necessity, so we followed that pattern as well.

The tasks are referred to in the following order later in the text:

- A. Create file processing script
- B. Create data processing script
- C. Set up Django REST application
- D. Create plotting class
- E. Set up database configurations
- F. Create database tables
- G. Create visualization render script
- H. Create API query scheme
- I. Create request / response scheme

6.5 Development of the Artifact:

Developing the artifact introduced a greater amount of information into the research project than we had first expected. Many of the assumptions we made about the environments were seemingly insufficient. Working with Task A proved that we had not taken into consideration how a file with data would reach this system. We completed the task under the assumption that some other process would place the file in an expected directory.

Task B was implemented in a simple fashion, with essentially a call to Python's CSV package and setting up a pipeline for pushing lines of data to MySQL. Task B through F provided no real challenges. Most of the database configurations were mock configurations, just to prove that it was possible to perform the exchanges we wanted to occur in the artifact.

Task G was set up as a mock class for this iteration, providing only with blank image files using the 2D plotting library Matplotlib.

Further on, H and I were completed but in a simple, rudimentary fashion. Users could pass a student number and an analysis type as parameters to the API, leading to a response with meta-information and a link to the requested visualization. These tasks were implemented to satisfy stories A through D. Essentially, all matters of selection and viewing would happen through making requests to the API.

6.6 Evaluation of the Artifact

The artifact prototype was subsequently ready for evaluation, but it was at the onset of the evaluation very clear that the artifact did not actually address the problem directly. In this iteration, it was only an utility to facilitate the production and deployment of the visualizations. While it technically could produce some bar charts from the test data, it was not tested in the systems environment that the course existed in.

For this reason, we created an evaluation scheme that would evaluate the utility, robustness and performance of the artifact. This evaluation was conducted in three stages:

1. Use Case Evaluations
2. System Usability Scale surveys
3. Performance tests

For the first two evaluations, we recruited three web developers at different stages in their career, having 1 year, 5 years and 8 years of professional experience respectively. The artifact was deployed on a virtual machine and PEM-keys were issued to each of the developers, so that they could access the artifact using secure shell (SSH) terminals. These evaluations were handled remotely over Discord - an application which facilitates Voice over IP communications, amongst others. For each participant, each stage of the evaluation process lasted roughly an

hour. The last evaluation was simply a benchmarking test where we measured the response times for a set of queries, and the participants were not involved in this process.

The Use Case Evaluations were based on the user stories, where each user story was used as a basis for prescribing a use case. Although the artifact at its current state had no user interface to speak of, it had modes of interaction through either a browser or command line (by use of cURL, for example). The results of this evaluation are found in the “Use Case Evaluation Results” section later in this chapter.

The System Usability Scale was introduced after the Use Case Evaluations. Normally System Usability Scale surveys requires at least two participants, so three participants is just above the minimum requirement in this case. This same survey would be issued to participants in future iterations, so we expected that the results from the current survey might be less congruent with the results of future surveys. The results of this evaluation are found in the “System Usability Scale Results” section later in this chapter.

6.6.1 Use Case Evaluations Results

The Use Case Evaluation was set up using five use cases pertaining to the user stories (see appendix *). In brief the following tasks were presented to the participants:

- A. Retrieve the analysis A results of student 12.
- B. Retrieve the analysis A results of all students.
- C. Retrieve the path to the most visited site of the course ware.
- D. Retrieve the path to the least visited site of the course ware.
- E. Specify a file to be analyzed.

The participants used the the Nielsen Heuristic to identify usability issues during the systematic inspection, and by basis of selection we expected participants expertise on web development would be involved. At the conclusion of their evaluations, the participants wrote down usability issues they found, noted with the particular guideline they believe is violated and a note explaining the issue they found.

The results were as follows:

| Task | Observations | Instance | Note |
|------|--------------|---------------|--------|
| A | 3 | 1, 2, 8 | A1 |
| B | 3 | 1, 2, 8 | A1 |
| C | 4 | 1, 2, 4, 8 | A1, A2 |
| D | 3 | 1, 2, 4, 8 | A1, A2 |
| E | 6 | 1, 2, 3, 5, 8 | A1, A3 |

Tab. *.1

| Task | Observations | Instance | Note |
|------|--------------|----------|------|
| A | 1 | 3 | B1 |
| B | 1 | 3 | B1 |
| C | 1 | 4 | B2 |
| D | 1 | 2 | B3 |
| E | 2 | 1, 2 | B4 |

Tab. *.2

| Task | Observations | Instance | Note |
|------|--------------|----------|------|
| A | 0 | - | - |
| B | 2 | 4, 6 | C1 |
| C | 0 | - | - |
| D | 0 | - | - |
| E | 3 | 1, 3, 8 | C2 |

Tab. *.3

The following notes were compiled from observations made in the evaluation:

A1:

Several state queries have long delays with no indication as to why. The participant wonders if heavy algorithms are tied to general queries. Several of the JSON results have unintuitive key and value associations. Some intentionally bad queries yielded no error report.

A2:

In task C and D, the list returned gives no clear indication if it is in ascending or descending order. The participant noted that there should be an indication of what the path links to, without having to click it.

A3:

The artifact suggests files that it detects in its read folder, but it is possible to pass invalid file names to the request. This leads to an exception, which should be avoidable in the first place. The participant suggests a more rigid and fault-free file selection system. In addition, it is not clear if files have previously been read or not - this would be greatly beneficial to the users.

B1:

When attempting to retrieve Analysis A for students, the participant purposefully queried student indexes beyond the list size. After receiving an exception error, it was not possible to backtrack or simply query a student number within the list size. The participant had to restart the artifact.

B2:

Query results had unclear or strange key vs value names. While some names make sense to programmers, the metadata may be unclear to academic staff.

B3:

The participant commented on how this task produces a list of URLs, but not the names of the sites. Additionally, the participant commented on there being no visibility of the list's order.

B4:

The participant noted that the return values when a file was specified were unclear, as the result was not in natural language or informed the reader poorly. In addition, the API does not make it clear what files have been read when the base-dialogue we accessed.

C1:

The requirements for querying all students in analysis A was ordered differently compared to the result from task A. The participant suggested that the queries should be more uniform and return a list, regardless of how many students are polled with the query.

C2:

Error handling was difficult with the artifact. Specifying which files to read, or accidentally hitting the query twice caused disruptions in the system, including broken data points in the analyses. It was not possible to "undo" erroneous file choices, or accidental queries. The participant desired a solution to reset the API or reverse the analysis of a specified file.

6.6.2 System Usability Scale Results

After having implemented the Use Case Evaluations, we asked the participants to fill out System Usability Scale surveys. In this iteration there were only three participants filling out this survey, and even so the SUS survey could provide with useful assessments about the artifact.

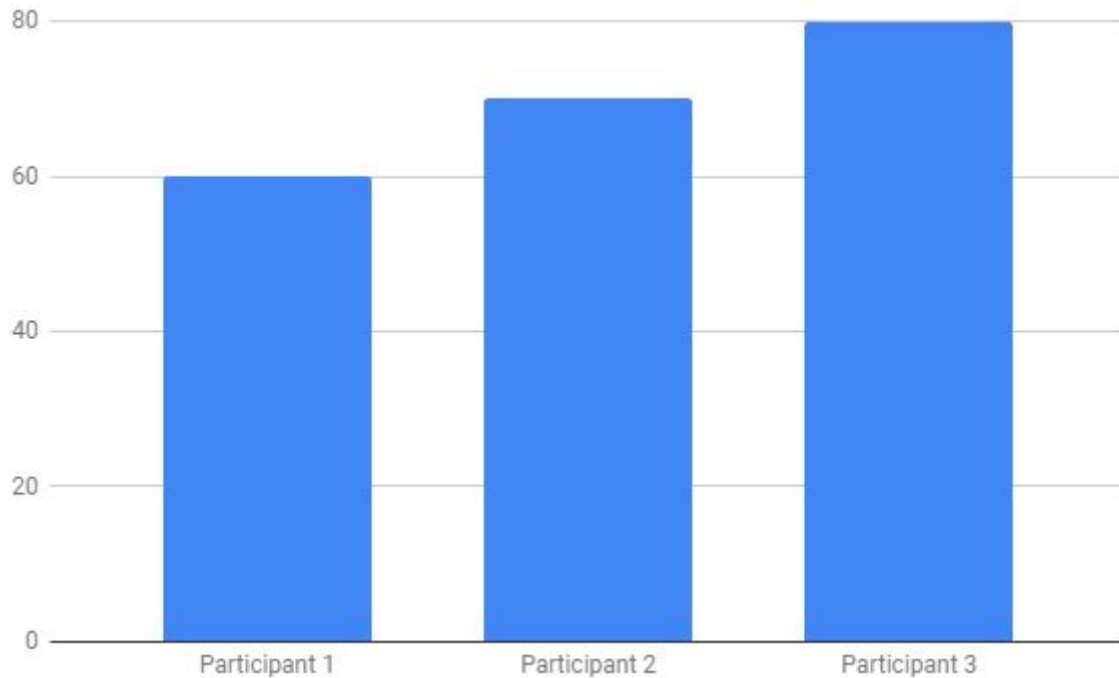


Figure 6.5 Results from the initial SUS survey.

Overall, the results were not particularly favorable for the artifact, with an overall average of 70 points and a maximum difference of 20 points. The scores matched my expectations of the artifact as well as the results from the Use Case Evaluations. The artifact had to be improved considerably.

6.6.3 System Performance tests:

The system performance tests were set up to measure the responsiveness of the artifact and to see if it was possible to scale up. Each test was run three times and recorded the average time as the result. In several cases, each run had the same response times. The following tests were set up:

- A. Start up time.
- B. Process 10 students.
- C. Process 100 students.
- D. Process 1000 students.
- E. Process 10000 students.
- F. Response time on visualization request.

With test A, we wanted to see how long the artifact would take to start up, which included launching the entire stack except for the operating system. This meant restarting MySQL and initiating the Django project.

Tests B through E were in this iteration tested with mock-up student data, which means the results will not be entirely comparable with the later iterations.

The results of the test were the following:

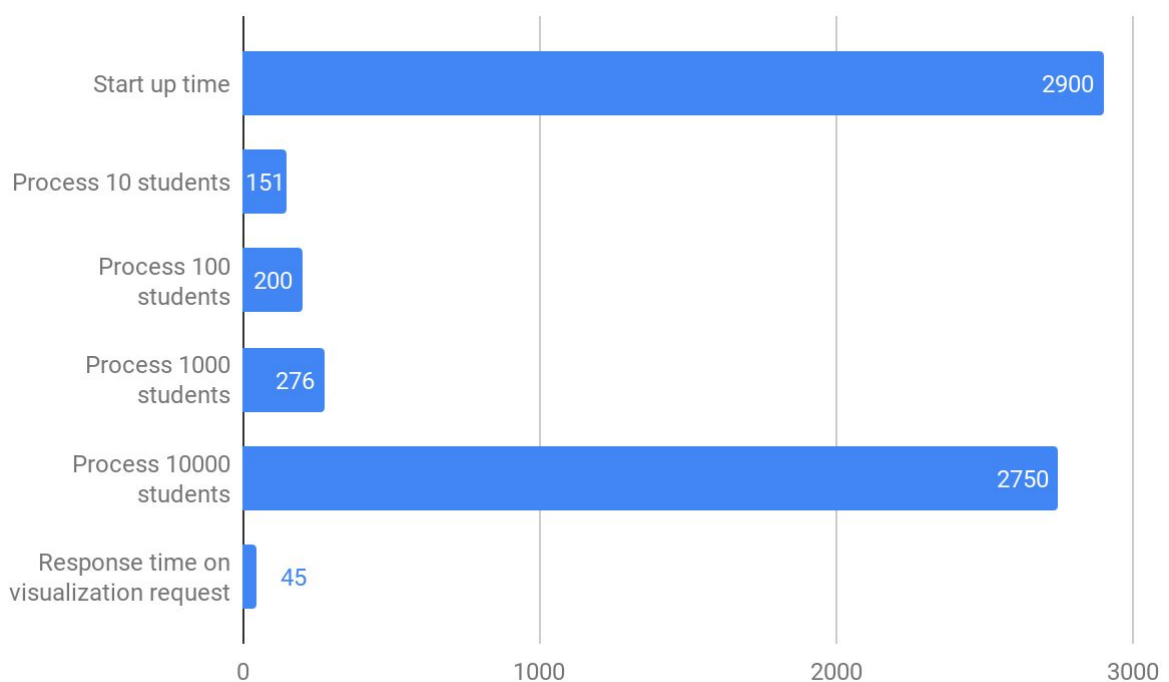


Figure *.6 - performance results from artifact activities. The horizontal axis represents time in milliseconds.

The results in themselves indicated that the artifact slowed down with scale. If this duration continued in the third iteration, we would have to look into applying threading in the system.

6.6.4 Heuristics Extraction:

After the evaluation of the artifact, we also made a point out of generating contingency heuristics for the following iteration. To do this, we consolidated the notes from the Use Case Evaluations and attempted to look for common themes and descriptions. A quick tool we used for sensemaking was to make a word cloud out of the sentences in the notes. The product of this word cloud is shown in figure 6.6.

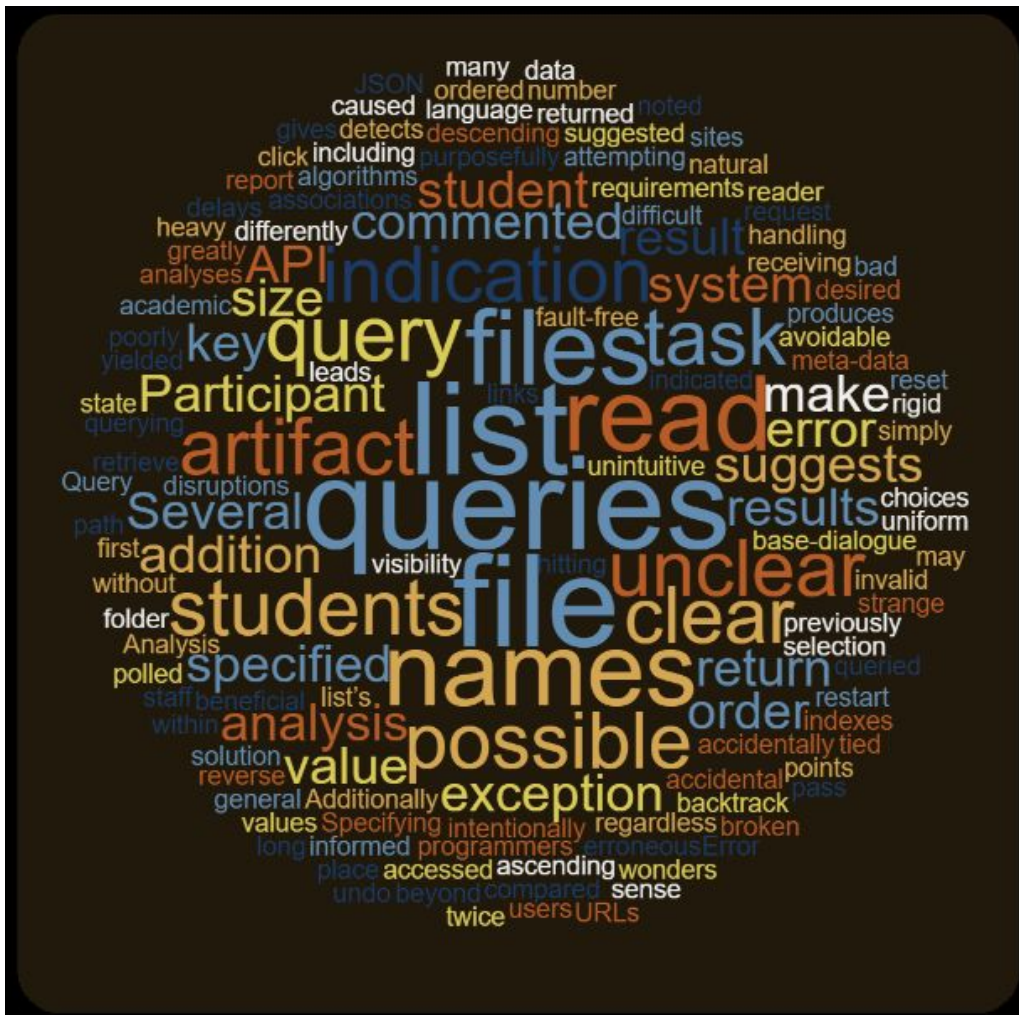


Figure 6.7 - Word cloud of the UCE notes.

Using the word cloud as a guiding heuristic, we sorted the comments as best as possible to common topics. The resulting contingency heuristics were found on a basis of this evaluation.

CH2.1:

Query parameterization and error-handling needs to be improved with consistent parameter arrays and useful recovery aids.

CH2.2:

Return values for queries need to be consistent across the system, including conventions and data structures. It should not be necessary to code different solutions for very similar queries.

CH2.3:

File analysis should be separated and made asynchronous from interactions with the API. In addition, the feedback of this process should be legible and clear in the API.

CH2.4:

It should be possible to revert the system to a state before a file was selected and analyzed.

CH2.5:

Files that have already been analyzed should clearly be indicated as such.

Generally, the artifact was troubled with flaws that we attribute to my own inexperience with making REST APIs. With the contingency heuristics we identified from the evaluations, we had some clear guides for developing the artifact itself. However, the artifact had not at its current stage satisfied the problem as earlier identified. Only to a degree had it solved some of the structural needs to satisfy the problem.

Further development of the artifact would have to address the academic content of the visualizations. At its current state, the artifact had a flexible model for file consumption, but without any prior knowledge of the layout of the datafiles, building visualizations would be infeasible.

However, a segment of the problem had been satisfied. The artifact had now established a relation between the Akademix Log Database and the OsloMet teacher staff.

Thus some final contingency heuristics were formalized to address the direction the artifact had to take:

CH2.6:

The artifact cannot at this moment satisfy the stakeholders Learning Analytics needs.

