The Evaluation Framework for Learning Analytics

Maren Scheffel

The Evaluation Framework for Learning Analytics

The research reported in this thesis was carried out at the Open Universiteit in the Netherlands at Welten Institute – Research Centre for Learning, Teaching and Technology,

Welten Institute
Research Centre for Learning, Teaching and Technology

Open Universiteit
welten-institute.org

and under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

SIKS Dissertation Series No. 2017-34

Printed by Datawyse
Cover design: Chris Peeters
Typeset in LaTeX

ISBN/EAN: 978 94 92231 81 9

The Evaluation Framework for Learning Analytics

Proefschrift

ter verkrijging van de graad van doctor
aan de Open Universiteit
op gezag van de rector magnificus
prof. mr. A. Oskamp
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 22 september 2017 te Heerlen
om 13:30 uur precies

door

Maren Scheffel
geboren op 5 juni 1983 te Bad Pyrmont

Promotor
Prof. dr. M.M. Specht
Open Universiteit

Co-Promotor
Prof. dr. H. Drachsler
Open Universiteit

Overige leden van de beoordelingscommissie
Prof. dr. C. Delgado-Kloos
Universidad Carlos III de Madrid

Prof. dr. D. Griffiths
University of Bolton

Prof. dr. F.L.J.M. Brand-Gruwel
Open Universiteit

Prof. dr. M. Kalz
Open Universiteit

Dr. R. Ferguson
The Open University

Finally, if there isn't a relevant questionnaire for your situation, you can devise one.

John Brooke (2013)

Contents

General Introduction

> *Insight is the buzzword in the data revolution. People who have insight into their behaviour will be able to work more efficiently from now on. This equally applies to organisations, higher education institutions and nations. The digital revolution has brought about a tremendous increase in the volume of data and it is comparatively easy to access. It would seem to be a lost opportunity to ignore such data.*
>
> Marjolein van Trigt, SURFnet (2016)

The amount of data being produced by digital devices increases exponentially every year. It is expected that by the year 2020 there will be three times as many devices as there are people on earth, that global IP traffic will reach several zettabytes, i.e. one trillion gigabytes, per year, and that mobile and wireless devices will account for two thirds of the global IP traffic (Cisco, 2016). The United Nations are calling for actions by mobilising the data revolution for sustainable development as according to them "data are the lifeblood of decision-making and the raw material for accountability" and "governments, companies, researchers and citizen groups are in a ferment of experimentation, innovation and adaptation to the new world of data, a world in which data are bigger, faster and more detailed than ever before" (UN Independent Expert Advisory Group, 2014, p. 2).

Analysing collected data to detect patterns and improve processes has been a staple in business and commerce for decades, e.g. for reporting purposes, customer handling, production controlling, success measurement, etc. However, doing so in the educational domain is in comparison a rather new field of research. Although students' enrolments, course and grade records have been collected and also statistically analysed in one way or another in the pre-digital era, it was only the emergence of the internet in combination with the development of online learning management systems and virtual learning environments that facilitated the forming of a research field called *learning analytics*.

But what exactly is learning analytics? The Oxford Dictionaries (2017) define the noun *learning* as the "acquisition of knowledge or skills through study, experience, or being taught" and the noun *analytics* as the "systematic computational analysis of data or statistics" as well as "information resulting from the systematic analysis of data or statistics". According to these definitions, the term *learning analytics* can thus be used to describe both the analysis of knowledge acquisition processes and its results. A more precise definition that has commonly been accepted and used by those involved in the field itself is the one provided by the organisers of the first

International Conference on Learning Analytics and Knowledge (LAK) in 2011 where learning analytics is “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” (Siemens, 2011).

Learning analytics can thus be seen as a four-step cycle (Clow, 2012): learners, data about those learners, metrics applied to the data, and interventions based on the metrics’ results. As it is further explained, learners can be pupils or students in schools or at university; they can take part in formal or informal learning. The data collected about the learners can come from various sources, some being collected automatically and some being generated by the learners themselves, e.g. demographics, grades, log data from interactions with a learning environment, or forum posts. The metrics, or analytics