CH2.7:

The developments on the artifact has changed and reduced the problem's scope.

With contingency heuristics identified, we concluded this iteration.

7 Third Iteration

. In this iteration, we had established contact with the teaching staff of the OsloMet course. This meant we could conduct planning game operations.

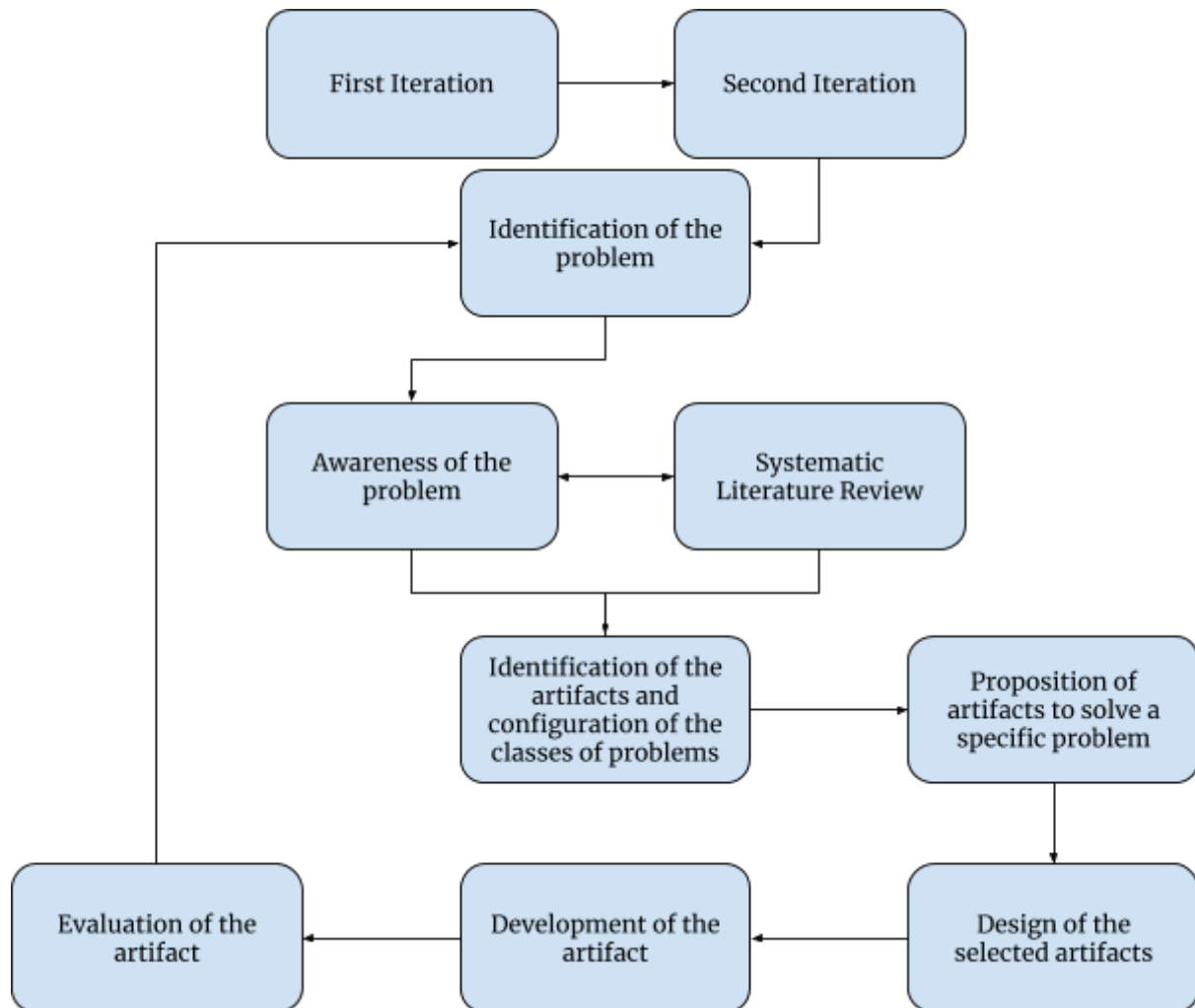


Figure 7.1 Illustration of the project progress in this iteration.

7.1 Identification of the Problem

With several contingency heuristics derived from the second iteration, there were plenty of problems to address with the artifact itself. In this iteration, the problem could be narrowed down to address the specific learning analytics needs.

As it stood now, the problems were specifically:

- The stakeholders need to identify the Learning Analytics needs they have.

- The artifact has to change to meet these needs.
- The artifact has to be able to consume the dataset from Akademix.

7.2 Awareness of the Problem

In this iteration, the problem had been refined and it became pertinent to describe the systemic operations of the artifact and how it related to the teachers operations. It would also be crucial to understand the teachers general workflow in relations to the course, and how they could best employ the visualizations that they were supposed to receive.

In addition to the relations of the artifact's functions and the operations of the Teachers, it is necessary to understand what traceable actions Learners make in the system. With an oversight of how these operations relate, it might be easier to identify where the artifact's visualizations can be used.

Without an artifact in place, the feedback cycle between the Teacher and Learner is manual feedback, as shown in figure 7.2.

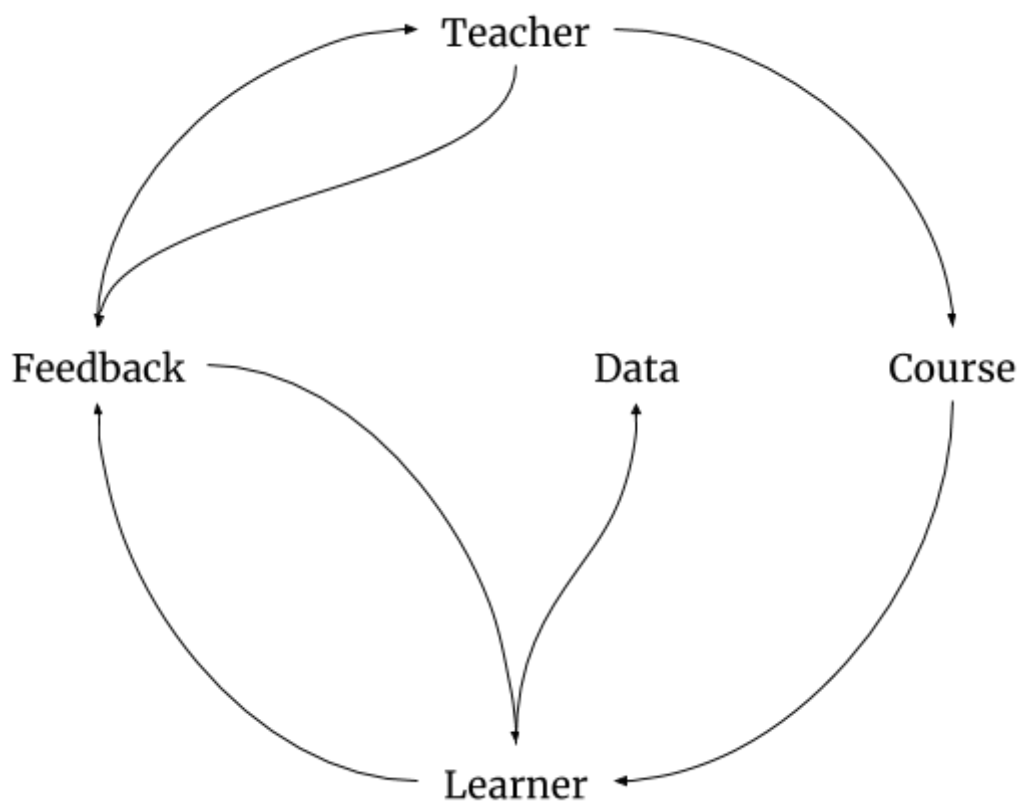


Figure 7.2 Illustration of the feedback between Teacher and Learner.

As shown, the Learner produces data, but this is not caught in the Feedback loop that exists between the Learner and the Teacher. Otherwise, the Feedback the Learner provides can be used by the Teacher to influence the Course.

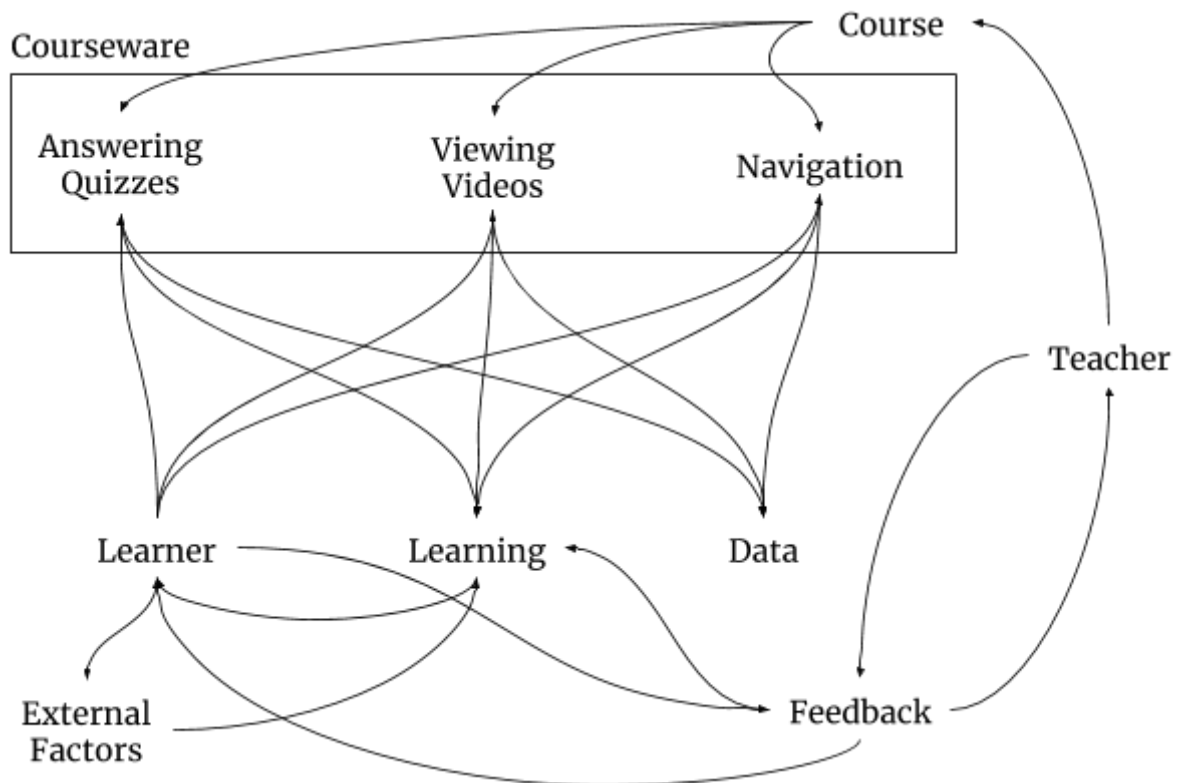


Figure 7.3 - Illustration of the interactions between the elements in the learning environment.

As shown in figure 7.3, there are several impressions that could plausibly lead to learning. The figure also shows that there are elements involved in the Learner's learning process that do not leave direct data traces. Activities in the classroom or activities such as reading offline curriculum do not leave data traces, which is an important aspect to consider.

7.2.1 Assessment of the Data

In addition to having established contact with the stakeholders, we had received data from Akademix. The data attributes were the following:

- line - the line number.
- accept_language - the client language.
- agent * - information about the browser-type being used.
- context * - the context which is packaged with the server request.

- event * - a packet that holds event information, such as video slide-seek or if a button was clicked.
- event_source - the source of the event, either the browser or the server.
- event_type - the type of event.
- host - the server's address.
- ip - the server's IP address.
- name - the event name, occasionally blank.
- page - the page the agent is acting on.
- referer - the page the agent is making the request from.
- session - the id of the session the agent is currently in.
- time * - date and time for the agent.
- username - the name of the agent.

Several of the fields above are atomic, however some fields hold compositions of data (these are marked with asterisks). Most of the atomic data could be subjected to simple analysis. For example, the amount of occurrences a certain username appeared, or which sites the user had visited. The composite data fields appeared to have formatting problems, such as invalid JSON syntax.

With some simple data analysis scripts, we evaluated the composite data fields for inconsistencies and found that n in x contained syntax errors of some sort. These data instances had to be repaired to some degree. When repaired, we found that the event and context objects contained information about how the user had interacted with videos and the multiple-choice quiz elements.

Many of the fields were not particularly useful at a glance either. For example the "accept_language", "agent" or "ip" fields did not change across lines and did not reveal anything interesting. For this reason, these fields were simply discarded.

The review of the data allowed us to make judgements about what types of visualizations were possible to make. These findings were also communicated to the stakeholders, such that we could develop user stories which were feasible.

7.3 Proposition of Artifact

With the perspectives gained from the previous steps, and following a meeting with the stakeholders, we were able to establish some user stories. The planning sessions omitted making a release plan with concrete dates, since we both understood that the work was exploratory research.

The user stories we developed were as follows:

- A. As a user, I want to track a student's progression from day to day.
- B. As a user, I want to see the student's amount of activities per day.
- C. As a user, I want to see the student's amount of time spent per day.
- D. As a user, I want to see how many times a video has been opened.
- E. As a user, I want to see how time is spent on each video.
- F. As a user, I want to compare time spent on videos associated with each module.
- G. As a user, I want to see if there are differences between movies with one or more persons as subject.
- H. As a user, I want to see the amount of attempts a student has made on quizzes.
- I. As a user, I want to see the amount of quizzes a student has completed.
- J. As a user, I want to compare the students in time spent on quizzes.

The above user stories had three distinct themes, which we decided could be divided into separate iteration plans. For the current iteration, the stories A, B and C became the main focus. We also emphasized in the proposition that the rest of the user stories inform the design of the artifact, such that significant changes could be avoided later in the research lifecycle.

7.4 Design of the selected Artifact

The new user stories did not pose a particular challenge to the infrastructure we had in place. However, the new data format required some changes in the data storage and modifications to how the data was read. With the fields that had composite-data, we specified a object-model that could read the data points into a dictionary.

With user stories A through C selected for this iteration, translating them to tasks seemed to be relatively simple. For each of the tasks, a query would have to be developed to

retrieve the relevant data. However, the more demanding part would be to determine the best way to visualize the data.

New tasks were generated based on the user stories and preliminary discussions from the planning session:

- A. Formalize classes for visualization pipeline.
- B. Establish necessary data, queries and lambda-operations.
- C. Survey matplotlib library for suitable visualizations.
- D. If no visualizations found, create a custom visualization.

In addition to these tasks, there were also technical debt tasks from the previous iteration that had to be addressed. These were also added to this iteration in order to pursue performance excellence for the artifact.

In addition to having established tasks, we created some specifications for the classes that we expected to need in order to satisfy the user stories.

7.5 Development of the Artifact

Implementing task A proved to be a fairly simple procedure. We kept the class structure to a minimum and focused on specifying types and custom methods that could be injected into the API framework. This dependency injection allowed for minimal change in the existing codebase. We discovered that Task A partially depended on Task B, so we decided to put Task A on hold and complete Task B first.

We specified queries that would retrieve all activities tied to each username, which were then ordered by datetime. For Story A, we first attempted to simply accumulated activities per date. The resulting data set was then tested out using spreadsheet diagrams to see if they satisfied basic visualization needs, as shown on figure 7.4.

| Month X | | | | | | |
|---------|----|----|----|----|----|----|
| M | T | W | T | F | S | S |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| 29 | 30 | 31 | | | | |

Figure 7.4 Sketch of calendar heat map visualization.

Having finished task B, we returned to task A and generalized the code in task B to satisfy this task. The refactor allowed for an abstract class to be created with a set of static functions to call on.

Task C and D resulted in a custom visualization. After having it consume the test set, it produced the following visualization:

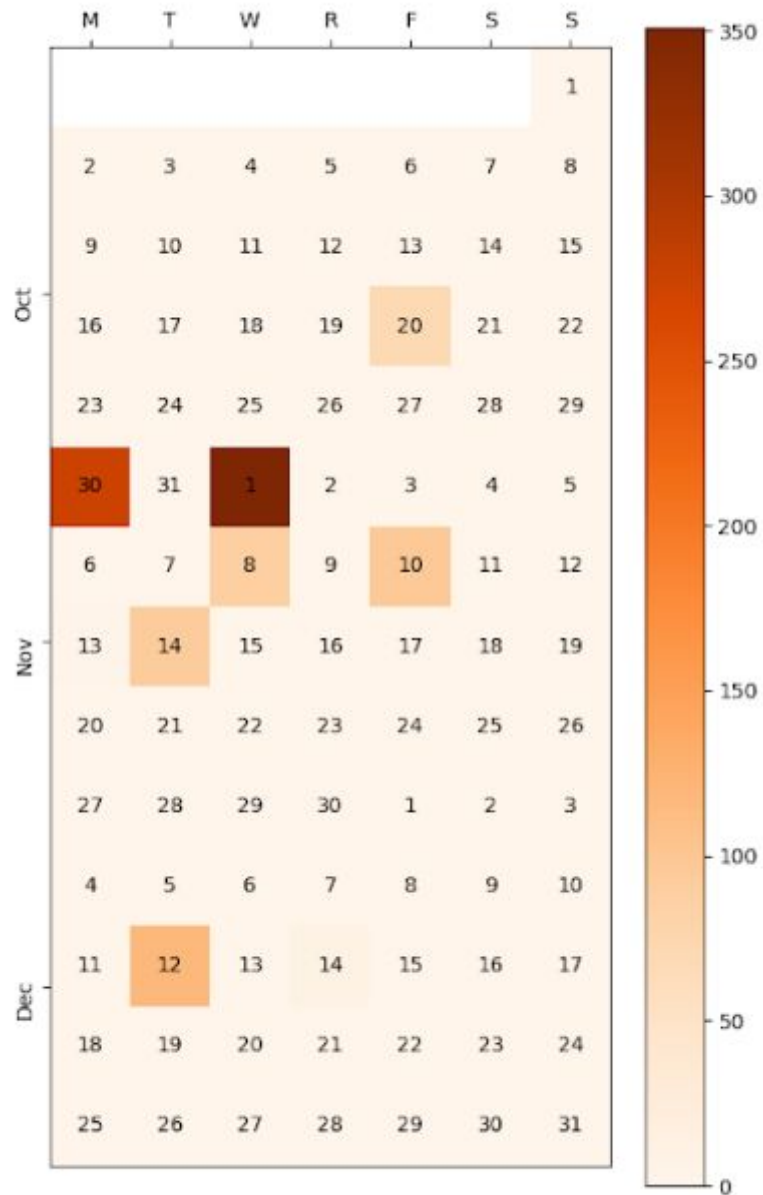


Figure 7.5 Illustration showing registered activities on given dates.

With the visualization code completed, we conducted a refactoring process to account for the contingency heuristics from the second iteration and to smooth out the newly created code as best as possible.

7.6 Evaluation of the Artifact

Aside from responding to specifications from the stakeholders, we wanted to make sure that the artifact still had a potential to be generalized. If the visualizations followed a very distinct visual vocabulary that only we and the stakeholders shared, then it could potentially limit the artifacts broader usefulness. For this reason, we recruited a cohort of ten participants to participate in System Usability Scale surveys. In addition, we executed a Performance test in the same style as in the second iteration.

The System Usability Scale surveys were implemented in two phases, where we first showed the participants a selection of visualizations and how they were accessed. we informed the participants that * Afterwards they filled out the survey.

7.6.1 System Usability Scale results

The System Usability Scale results came as a surprise. The average usability score sank by 4,75 points. The participants were not the same as the previous survey, so there were no expectations of continuity in that sense.

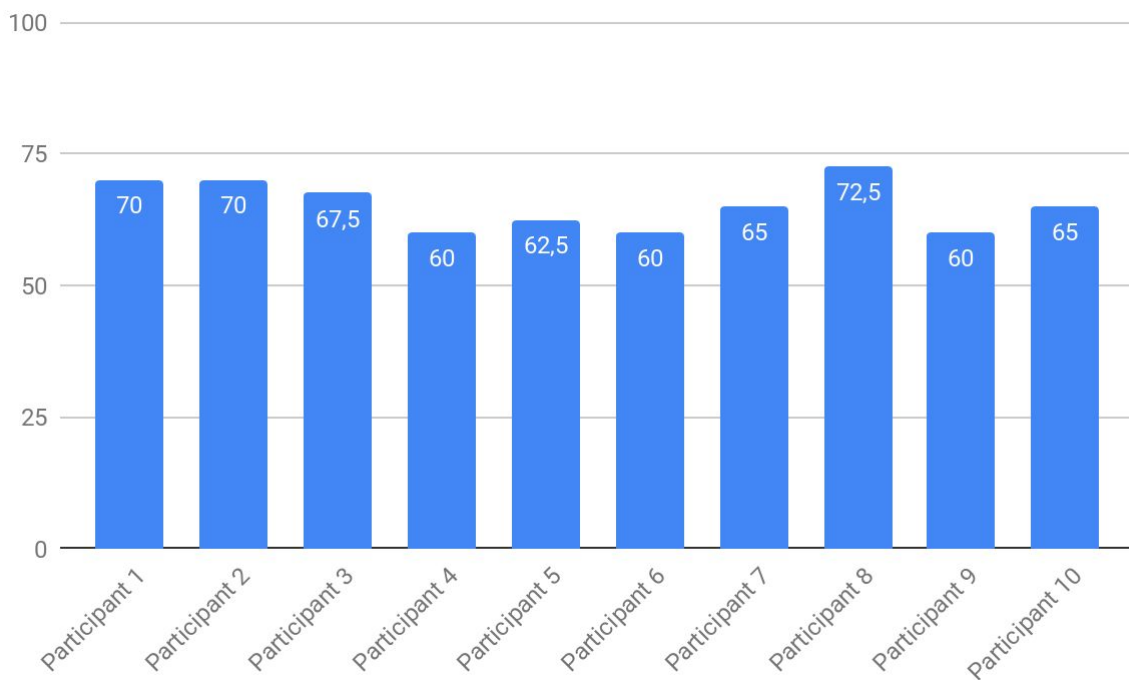


Figure 7.6 Results from the System Usability Scale survey.

Due to the average having settled on 65,25%, we looked into the averages for each question as well. In a high-scoring average System Usability Scale survey, the difference between each of the questions should be high, so we looked for the inverse.

Average vs. Question

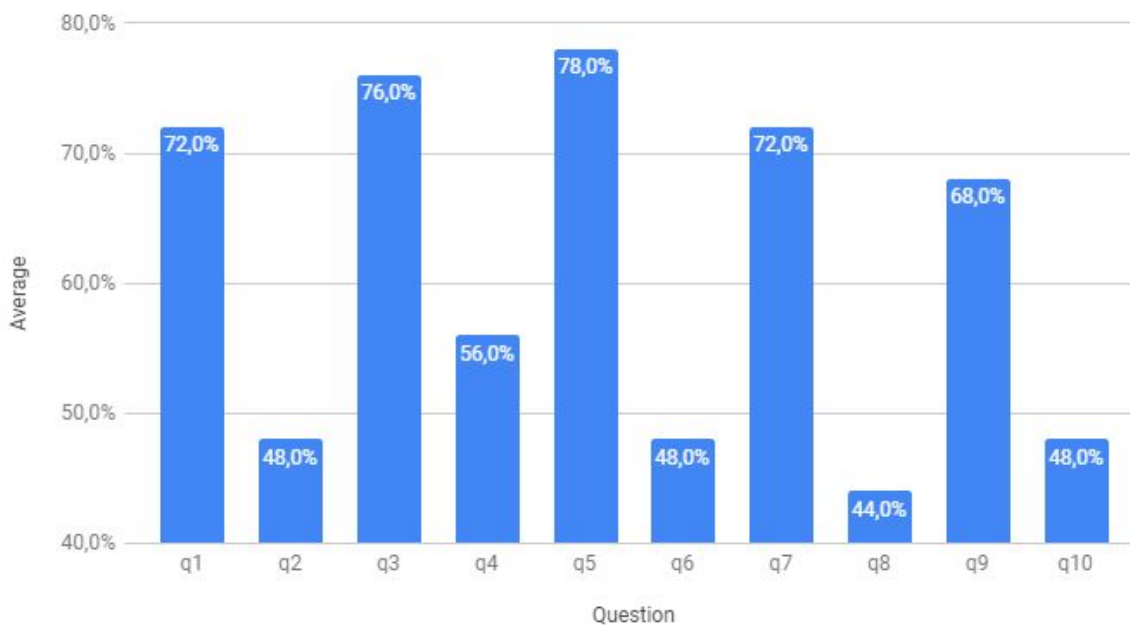


Figure 7.7 The average response valuation for each questions. Odd-numbered questions are positively toned, while even-numbered questions are negatively toned.

The questions with the lowest difference were Questions 3 and 4, and Questions 9 and 10.

As a reminder, those questions were:

q3) I thought the system was easy to use.

q4) I think that I would need the support of a technical person to be able to use this system.

q9) I feel that the visualizations help me with my work.

q10) I do not think I will use the visualizations to inform my opinion.

While most negatively toned questions were below the 50% mark - something we considered positive - the fourth question raised some concern, although not as much as the responses to the ninth and tenth questions. If the artifact is supposed to be used in an Learning Analytics context, then having a more distinct difference between these questions is important.

Question four scored relatively high compared to the other negatively phrased questions. Given the nature of the artifact, this did not come as a surprise. A REST API is not generally accessible unless the user is familiar with HTTP requests.

7.6.2 System Performance Results

Unfortunately, the implementation we chose for the visualization caused an increase in processing time. This was not unexpected, as the matplotlib module did take some hundred milliseconds to complete. This also influenced the results of the scale tests, showing that the artifact would not perform well with a large volume of students.

Running the system performance tests showed the following results:

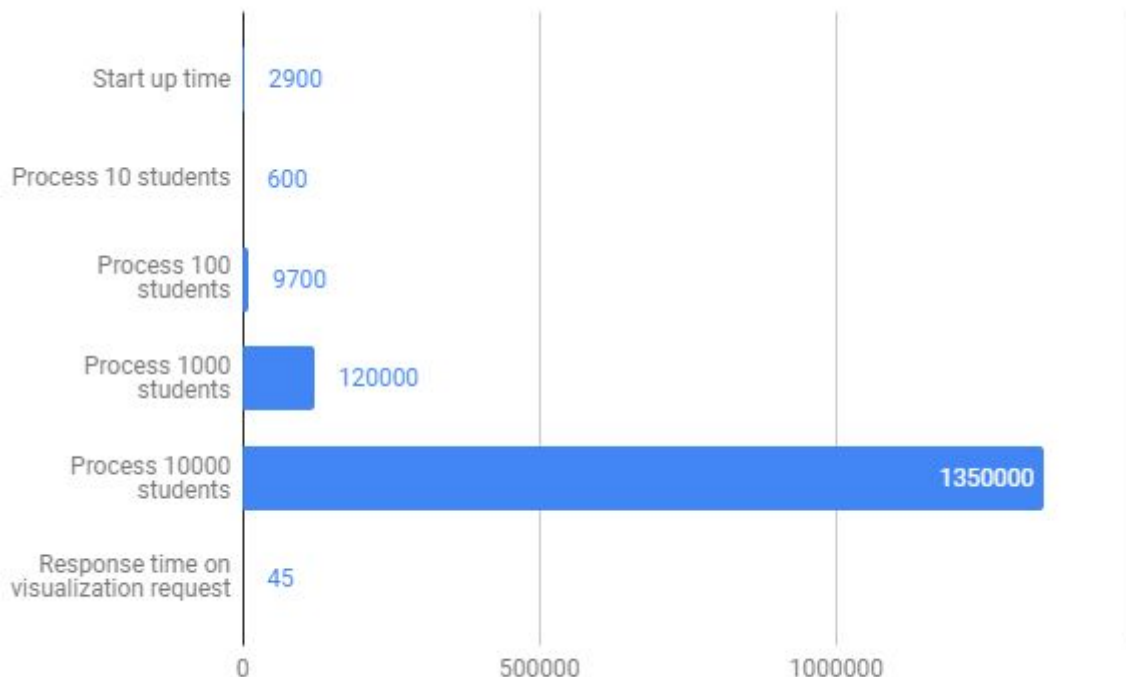


Figure 7.8 Diagram of system performance.

This test showed that the artifact handled volume scaling badly. Even at 1000 students, the processing time took 2 minutes. 10000 students required 22 minutes and 30 seconds to complete. With some breakdown of the tests, we found that matplotlib was the major constraint on the system's processing time. Despite each image being rendered in less

than a second, or 135 milliseconds on average, the sheer amount of students prompted a considerable total render time.

Considering that this performance test was only subject to one visualization, and that there would be more more visualizations to add, then the process time would most likely be quadratic. In future iterations, the artifact might use several minutes to complete an analysis of thousands of students, if not hours or more.

Feedback from Stakeholders

The visualizations were exported into folders ordered by student names and sent to the stakeholders for review. Afterwards we scheduled a skype conversation to conduct a semi-structured interview. In general, the visualizations were received with positive sentiments, but some criticism was offered as well.

The main points of criticism were:

1. Each visualization had unique gradients and color map scales. This made them hard to read.
2. Certain visualizations were difficult to read due to the coloring.
3. It was difficult to interpret exactly how many activities were observed on a given date.

The stakeholders also noted that certain students were active immediately following or on the same dates as the classroom lectures. We found this heuristic to be interesting, and as such we codified it in the heuristics extraction.

7.6.3 Heuristics Extraction:

In this iteration, the evaluations provided less information to translate to contingency heuristics than in the previous iteration. Despite this, there were still some important lessons that could be translated to contingency heuristics.

CH3.1:

56% of the SUS survey participants responded that they would need support to use the system. Can the artifact be simplified?

CH3.2:

The overall helpfulness or decisiveness of the information presented was rated poorly compared to other aspects of the artifact. How can we increase this score?

CH3.3:

The artifact is slow to process high amounts of students. Is it possible to reduce the processing duration, or is this a limitation in the artifact?

CH3.4:

The activity visualization we had produced revealed behaviour patterns in the students. Some students were particularly active on the same dates as the physical classroom sessions.

8 Fourth Iteration

The fourth iteration began on the 22nd of January 2018.

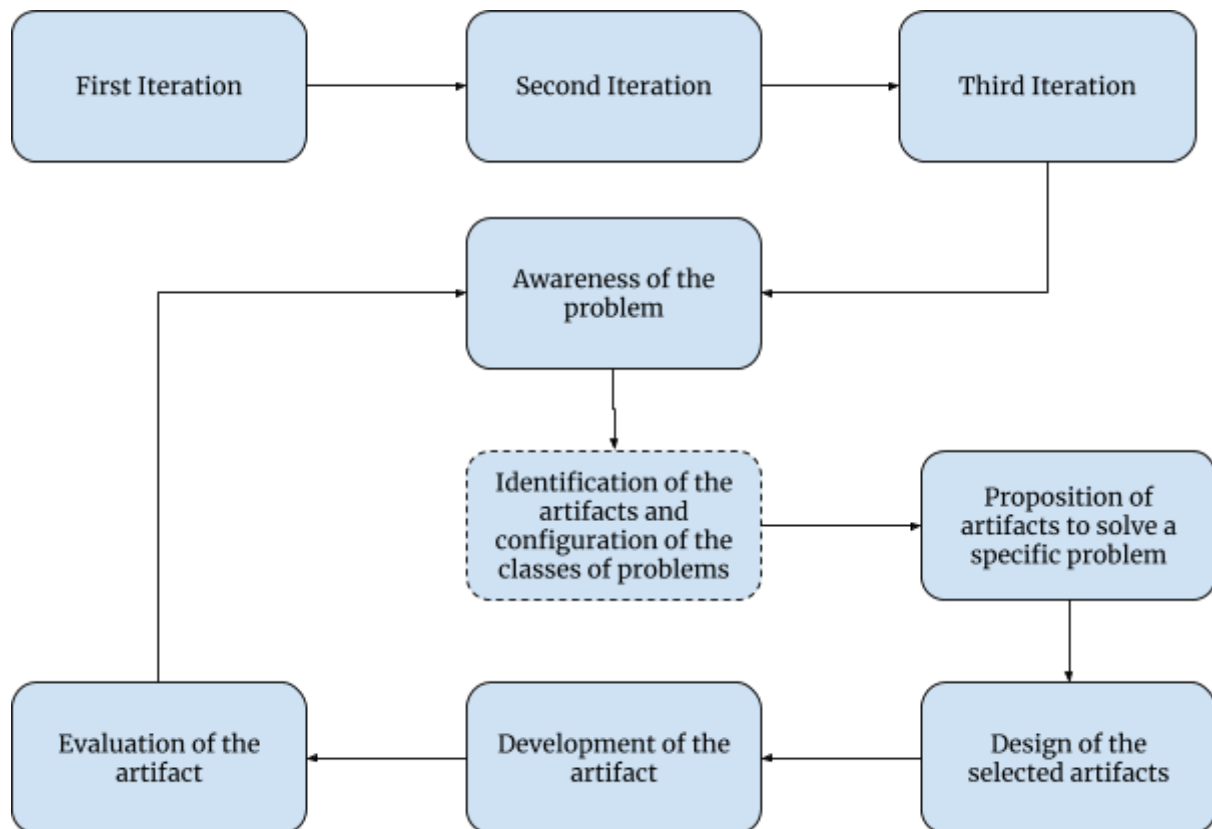


Figure 8.1 Illustration of the project progress in this iteration.

8.1 Awareness of the Problem

From the previous iteration, we had learned that the artifact already had a scaling problem. For the purpose of addressing this, we looked into options for implementations of threading or multiprocessing. After reading upon documentation and responses to similar problems in sites like Stack Overflow, we reached a conclusion that the scaling issue would have to be postponed. Since the course the stakeholders were hosting had a maximum of 20 students, they did not need the capacity to tackle thousands or even hundreds of students.

Another contingency heuristic that seemed important was CH3.2, which indicated low helpfulness or support for decision-making. In conference with the stakeholder, we decided that the issue could be set to a low priority and rather be assessed once we had more visualizations to consider and compare.

Finally, CH3.1 indicated that the majority of the users might require technical support. This perspective on the system made us prioritize looking into possible ways to make the artifact simpler to use. The stakeholders made a request for a more interactive solution as well, although we both agreed that this could be introduced at a later in the development cycle.

With the contributions of the contingency heuristics assessed, the research project still had a backlog of user stories which had not been satisfied. Most of the awareness we built during this stage turned out to be more relevant for a later iteration.

8.2 Proposition of Artifact

For this iteration, the topic of video interactions became a priority. That is, the following stories were selected for the artifact:

- D. As a user, I want to see how many times a video has been opened.
- E. As a user, I want to see how time is spent on each video.
- F. As a user, I want to compare time spent on videos associated with each module.
- G. As a user, I want to see if there are differences between movies with one or more persons as subject.

8.3 Design of the selected Artifact

During this iteration, the design stage revolved mostly around adding new visualizations to the artifact. The previous iterations had put in place the infrastructural and software engineering needs for the artifact

Based on the user stories, we created the following tasks:

- A. Create queries for video events.
- B. Create video sorting algorithms.
- C. Create visualization that shows amount of times a video has been opened.
- D. Create visualization that shows viewing time per video.
- E. Create visualization that lists video viewing time, grouped after module.

- F. Create visualization that shows one student's video activities compared to a variable other student.

User Story G was a bit difficult to interpret, and as such we decided on a flexible visualization scheme that would allow the stakeholders to compare the video activities of two students. Due to the response times between the researchers and the stakeholders, we decided to move on with the development and rather handle clarifications whenever they would appear.

Tasks C through F only required to display a certain quantity per entity, meaning most visualizations in this iteration could be handled with simple bar charts. For these reasons, the design phase was not particularly long.

8.4 Development of the Artifact

The previous iteration had established that the artifact could reliably deliver visualizations. In this iteration, the analyses would revolve around gathering events related to the videos and extracting useful data. A typical example of video events is shown in figure 8.2. One of the issues we found with these types of events is that they did not always give consistent records. With some videos, we found that time-stamps were duplicates, while in other cases there were gaps in the play records.

```
{
  "code": "65LflKmgHug",
  "id": "3a41ad4a359849dca5fb2a3e378754d9",
  "currentTime": 91.33140802288818}
```

Figure 8.2 Example of a video event.

Due to these data inconsistencies, we realized that there was a significant amount of work in just establishing a consistent timeline for each video per user. One of the events we could replicate was how much a video was played in absolute play time.

Another issue we encountered was that there were no semantic identifiers for the videos in the data. To find legible titles for the videos, we had to manually search for them in the course-ware. Many of the video-IDs matched with videos hosted on YouTube, but the

course also had videos on other content platforms. We found bar charts to be best suitable for displaying play times for videos. Along with limitations in matching video identities, we also encountered a problem in finding the durations of the videos. In the initial visualizations, we had no alternative but to post accumulations of play time in seconds.

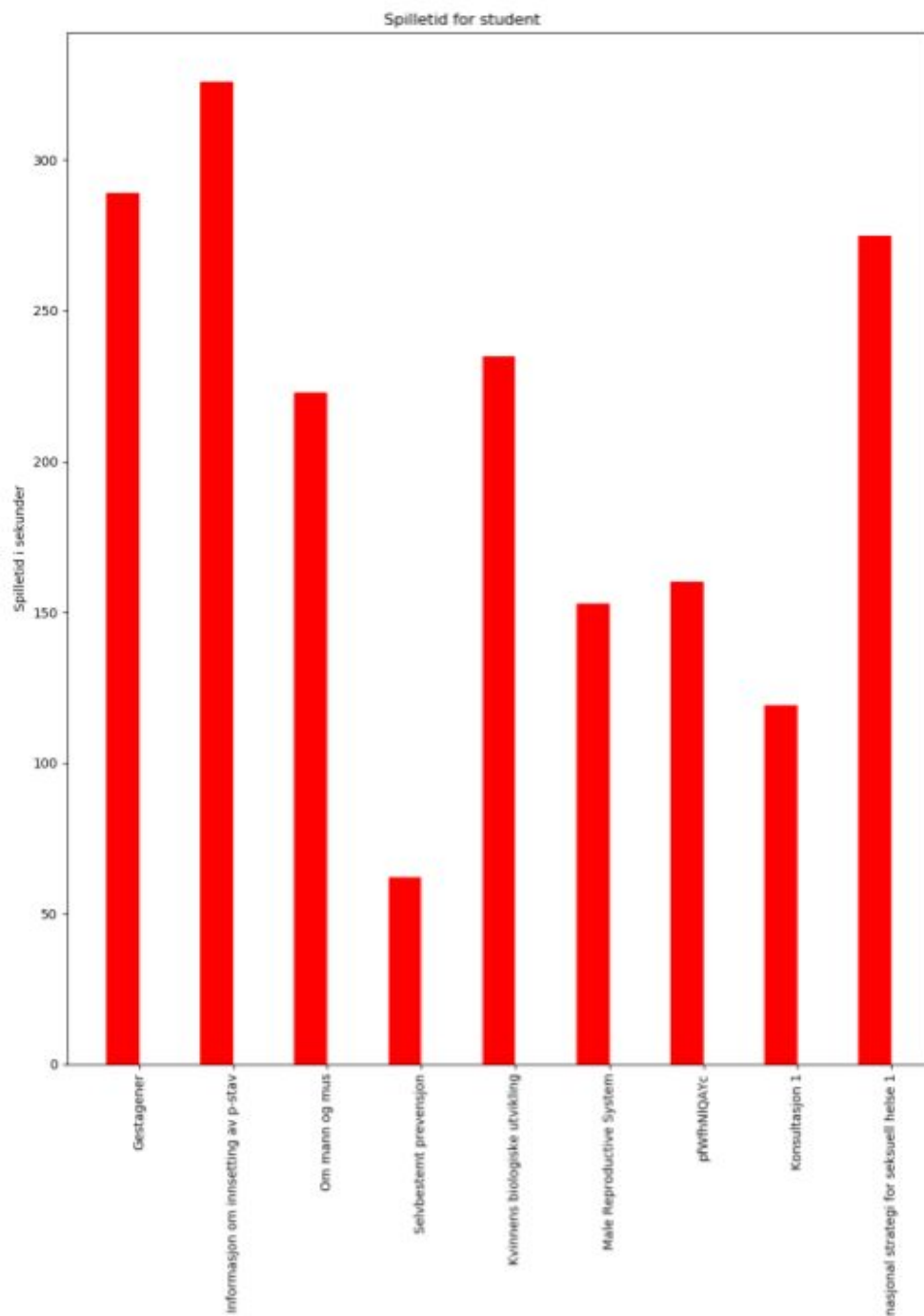


Figure 8.3 Visualization of video play times (in seconds) for a single student.

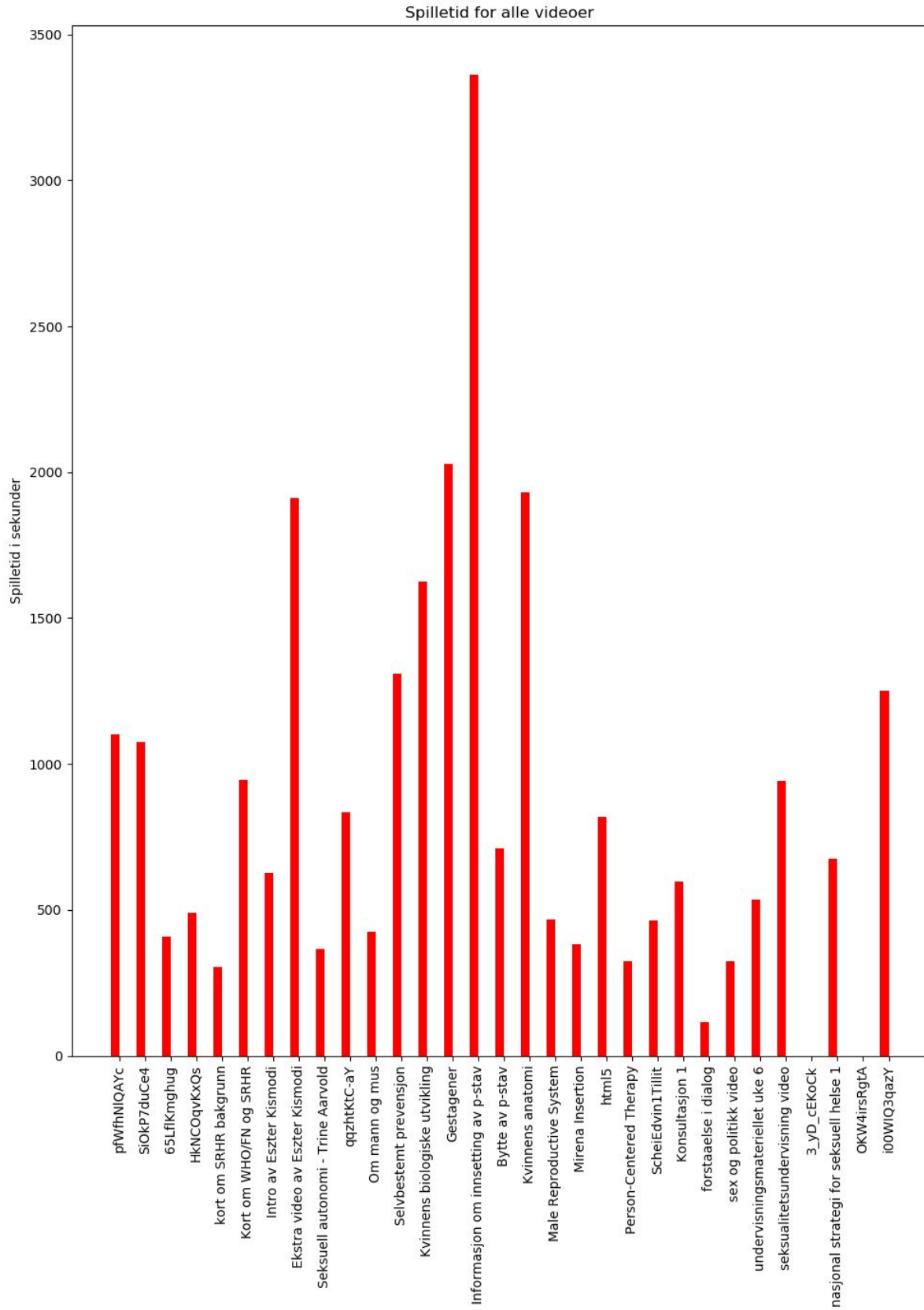


Figure 8.4 Visualization of total playtime for all videos.

Although we did not manage to resolve all of the video titles, most of them were included into the visualizations. This did reveal a potential gap in the alignment of motivations for developing courses and information visualization artifacts. Had the problem of video-IDs been identified earlier, we could have suggested some actions that could circumvent the problem. One such suggestion was to simply write a table in an spreadsheet style file, with a column for video-ids and video display titles.

User story G appeared to be difficult to satisfy, as several users had incomplete logs. In order to make visualizations that matched this description, we attempted to make one visualization that showed a bar per user for each video. After conversing with the stakeholders, this idea turned out to be less feasible and eventually we agreed to drop it.

8.5 Evaluation of the Artifact

As with the previous iteration, a System Usability Scale survey was implemented. We recalled the cohort we previously had recruited. The evaluation was carried out in two stages, first showing the new visualizations and then the complete set of visualizations, each time issuing a separate survey. The goal of this staging was to be able to assess the new visualizations on its own, before assessing them as part of the complete set of visualization. Hopefully, this would allow us to create a nuanced view of the artifact's development.

In addition to the System Usability Scale surveys, we also performed a Performance test.

8.5.1 System Usability Scale results

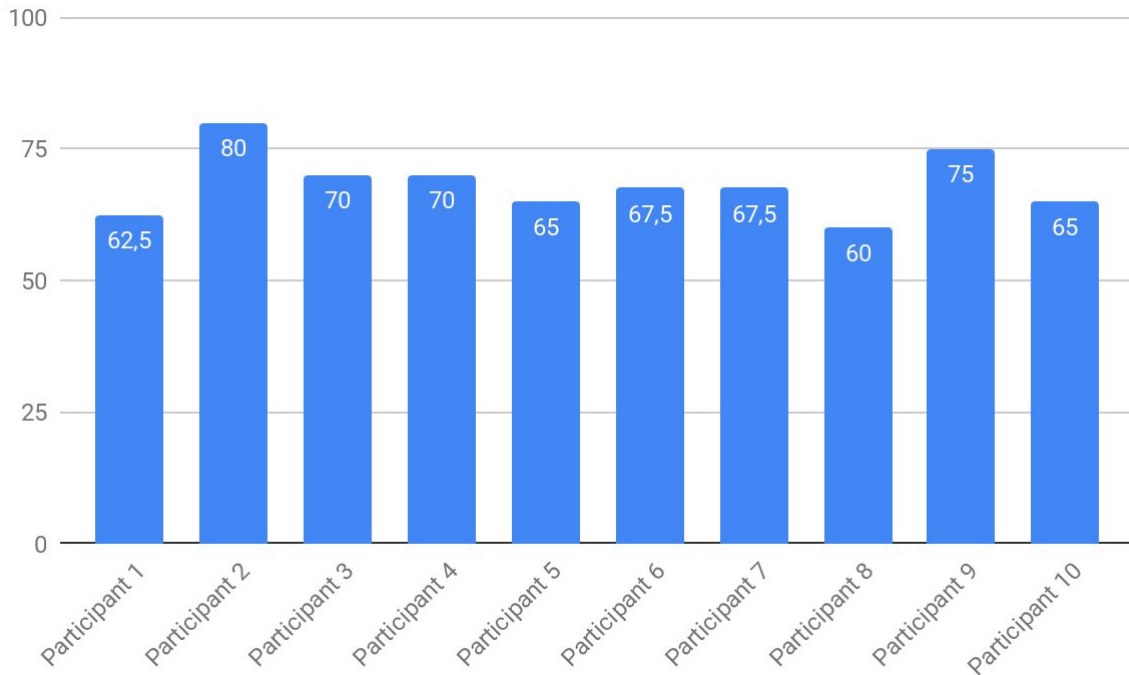


Figure 8.5 - System Usability Scale survey based on video visualizations.

The video visualizations scored higher than the activity visualizations from the third iteration, with an average score of 68.25. This put the video visualizations squarely in an above average bracket, but we noted a problem had persisted from the previous iteration, which related to Question 4.

Average vs. Question

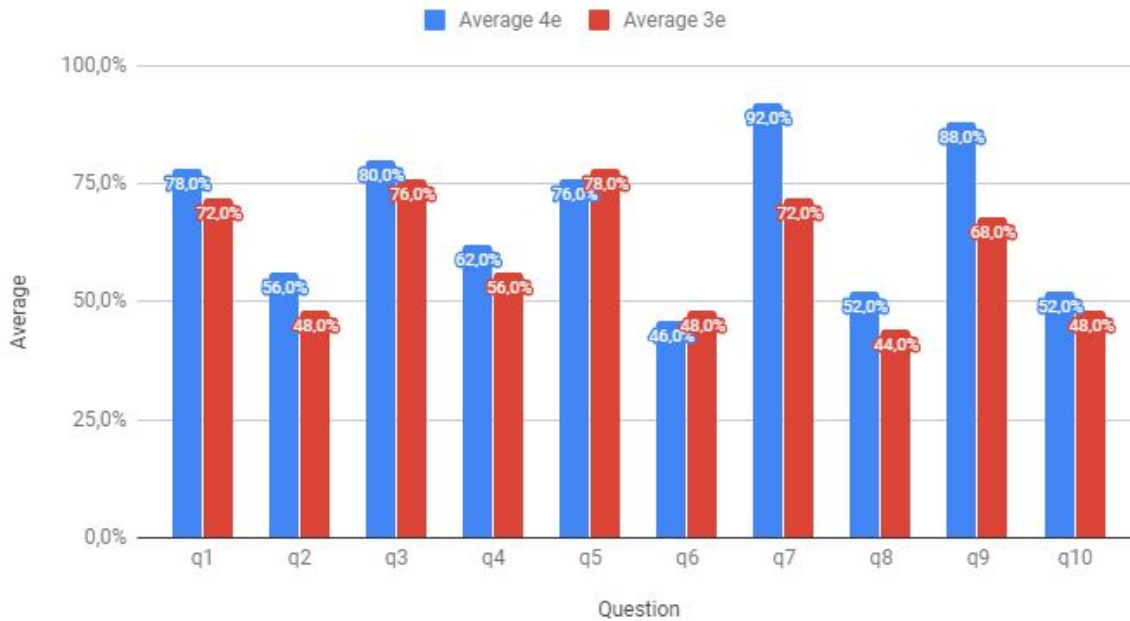


Figure 8.6 The average response valuation for each questions. 4e represents the values from this iteration, while 3e are the values from the previous iteration.

Based on the responses, we could tell that the participants found these visualizations to be more intuitive or meaningful, and that they would be helpful for the participants' work situation. We noted that the participants were responding consistently with questions 3 & 4 and 9 & 10, as with the previous iteration.

Iteration comparison

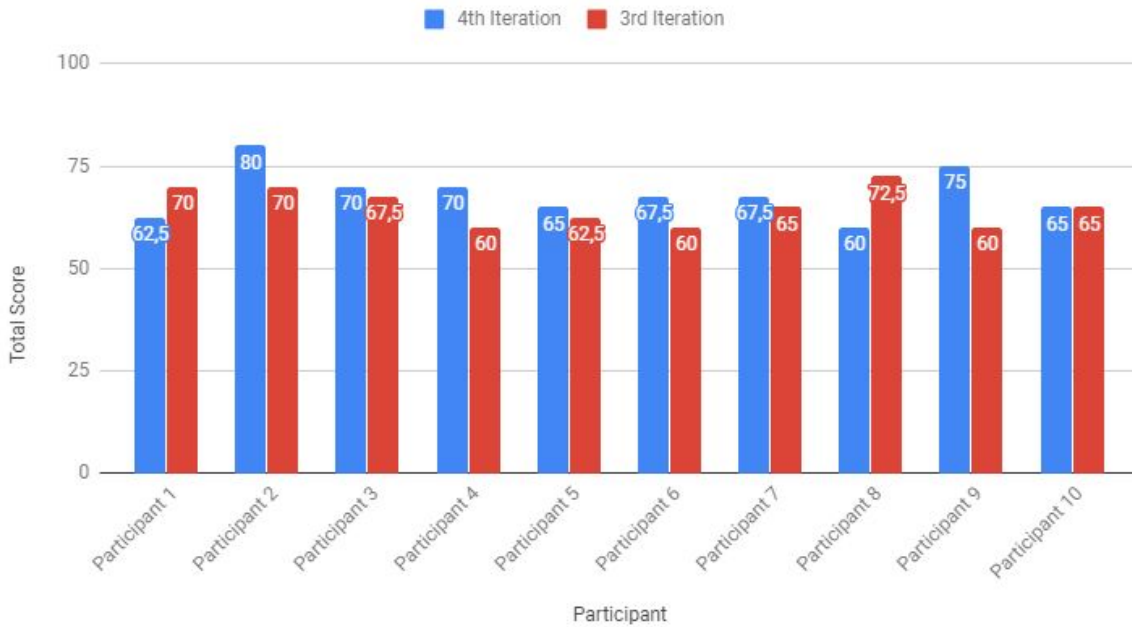


Figure 8.7 Activity results (red) and video results (blue) in comparison.

We noticed with the results that the participants were slightly more approving of the video visualizations. Despite having two participants with a declining approval, there were also two participants with a significant increase in approval that led to a net increase in approval.

We also implemented an evaluation to assess the artifact with both visualizations active. In this survey, the results marked a steady increase from both the previous iteration and just the video visualizations. The average had risen to 70,5, which placed the artifact well into the 'above average' bracket.

4th Iteration vs. 3rd Iteration

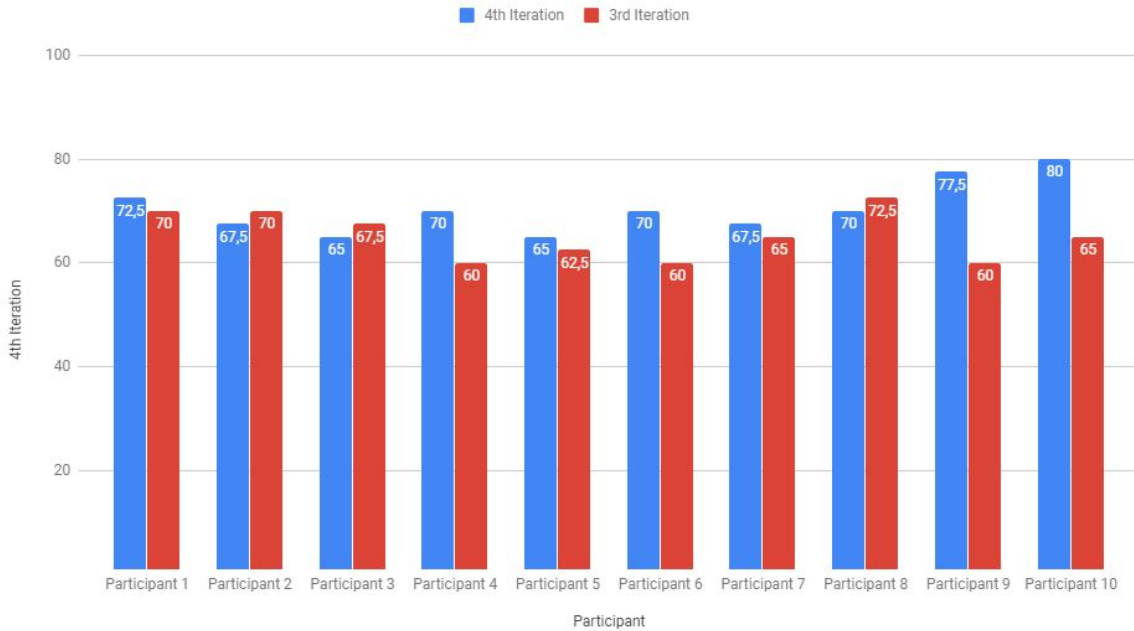


Figure 8.8 Complete survey of the 4th and 3rd iterations.

4th Iteration vs. 3rd Iteration

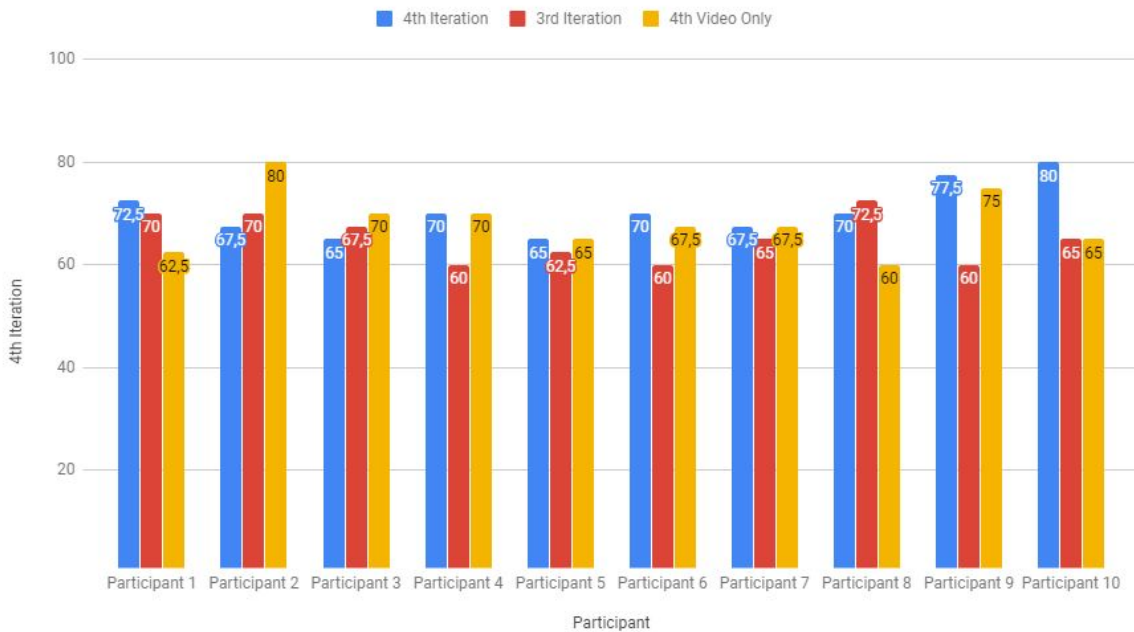


Figure 8.9 Comparing all survey results taken for the third and fourth iteration.

Although the surveys indicated a rise in usability rating, the evaluations did not reveal much new information. However, with another comparison of how the questions were answered, we saw some new indications in the participants' perception of the artifact.

Iteration Question Comparisons

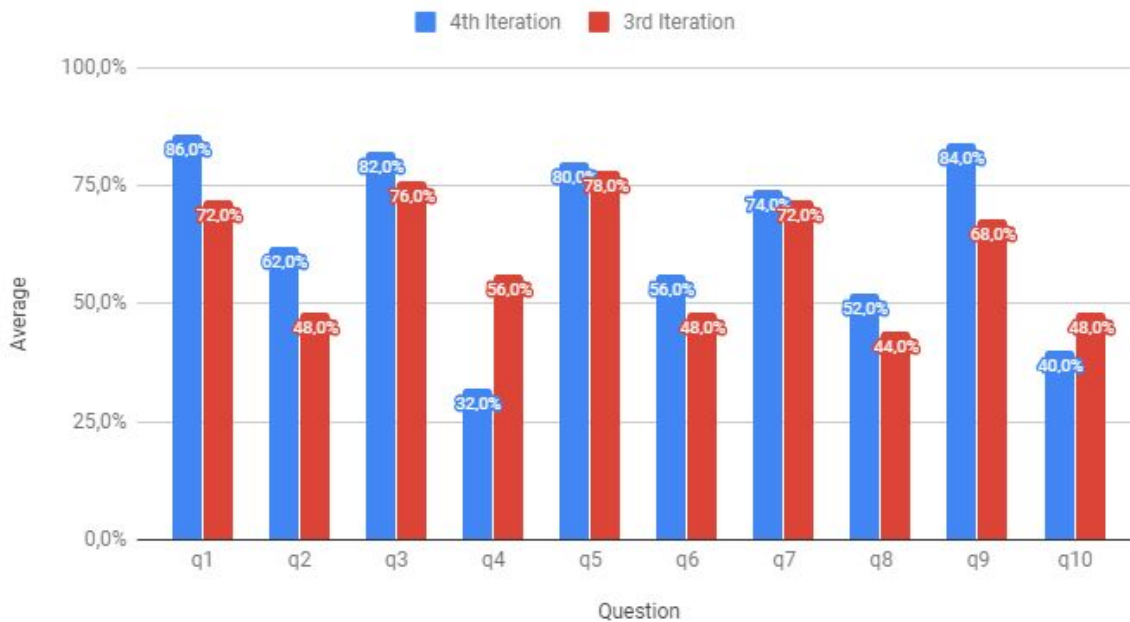


Figure 8.10 Comparing rating scores for the survey questions between iterations.

Three new issues appeared, as negatively formulated questions had risen above the 50% line. According to the System Usability Scale survey for the 4th iteration, the artifact was now likely to be perceived as “unnecessarily complex” (q2), some of its visualizations “unclear or noisy” (q6) and that the data needed to be “better presented” (q8).

Another change between the iterations, even though no changes were implemented to deal with them, we noticed that questions 4 and 10 had declined. These questions had previously been noted as concerns, especially with question 4 reaching above the 50% line. One interpretation that could be made here is that the participants felt more confident in using the artifact after having been exposed to it a second time.

8.5.2 System Performance tests

In this iteration, the System Performance used considerable time in processing the student volumes. At 10000 students, the artifact used over 11 hours to finish all the visualizations.

Processing Time in Minutes

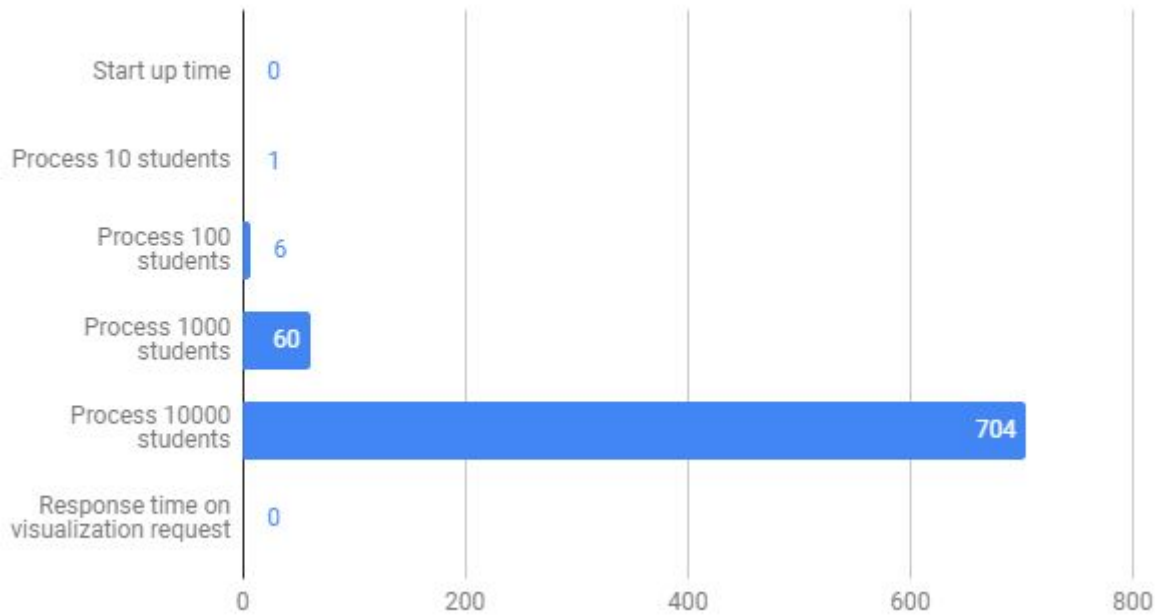


Figure 8.11 Processing times in minutes from the System Performance test.

At this point, it became more evident that the artifact would not be quite so capable of handling large volumes of students unless we invested time in parallelization or perhaps changing the method for producing the graphics. Since the artifact is aimed towards deployment on the web, a potential solution could be to shift towards producing vector graphics in a SVG format.

8.5.3 Generalization to Heuristics

< to be written >

9 Fifth Iteration

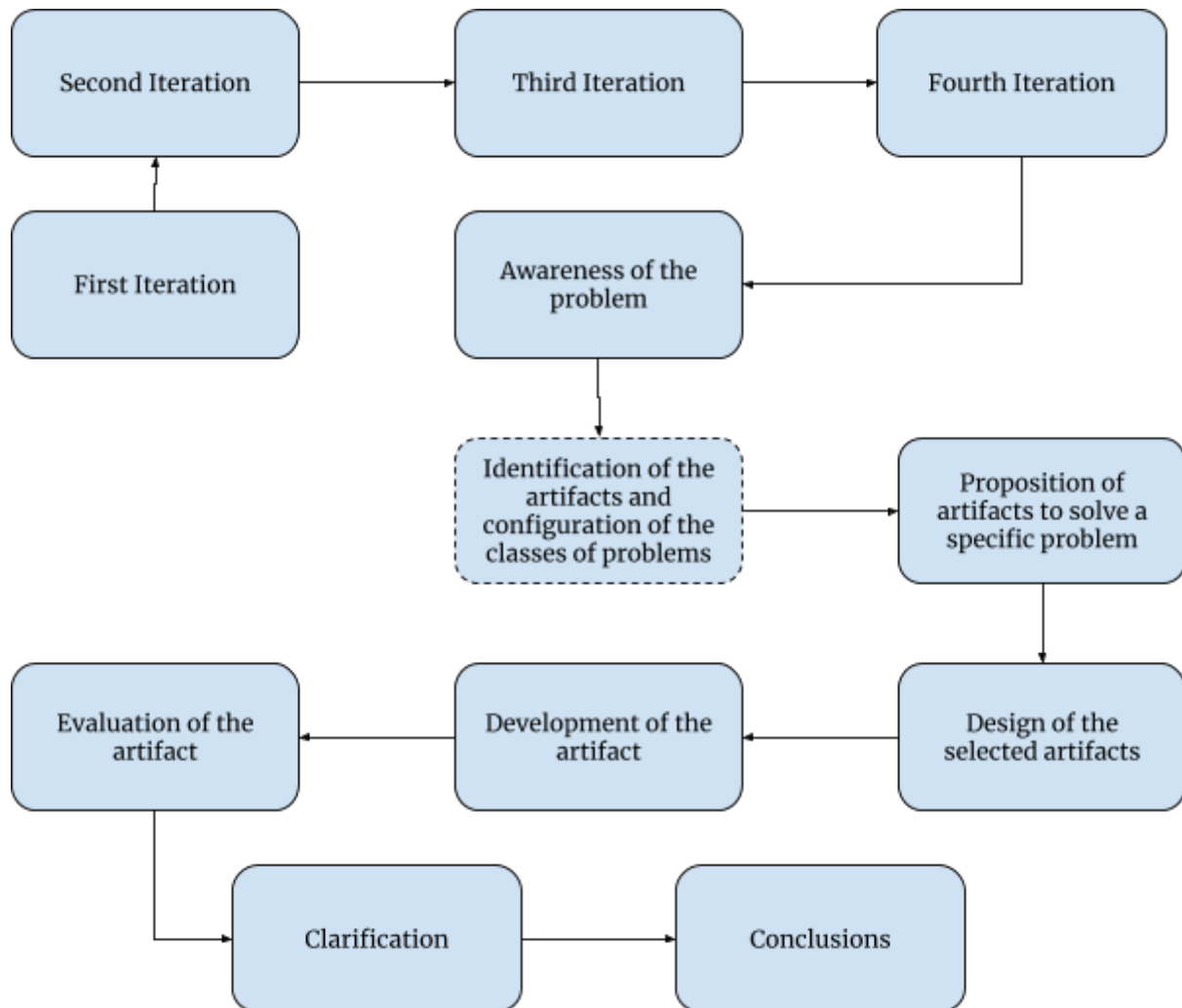


Figure 9.1 - Illustration of the project progress in this iteration.

9.1 Awareness of the Problem

The previous iteration produced six contingency heuristics. The first three of them were from the findings in the SUS survey for that iteration, which informed our design process as guidelines. They were clear indications that there was a moderate sense of noise in the visualizations, and that they had to be made as simple as possible.

Contingency heuristic CH4.4 highlighted the processing duration for the artifact. We decided to look into this by experimenting with using google charts. Since processing the

data itself did not require much computing time, the artifact could circumvent the long image rendering durations.

The remaining contingency heuristics, that is CH4.5 and CH4.6, were the basis for their own user stories. Given how the previous iterations had fared, we decided that these would have to wait until we could finish the first list of user stories. Still, we codified them to user stories in order to make informed decisions about the proposition of a new artifact design.

The new user stories were:

- K. As a user, I want to read video titles along the Y-axis of the visualizations.
- L. As a user, I want to see the duration of videos in the visualizations when relevant.
- M. As a user, I want to see where in a video timeline the most playback occurred.
- N. As a user, I want to see the visualizations in an on-site dashboard.
- O. As a user, I want to see how many interactions each syllabus-page has.
- P. As a user, I want to see how many interactions students have with off-site syllabus content.
- Q. As a user, I want to see the progress of a student.

9.2 Proposition of Artifact

With much of the video and activity analyses completed, the only remaining “family” of user stories was the quiz-related user stories. These became subject for the final iteration before the research was concluded.

The final user stories left were the following:

- H. As a user, I want to see the amount of attempts a student has made on quizzes.
- I. As a user, I want to see the amount of quizzes a student has completed.
- J. As a user, I want to compare the students in time spent on quizzes.

In addition to the user stories above, we also codified some user stories based on the contingency heuristics. These were not added to this proposition, but we informed our

design decisions to account for them where relevant. For example, we decided not to use vertical labels in any of our visualizations.

9.3 Design of the selected Artifact

In the same style as previous the previous iterations, we generated tasks based on what we experienced had worked with previous user stories. We knew at this point in the research process that we had little time and that there was not much room for experimentation. Just producing any visualizations would have to suffice.

The following tasks were generated from the user stories:

- A. Create queries that find:
 - a. All quiz activities.
 - b. Quiz activities grouped by users.
 - c. Quiz activities grouped by quiz-identity.
- B. Identify valuable data points in the quiz-related data.
- C. Create visualization algorithm for:
 - a. Amount of student quiz attempts.
 - b. Total amount of quizzes a student has completed.
 - c. Find the time a student has spent on quizzes.
- D. Explore quiz HTML-data.
- E. Test bar chart solutions.

We had previously assessed the data and found that there was a lot of excessive or verbose data. Fortunately, the user stories called for relatively simple insights, which meant that many of the analyses would be supported by simple counting functions.

9.4 Development of the Artifact

The manner which the quiz data was stored had similar issues to that of the video events. In addition to some meta-data, the data packets also contained entire chunks of HTML code. Considerable time went into expanding the data and applying sense-making. While some simple observations could be made at a glance, such as a sum of interactions with a

given quiz element, nuances such as completed quizzes or time spent became a protracted process.

Eventually we narrowed the focus in on activities with the 'problem_check' type. The data contained in those activities appeared to have the most useful data for our purposes. In figure 9.3 we found data such as the choice answer, whether the answer was correct, which site the question was checked and the grading values tied to the particular quiz.

```
input_5af208eb66154e52b7dd7069675abcd2_2_1=choice_0
```

Figure 9.2 Example of how a quiz attempt looks like.

```
{'answers': {'5af208eb66154e52b7dd7069675abcd2_2_1': 'choice_0'},
'attempts': 1,
'correct_map': {'5af208eb66154e52b7dd7069675abcd2_2_1': {'answervariable': None,
'correctness': 'incorrect',
'hint': '',
'hintmode': None,
'msg': '',
'npoints': None,
'queuestate': None}},
'grade': 0,
'max_grade': 1,
'problem_id': 'block-v1:HiOA+PREV6000+2017H+type@problem+block@5af208eb66154e52b7dd7069675abcd2',
'state': {'correct_map': {},
'done': None,
'input_state': {'5af208eb66154e52b7dd7069675abcd2_2_1': {}},
'seed': 1,
'student_answers': {}},
'submission': {'5af208eb66154e52b7dd7069675abcd2_2_1': {'answer': '22,8%',
'correct': False,
'input_type': 'choicegroup',
'question': 'Hvor
'mange av
'de
'norske
'kvinnene
'oppgav
'at
'svangerskapet
'ikke var
'planlagt?',
'response_type': 'multiplechoiceresponse',
'variant': ''}},
'success': 'incorrect'}
```

Figure 9.3 An example of a server data response to a quiz attempt.

By going through all the quiz data in the dataset, we found the following questions had been engaged with:

- A. En effekt av p-sprøyte er å:
- B. Hormonspiral virker ved å
- C. Hvor settes p-stav når kvinnen er venstrehendt?
- D. Hvor mange av de norske kvinnene oppgav at svangerskapet ikke var planlagt?
- E. Hvor mye større var sjansen for å rapportere om at svangerskapet ikke var planlagte blandt kvinner som hadde opplevd nylige seksuelle overgrep (innen det siste året) sammenliknet med de som ikke hadde vært utsatt for overgrep?
- F. Hva var karakteristika hos kvinner som rapporterte om et ikke planlagt svangerskap?
- G. Hva er IKKE en virkning av gestagenpreparater?
- H. Hvilke kontraindikasjoner er de samme på både kombinasjons p-piller og gestagen p-piller?

During development we found that few of the students had engaged with the quizzes. Out of 16 enrolled students in the course, only 7 of them had attempted to answer the quiz set.

```
for student, student_activities in activities.items():
    quiz_attempts = {}

    for result in student_activities:
        quiz_alias = quiz_map[result['problem_id']][0]

        if quiz_alias not in quiz_attempts.keys():
            quiz_attempts[quiz_alias] = 1
        else:
            quiz_attempts[quiz_alias] += 1

    print("{} - Score: {}".format(student, quiz_attempts))
```

Figure 9.4 Code snippet of how student quiz activity was aggregated.

Using the product of the code in figure 9.4 we manage to transpose the results. These results were then sliced for visualization of each student, and for a joined visualization of all students. The data example shown in figure 9.3 was reduced down to aggregation and identity resolution. 'problem_id' was used as keys for the 'quiz_map', where the values were the question content as found in 'question' entry.

Even with a relatively low volume of students, some of the visualizations were less feasible than others. For example showing the attempt activity of the entire enrolled class of students per quiz question or quiz group would open for issues if the student cohort exceeded 20 students.

As we neared the end of the development phase, we also tested out just transmitting the data necessary for the visualizations, but not actual images.

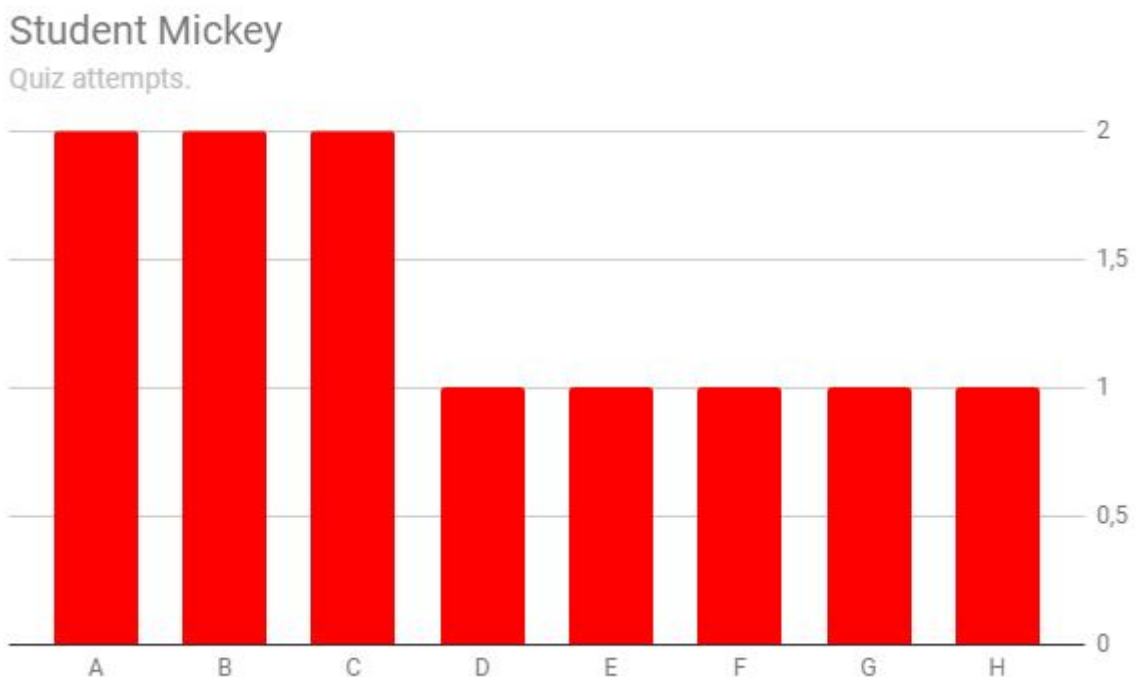


Figure 9.5 The quiz participation of student "Mickey".

Attempts per quiz question

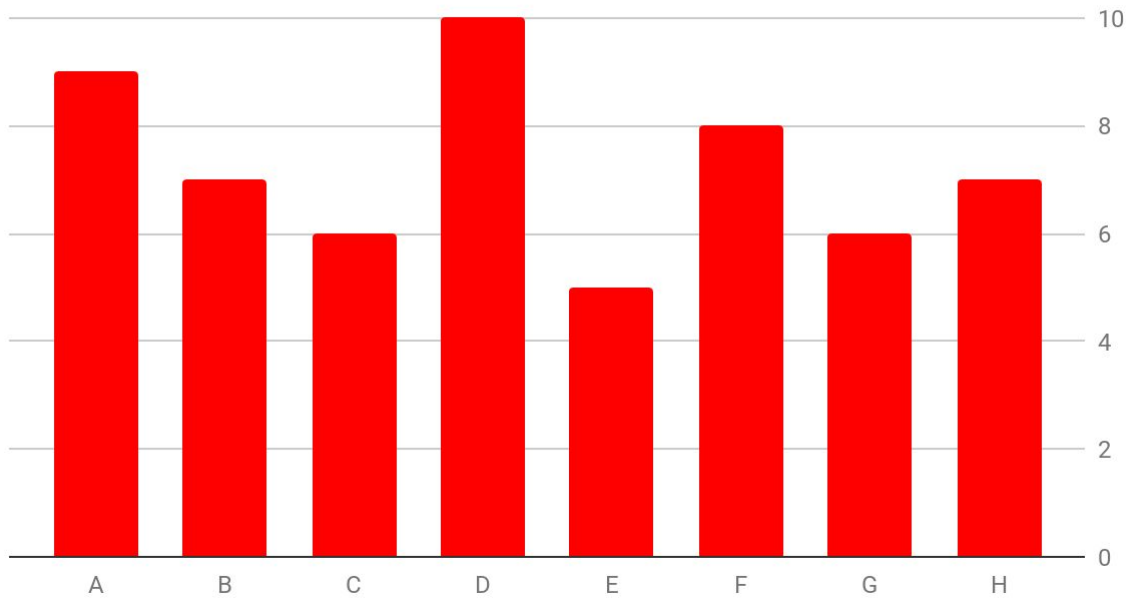


Figure 9.6 - Visualization of attempts made on each quiz question.

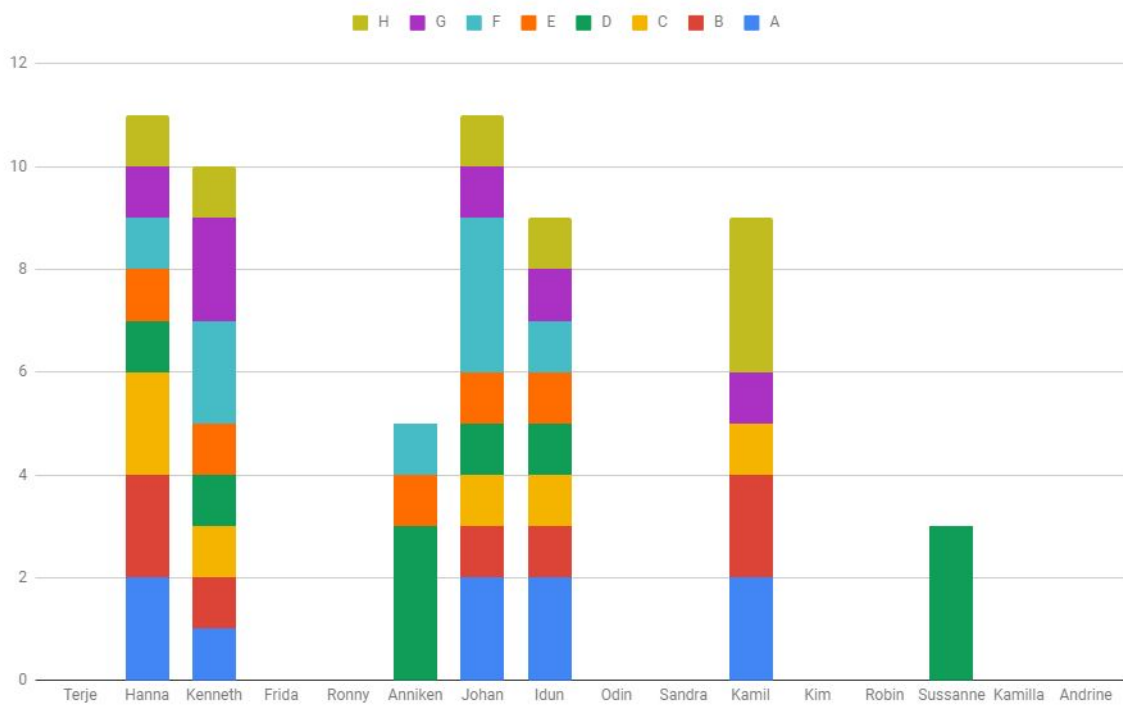


Figure 9.7 Quiz attempts per student in each quiz question. Names are fictional.

Quiz attempt durations suffered from what we considered unreliable accuracy. While we had timestamps for each grading event, there was little we could trace in terms of time spent on the quizzes. At best we could infer some time based on the difference between the timestamps and presume that the student spent at least 10 seconds or so on the first quiz question.

```

for student, student_activities in activities.items():
    event_times = []
    previous = None
    time_deltas = []

    for result in student_activities:
        event_times.append(result[1])
        quiz_alias = quiz_map[result[0]['problem_id']][0]

        if quiz_alias not in quiz_attempts.keys():
            quiz_attempts[quiz_alias] = 1
        else:
            quiz_attempts[quiz_alias] += 1

    for event in event_times:
        if previous is None:
            previous = event
        else:
            duration = event - previous

            if duration.seconds > 600:
                duration = timedelta(seconds=300)

            previous = event
            time_deltas.append(duration)

    if len(time_deltas) <= 0:
        seconds = 0
    else:
        seconds = 10

    for delta in time_deltas:
        seconds += delta.seconds
    print("{} - {}".format(student, seconds))

```

Figure 9.8 The code used to establish quiz durations.

Seconds spent on quizzes

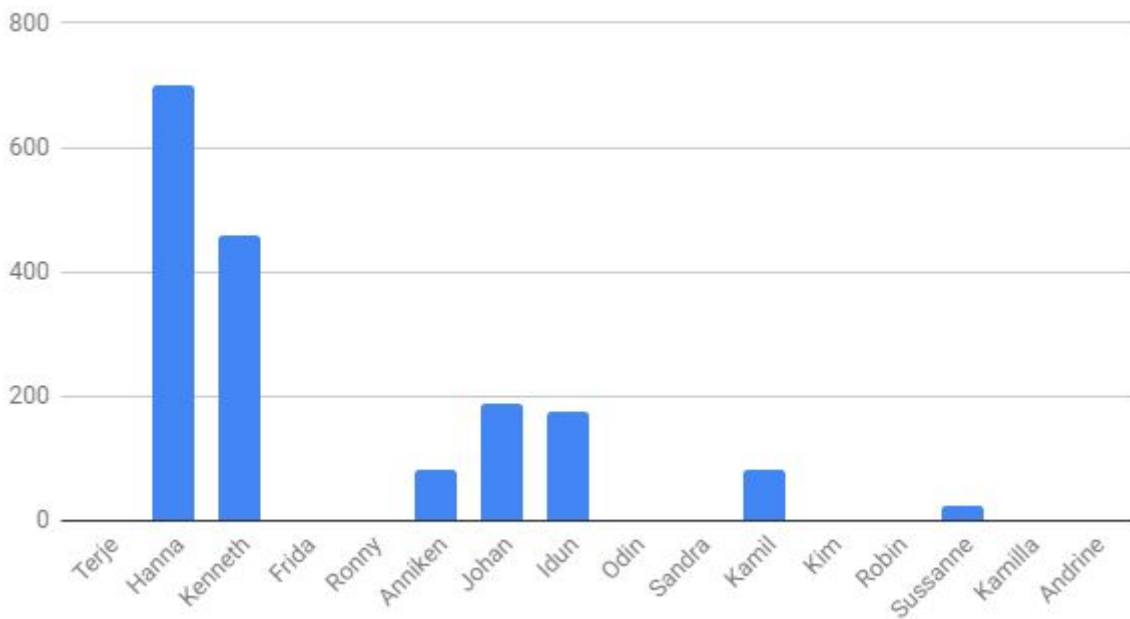


Figure 9.9 Durations each student spent on quizzes in seconds.

Minutes spent on quizzes

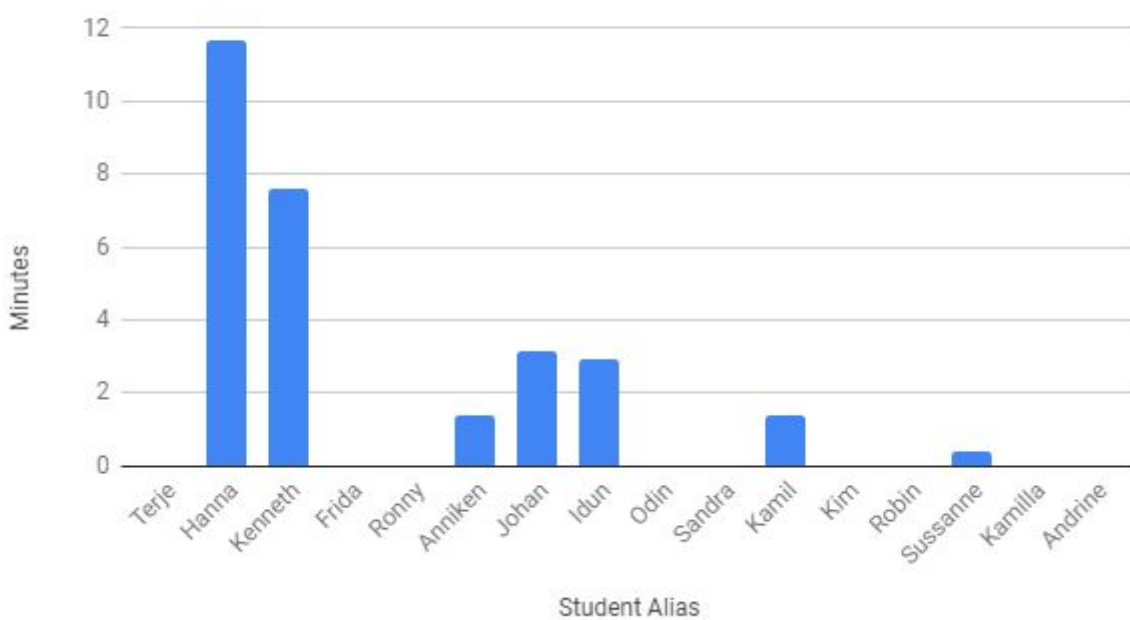


Figure 9.10 Duration each student spent on quizzes in minutes.

We felt that the minute count was more readable than the seconds count, on the assumption that it is rather rare to think of longer durations in terms of seconds.

9.5 Evaluation of the Artifact

In this iteration, the final evaluations would be carried out. In an attempt to gauge usefulness for a teacher to make informed decisions about the structure of the course, we implemented a survey using Likert scale statements. In addition, we also carried out a System Usability Scale survey to evaluate the latest additions to the artifact's visualizations. A system performance test was carried out as well. A final think aloud evaluation was performed as well, in order to capture some thoughts about the artifact.

9.5.1 System Usability Scale survey

The System Usability Scale survey was implemented in much the same way as in the third and fourth iteration. We first displayed the new types of visualizations and evaluated these, and then evaluated the complete set of visualizations afterwards.

System Usability Scale Results

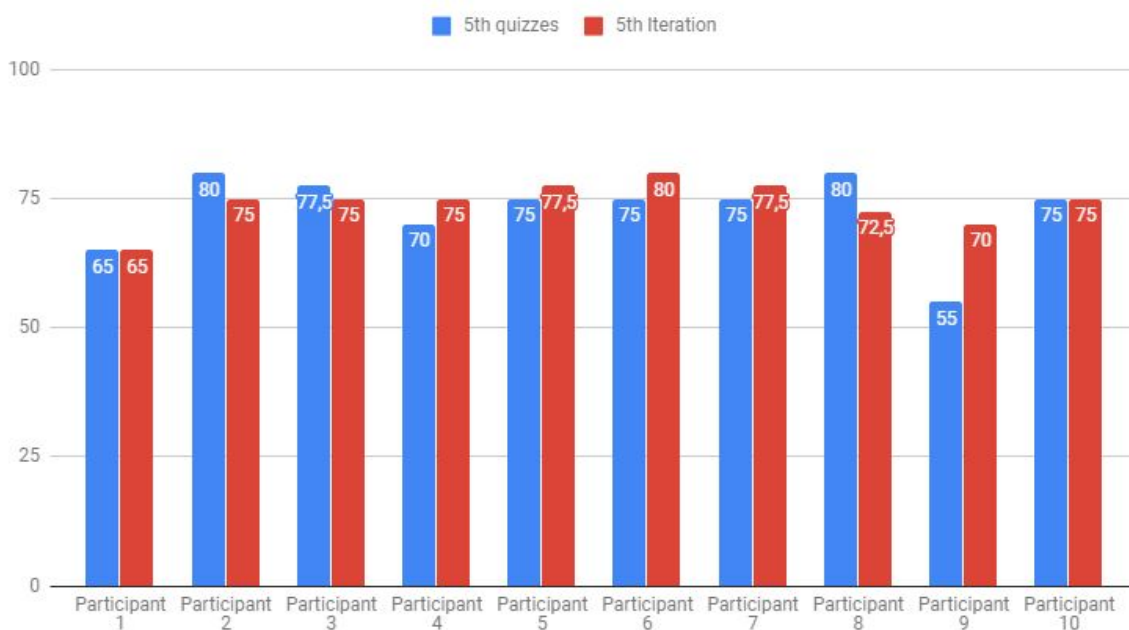


Figure 9.11 System Usability Scale results.

In these iterations, the scores improved by a fair degree above what the previous iterations had been. The quiz visualizations garnered an average score of 72,75, which is well above the average score. The score for the entire artifact in its fifth iteration scored 70,5.

5th Iteration, 4th Iteration and 3rd Iteration

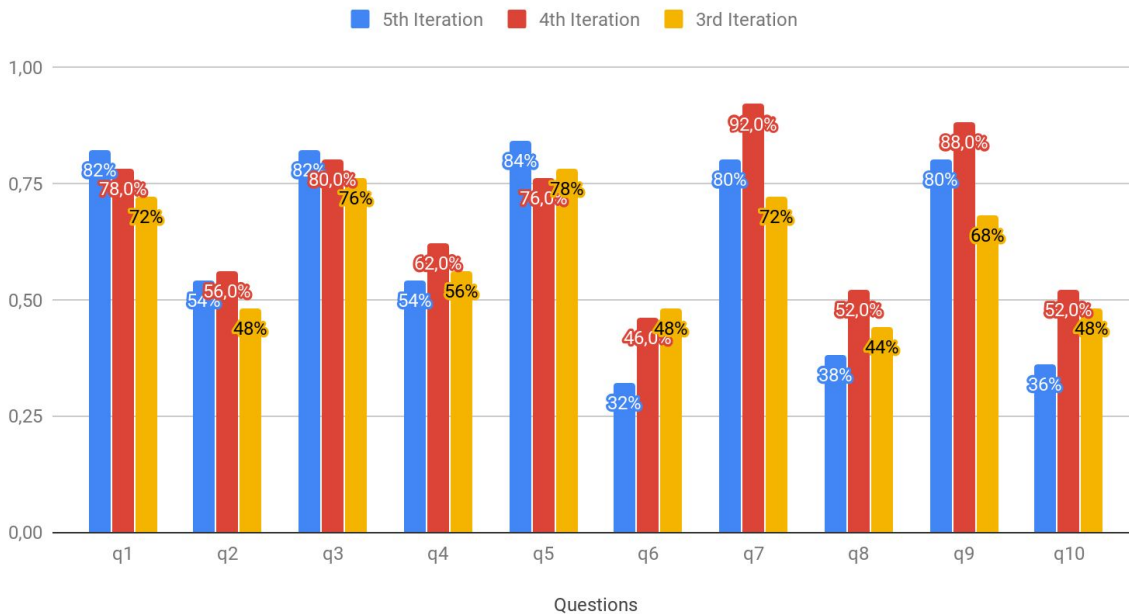


Figure 9.12 Response ratings per question in the SUS questionnaire.

By looking at the rates of the questions in the SUS survey, we could also see what impressions had been made on the participants. The percentages for the positive questions were close to 80% average, while the negative questions were dropping off considerably. The more persistent theme in the development cycle of the artifact was that the participants moderately agreed to that the system was unnecessarily complex and that they might need the support of a technical person to use it.

9.5.2 Likert survey

The Likert survey contained 26 questions, each with 5 possible statements ranging from Strongly Disagree to Strongly Agree. Ten of these questions were classed as “calls for

improvement” while 16 of the questions were classed as “neutral” statements. For the purposes of RQ1, we want to see higher averages in the former class.

An average score in this survey is at 60%. For us to classify a participant as likely to take action to improve the course based on the visualizations, we need their score to satisfy the following conditions:

$$\text{Average Score} < \text{Neutral Score} + 5\% < \text{Improvement Score}$$

Figure 9.13 Criteria for improvement action.

We asked the same participants as those in the System Usability Scale surveys to respond to our surveys. However, only 6 of the participants accepted and filled out these surveys. The results are as follows:

Average responses

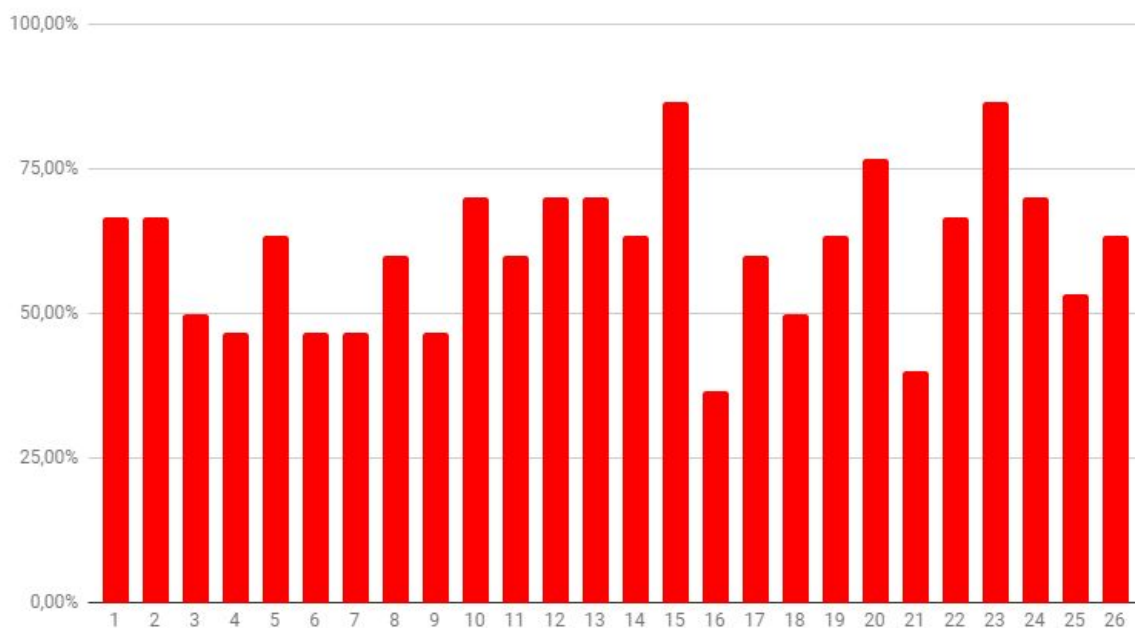


Figure 9.14 Averages per question. Each question is represented by a number.

Average according to class

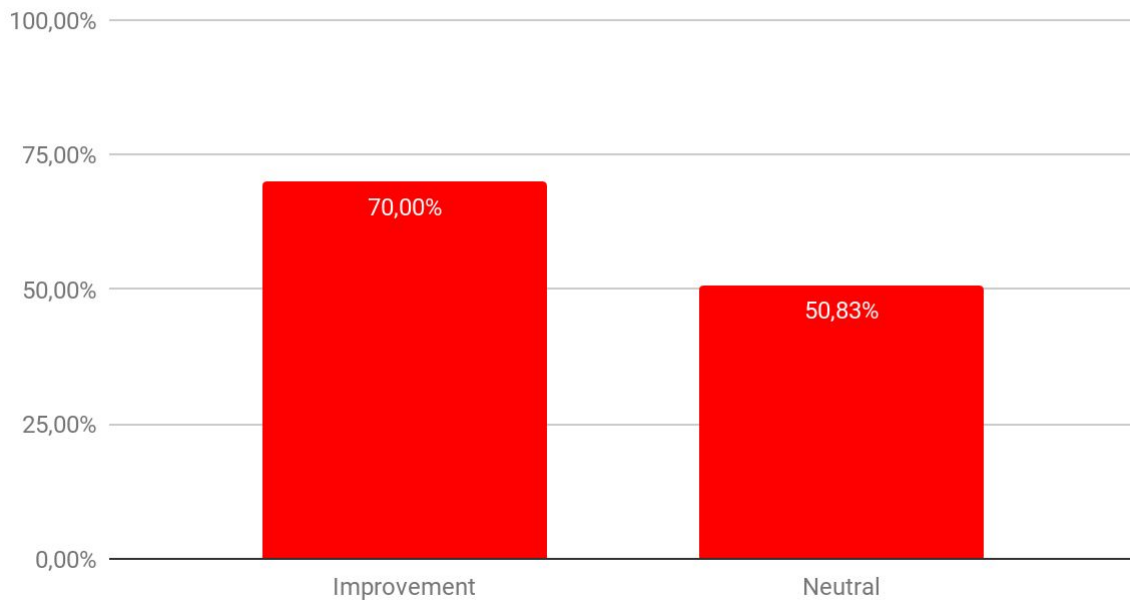


Figure 9.15 Averages of responses grouped by class.

The average score for responses in the “calls for improvement” class were decidedly higher than either neutral responses and the expected average of the questionnaire. While these results were promising, there are some questions which may be obvious to disagree with, such as the “I would do nothing” statement. We purposefully inserted a minority number of “calls for improvement” statements, that is only ten out of twenty six, such that there would be a fair margin of neutral statements that were also reasonable options.

9.5.3 Feedback from Stakeholders

In this iteration, the development process was protracted and little time was left for an evaluation with the Stakeholders. An example document with the new types of visualizations was sent to the stakeholders, where they could evaluate the new additions. The feedback was limited, mostly due to scheduling constraints, and as such contained only a few positive remarks on the visualizations.

9.5.4 Heuristics Extraction

The final iteration had come to an end, but we still codified some contingency heuristics for possible future work.

CH5.1:

Participants scored high on statements and questions which we had classified as “calls for improvement”.

CH5.2:

Participants of the SUS survey reported a slight decrease in question 2, which indicated moderately that the artifact was unnecessarily complex.

CH5.3:

Participants of the SUS survey reported an increase in question 4, which indicated moderately that the participant might need a technical person to use the artifact.

In addition to Contingency Heuristics, we also codified a final Construction Heuristic, which was essentially a UML-like diagram of the artifact’s internal structure.

ConH5.1:

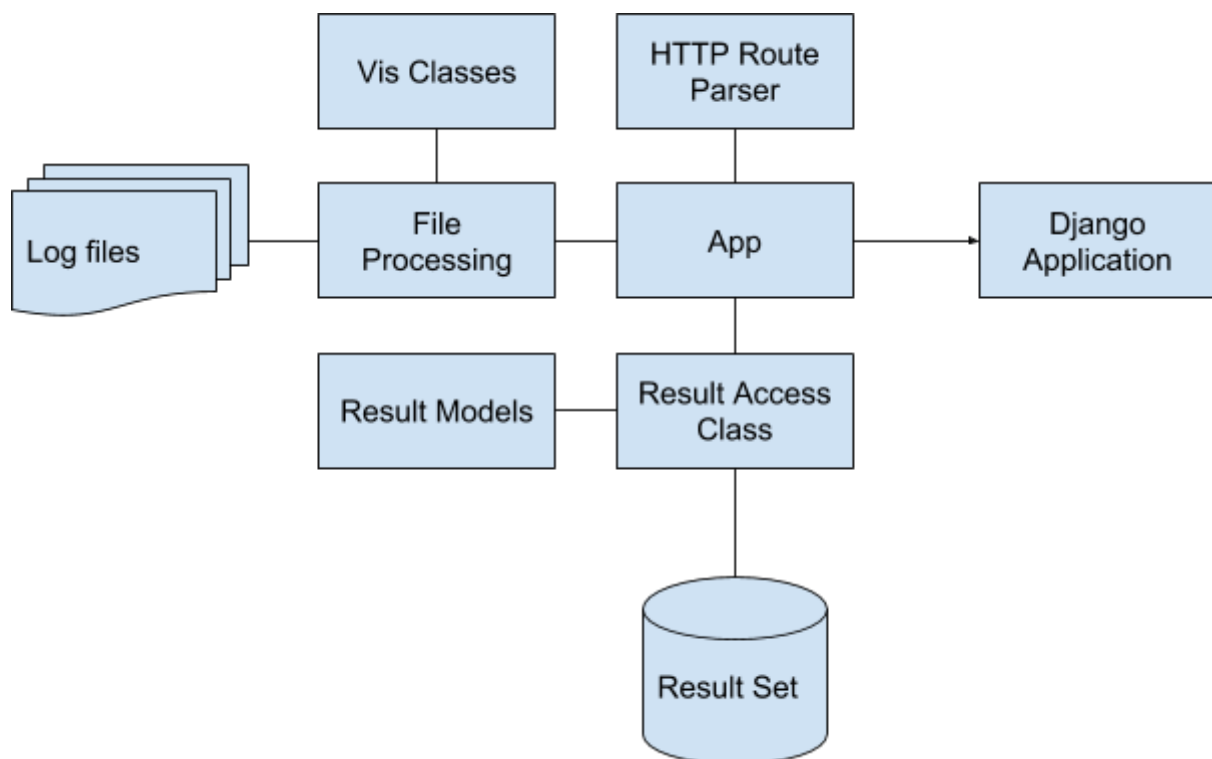


Figure 9.16 Construction Heuristic of the artifact.

10 Discussion

10.1 Findings through application of Design Science

In the course of the research effort, several heuristics were codified based on encounters with the design process and information gathering. While many of these heuristics were not directly relevant for the research questions that guided the research process, they still highlight features that are important to consider in other research projects.

10.1.1 Constraints of Policy

In the second iteration, it was found that policy influenced our capacity to conduct the research we wanted to do. If we had established the necessary infrastructure to host the course ourselves, then we would not have been limited in the same way. While this would have been ideal for our research, it would not necessarily have been viable for the course or the educational institution responsible for the course.

If we had been able to access the infrastructure directly, then the artifact itself would have been significantly different. We could have set up Insights, which would have informed our approach differently than the one we took. The artifact could also have been tested in conjunction with the MOOC software, allowing us to evaluate the artifact as a complement to the Open edX software. Since there were several teachers involved in the course, we could have performed A/B testing where only some of the teachers could view the artifact's visualizations.

10.1.2 Constraints of Infrastructure

The Open edX software set many terms for the data we could access and in many ways also set the frames for what we could analyse. This raised a question of what should set the terms for our methods of observations? Having server logs is certainly valuable as a source of observations in some regards, and we managed to produce several visualizations based on just them. However, is this sufficient?

Based on the development cycle in the research project, we did encounter several points where the data were hard to align with our user stories. For example the video identity table could have been easily circumvented with a proper content manifest in the Open edX software. Essentially, it was possible to develop courseware for which it could be difficult to implement Learning Analytics.

Based on such experiences, an infrastructure was identified as a constraint in itself. For future Learning Analytics research projects, this constraint should be considered and tackled in some way. A possible scenario for future research could be to fork the Open edX software and develop Learning Analytics solutions directly into the software. With such an approach, the researchers are better poised to pursue Learning Analytics research questions more freely.

10.1.3 Constraints of Evaluation

Throughout the research project several evaluations were conducted and the results raised confidence in the usability of the artifact. The evaluations were implemented with a cohort of teacher students and professional teacher whom had varying degrees of experience with teaching. Some also had experience with MOOCs, but none of them had experience with teaching through a MOOC. For this reason we had to consider that the participants responses would likely be equal or close to that of a teacher staff instructing a MOOC for the first time. Another factor was that the participants would not necessarily have the same experiences as a MOOC staff would in terms of developing a MOOC.

With each iteration we also passed the visualizations by the stakeholders of the MOOC course, which made measured responses and criticism. With each iteration, these stakeholders were also involved in the specification for each of the visualizations. In order to receive feedback which was not influenced by the researchers in dialogue or due to the cooperative effort of the research project, the System Usability Scale surveys remained one of the methods where we could receive feedback that were not involuntarily primed by the researchers.

To avoid phenomenon such as the Hawthorne effect, we did not show our research question to the participants, but rather presented the idea of a MOOC and what Learning

Analytics could do. For each new iteration, the participants were shown the most recent visualizations and afterwards asked to fill out a SUS survey.

While this scheme provided useful feedback, it did not directly help in answering the research questions. The System Usability Scale survey had explicit questions about using the visualizations to inform a participant's decisions, but the degree of ambiguity provided no solid basis for confirming whether the participant would make changes or attempt to make improvements.

This is why the fifth iteration featured a separate Likert scale survey in order to better gauge whether the participants would think to make improvements upon the course itself. While the Likert scale survey provided us with some signals as to whether the participants would attempt to improve the learning environment, exactly what changes or how strongly the participants felt about such actions proved hard to discern.

10.1.4 Outcomes of Heuristics

The research produced several heuristics belonging to two classes of heuristics; Contingency Heuristics and Construction Heuristics. A majority of the heuristics produced were Contingency Heuristics, many which had a limited scope and perhaps not much applicability or use outside the iteration process of the research. Throughout the research process, we codified 23 Contingency Heuristics. After having completed all the iterations, the heuristics were counted and classed after the source that reported them. Some of the Contingency Heuristics, especially the ones that originated from the Performance tests, were essentially duplicates or reporting the same phenomenon.

| Source | Count |
|---------------|-----------|
| Environment | 3 |
| Performance | 2 |
| Researcher | 2 |
| Stakeholders | 3 |
| SUS survey | 7 |
| UCE | 5 |
| Likert survey | 1 |
| Total | 23 |

Table 10.1 - Instances of Contingency Heuristics and the source of their reporting.

The majority of the Contingency Heuristics came from evaluation methods, with only 3 of them originating in the environment. Many of the heuristics were helpful for the following iteration, usually in the form of creating new user stories or tasks. We found that the most concrete Contingency Heuristics came from the Use Case Evaluations and the feedback from the stakeholders. At best, the System Usability Scale surveys served as guidelines for design process in the following iteration.

The reporting of the Contingency Heuristics did not have a clearly defined process in the literature [10]. The heuristics codified in the research process were mostly summaries of distinct phenomena in each evaluation process.

The research did not produce a lot of Construction Heuristics. None were explicitly codified in the same way as the Contingency Heuristics, but the research process did create several diagrams that documented the relationship between the various systems involved with the artifact. Its internal structure was also illustrated in the second iteration. The construction heuristics, especially those created by the systemic thinking processes, aided the researchers in understanding the artifact and the problem it was built to solve.

The heuristics produced in the research were considered valuable by the researchers for the process. However, as Dresch *et al* [10] never explicitly defined these heuristics, we could only produce our best approximations of such heuristics.

10.2 Response to the Research Question

In the outset of this research thesis, the question "*How learner data can be visualized to support teaching environment improvement?*" was asked. Through the course of the iterations an artifact has been developed that highlighted how the students engage with various elements in the teaching environment. There are several elements that are subject to improvement, ranging from the organization of syllabus pages, to the content in videos and the way quizzes are formulated. Our artifact shows how the students engage with these elements.

For a course stakeholder, there are visualizations the artifact can show. These visualizations could indicate that the element is not in frequent use or viewed extensively. By making clear and simple visualizations that show *contrast* between high and low use of course elements, we believe the artifact enables the course stakeholder to make *informed decisions* about the course content. The visualizations must also be relevant to the course stakeholders' needs and inform about the course elements which they are capable of changing. Not all elements of a teaching environment are malleable, so visualizations should strive to contextualize the elements that are malleable. We attempted to achieve this by engaging with the stakeholders to establish the visualizations they needed.

Through the design iterations, the artifact has produced visualizations that were increasingly scoring higher on System Usability Scale surveys. The survey's questions were formulated for the participant to consider visualization clarity and whether the participants would use the visualizations to form their opinions. In combination with the positive remarks from the stakeholders in this research, we believe the artifact is capable of producing such contrasts and enable informed decision-making. Especially in the Third Iteration, the Stakeholders were quickly able to spot that students went from the physical classroom to the online learning environment on the same dates, as indicated with Contingency Heuristic 3.4.

In the Fifth Iteration [Section 9.5.2], a subset of the participants from the SUS evaluations answered a Likert questionnaire, where the responses to elements we classed as "calls for improvement" scored above average and above the "neutral" class responses. In addition to indications mentioned above, we have seen that multiple participants have responded in favor of making improvements in the learning environment.

Despite what we consider strong indicators in regards to our research question, the artifact has not helped the stakeholders in identifying course elements in need of improvement. The body of research happened in the interim between two course runs. The fifth iteration concluded weeks before the second course had finished. Even though some of the visualizations reached the stakeholders before the second run of the course, neither we nor the stakeholders have been able to identify elements to improve. We consider this to be a point of ambiguity in our results, since we were not able to fully evaluate the complete artifact at the start of a course run.

Considering that Vulic *et al* [37] were able to identify elements to improve, we believe this is also a possibility with the artifact. By consuming the learner data from the second course, combined with the data set from the first course, the stakeholders may be able to identify improvable elements in future course runs.

11 Conclusions

11.1 Summary

The research question "*How can learner data be visualized to support teaching environment improvement?*" was posed for this thesis. To respond to this research question, we first attempted to apply an existing artifact and then created a new artifact instead. The artifact we designed created visualizations from learner data based on user stories defined in collaboration with the stakeholders of the learning environment we chose to address.

The research process developed over the course of five iterations, four of which involved the design and development of the artifact we created. In each iteration, an artifact and a set of Contingency Heuristics and Construction Heuristics were produced. These heuristics would constantly inform the following iteration with In the first iteration, we attempted to apply the Insights artifact to attempt to answer the research question, but due to policy constraints we were not able to do so.

In the second iteration, we began the design and development of our own artifact, which contained rudimentary visualization capabilities. We evaluated the artifact using Use Case Evaluations, System Usability Scale surveys and System Performance tests.

In the third iteration, we developed user activity visualizations, which showed which dates the learners were active. We evaluated the artifact using System Usability Scale surveys, System Performance tests and a Semi-Structured Interview with the Stakeholders.

In the fourth iteration, we developed video activity visualizations, which showed how much time the learners spent watching videos. We evaluated the artifact in the same way as with the third iteration.

In the fifth iteration, we developed quiz activity visualizations, which showed how the learners dealt with the quizzes on the course. In this final iteration, we evaluated the artifact using System Usability Scale surveys, a Likert survey and a Semi-Structured Interview with the Stakeholders.

With each evaluation we came closer to having a coherent response to the research question. At the end of the fifth iteration we had covered the original user stories which were created in the third iteration, and the visualizations gave a comprehensive overview of the course elements we considered as malleable. The Likert survey results indicated that the participants were likely to engage with statements that we considered “calls for improvement” in the teaching environment. Considering the heuristics and the results of the evaluations, we believe that the artifact is likely producing visualizations that could support improvement in the learning environment for the stakeholder's course. However, the stakeholder has not implemented any improvement based on information from the artifact.

11.2 Future Work

The artifact is by no means finished and there are many possibilities for future work with this artifact. The next steps would be to handle the user stories described in section 9.1, which we did not address before the conclusion of this research. These user stories would likely improve the artifact's visualizations and give the stakeholders an even better awareness of their learners and learning environment.

The artifact would also benefit from having a Learning Dashboard solution, which could be developed to provide the users with a more interactive experience. Currently the visualizations the artifact provides are of most use to the teachers, but that does not preclude that the learners could benefit from them as well.

A large part of the evaluations were quantitative studies, which we believe can form the basis for qualitative evaluations in successive iterations. Many of the visualizations are simple, but can likely be optimized with more in-depth evaluations. The Information Visualization discipline has a wide array of evaluation methods which we did not have the opportunity to use, and many of these could be interesting to apply in future evaluations.

The Learning Analytics life cycle is quite long term as well, with course outcomes produced annually. Using these outcomes as the basis for new visualizations could be beneficial to the optimization segment of the artifact's Learning Analytics cycle. Similarly, we agree with Dyckhoff *et al* [1] in their assessment that more indicators should be added to the learning environment.

12 References

- [1] A. L. Dyckhoff, V. Lukarov, A. Muslim, M. A. Chatti, and U. Schroeder. 2013. Supporting action research with learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*, Dan Suthers, Katrien Verbert, Erik Duval, and Xavier Ochoa (Eds.). ACM, New York, NY, USA, 220-229.
- [2] Abhyankar, Kushal, and Subhashini Ganapathy. "Technology-Enhanced Learning Analytics System Design for Engineering Education." *International Journal of Information and Education Technology* 4.4 (2014): 345.
- [3] Barwick, Daniel W. "Views: Does Class Size Matter?." Retrieved June 28 (2007): 2018.
- [4] Burge, Janet, et al. "SPOCs: what, why, and how." *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. ACM, 2015.
- [5] Card, Mackinlay. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.
- [6] Clow, Doug. "An overview of learning analytics." *Teaching in Higher Education* 18.6 (2013): 683-695.
- [7] Clow, Doug. "The learning analytics cycle: closing the loop effectively." *Proceedings of the 2nd international conference on learning analytics and knowledge*. ACM, 2012.
- [8] Cooper, Adam. *A Brief History of Analytics A Briefing Paper*. CETIS Analytics Series. JISC CETIS, November 2012.
- [9] Dawson, Shane, and Harry Hubball. "Curriculum Analytics: Application of Social Network Analysis for Improving Strategic Curriculum Decision-Making in a Research-Intensive University." *Teaching & Learning Inquiry: The ISSOTL Journal*, vol. 2, no. 2, 2014, pp. 59–74. JSTOR
- [10] Dresch, Aline, Daniel Pacheco Lacerda, and José Antônio Valle Antunes Jr. *Design science research: A method for science and technology advancement*. Springer, 2014.
- [11] Ferguson, Rebecca. "Learning analytics: drivers, developments and challenges." *International Journal of Technology Enhanced Learning* 4.5-6 (2012): 304-317
- [12] Fini, Antonio. "The technological dimension of a massive open online course: The case of the CCK08 course tools." *The International Review of Research in Open and Distributed Learning* 10.5 (2009).
- [13] Goral, Tim. "Make way for SPOCS: small, private online courses may provide what MOOCs can't." *University business* 16.7 (2013).
- [14] Govaerts, S., Verbert, K., Klerkx, J., Duval, E. (2010). *Visualizing Activities for Self-reflection and Awareness*. *Proceedings of the 9th international conference on Web- based Learning*, pages 91–100. Springer, 2010
- [15] Héctor J. Pijera Díaz, Javier Santofimia Ruiz, José A. Ruipérez-Valiente, Pedro J. Muñoz-Merino, and Carlos Delgado Kloos. 2016. A Demonstration of ANALYSE: A Learning Analytics Tool for Open edX. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale (L@S '16)*. ACM, New York, NY, USA, 329-330.
- [16] Helen Sharp, Yvonne Rogers, and Jenny Preece. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, 3rd edition edition, 2011.
- [17] Hornbæk, Kasper, et al. "Use case evaluation (UCE): A method for early usability evaluation in software development." *IFIP Conference on Human-Computer Interaction*. Springer, Berlin, Heidelberg, 2007.
- [18] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, J. A. Gascón-Pinedo and C. D. Kloos, "Scaling to Massiveness With ANALYSE: A Learning Analytics Tool for Open edX," in *IEEE Transactions on*

- Human-Machine Systems, vol. 47, no. 6, pp. 909-914, Dec. 2017.
- [19] Javier Santofimia Ruiz, Héctor J. Piñeira Díaz, José A. Ruipérez-Valiente, Pedro J. Muñoz-Merino, and Carlos Delgado Kloos. 2014. Towards the development of a learning analytics extension in open edX. In Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM '14). ACM, New York, NY, USA, 299-306.
- [20] Johnstone, Sally M. "Open educational resources serve the world." *Educause Quarterly* 28.3 (2005): 15.
- [21] Khalil, Mohammad, and Martin Ebner. "Learning analytics: principles and constraints." *EdMedia: World Conference on Educational Media and Technology*. Association for the Advancement of Computing in Education (AACE), 2015.
- [22] Long, Phillip. *Proceedings Of The 1St International Conference On Learning Analytics And Knowledge*. 1st ed. New York, NY: ACM, 2011. Print.
- [23] Marc Wells, Alex Wollenschlaeger, David Lefevre, George D. Magoulas, and Alexandra Poulouvassilis. 2016. Analysing engagement in an online management programme and implications for course design. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16). ACM, New York, NY, USA, 236-240.
- [24] Mazza, Riccardo & Dimitrova, Vania. (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. 154-161. 10.1145/1013367.1013393.
- [25] McGee, Patricia, and Abby Reis. "Blended course design: A synthesis of best practices." *Journal of Asynchronous Learning Networks* 16.4 (2012): 7-22.
- [26] Parr, Chris. "Mooc creators criticise courses' lack of creativity." *Times Higher Education* 17.03 (2013): 2013.
- [27] Pistilli, Matthew D., James E. Willis, and John P. Campbell. "Analytics through an institutional lens: Definition, theory, design, and impact." *Learning analytics*. Springer, New York, NY, 2014. 79-102.
- [28] Schmitz, H., Schefel, M., Friedrich, M., Jahn, M., Niemann, K., & Wolpers, M., (2009). CAMera for PLE. Learning in the Synergy of Multiple Disciplines. Proceedings of the 4th European Conference on Technology Enhanced Learning: Learning in the Synergy of Multiple Disciplines, 507-520. Berlin, Heidelberg: Springer-Verlag . doi:10.1007/978-3-642-04636-0_47
- [29] Schwendimann, Beat A., et al. "Perceiving learning at a glance: A systematic literature review of learning dashboard research." *IEEE Transactions on Learning Technologies* 10.1 (2017): 30-41.
- [30] Shedroff, Nathan. "Information interaction design: A unified field theory of design." *Information design* (1999): 267-292.
- [31] Shum, Simon Buckingham, and Rebecca Ferguson. "Social learning analytics." *Journal of educational technology & society* 15.3 (2012).
- [32] Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). USA: MIT Press.
- [33] Smith, Marshall S., and Catherine M. Casserly. "The promise of open educational resources." *Change: The Magazine of higher learning* 38.5 (2006): 8-17.
- [34] Sorman, John D. *Business dynamics: systems thinking and modeling for a complex world*. No. HD30. 2 S7835 2000. 2000.
- [35] Sumner, Jennifer. "Serving the system: A critical history of distance education." *Open learning* 15.3 (2000): 267-285.
- [36] Vigentini L., Clayphan A., Zhang X., Chitsaz M. (2017) Overcoming the MOOC Data Deluge with Learning Analytic Dashboards. In: Peña-Ayala A. (eds) *Learning Analytics: Fundamentals, Applications, and Trends*. Studies in Systems, Decision and Control, vol 94. Springer, Cham

- [37] Vulic, John, et al. "Data Analytics Informing MOOC Continuous Improvement." (2017).
- [38] Wiley, David A. "Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy." *The instructional use of learning objects* 2830.435 (2000): 1-35.
- [39] Zhang, H., Almeroth, K., Knight, A., Bulger, M. & Mayer, R. (2007). Moodog: Tracking students' Online Learning Activities. In C. Montgomerie & J. Seale (Eds.), *Proceedings of ED-MEDIA 2007--World Conference on Educational Multimedia, Hypermedia & Telecommunications* (pp. 4415-4422). Vancouver, Canada: Association for the Advancement of Computing in Education (AACE). Retrieved November 27, 2018 from <https://www.learntechlib.org/primary/p/26016/>.

Appendix A - Likert Scale Survey

| Question | Class | Statement | Variable Statement |
|----------|-------------|---|---|
| | | If a video has few views, I would think: | |
| 1 | Improvement | | The video might be better placed in the course ware. |
| 2 | Neutral | | The students fail to recognize the importance of the video. |
| 3 | Neutral | | The video's content is not engaging. |
| 4 | Improvement | | The video's content could be improved. |
| 5 | Neutral | | The video is sufficient - some other factor is causing the low views. |
| 6 | Neutral | | I would not take action. |
| | | | |
| | | If a video has many views, I would think: | |
| 7 | Neutral | | The videos position in the course ware is appropriate. |
| 8 | Neutral | | The students recognized that the video was important. |
| 9 | Neutral | | The video is engaging. |
| 10 | Improvement | | The video's content could be improved. |
| 11 | Neutral | | The video's high views is somehow a fluke. |
| 12 | Improvement | | I would consider this video when designing other videos. |
| | | | |
| | | | |
| | | | |
| | | If student engagement drops off, I would consider the reason to be: | |
| 13 | Neutral | | Natural decline in interest. |
| 14 | Improvement | | Course curriculum being difficult. |
| 15 | Improvement | | Few course ware items to engage with. |
| 16 | Neutral | | Schedule conflicts from other |

| | | | |
|----|-------------|--|--|
| | | | personal activities. |
| | | | |
| | | If student engagement is focused on certain days of the week, I would: | |
| 17 | Neutral | | Compare it to the schedule of the course. |
| 18 | Neutral | | Assume it is nothing special. |
| 19 | Improvement | | Consider it an opportunity to engage with students online in those days. |
| 20 | Improvement | | Schedule staff to be active on the same days. |
| | | | |
| | | | |
| | | | |
| | | If more than an average amount of students answers wrong on a quiz, I would: | |
| 21 | Neutral | | Want to know if the students had read the material properly. |
| 22 | Neutral | | Assume the students clicked randomly or not take the quiz seriously. |
| 23 | Improvement | | Consider the phrasing of the quiz. |
| 24 | Improvement | | Consider the placement of the quiz. |
| | | | |
| 25 | Neutral | I would want to know what students read before they answer a quiz. | |
| | | | |
| 26 | Neutral | I would want to know what students read after they fail a quiz. | |

Appendix B - Semi-Structured Interview Guide

1. Any criticisms from our former meeting?
2. Thoughts on the new visualizations?
3. Thoughts on the direction of the development process?
4. Any ideas or suggestions for the following iteration?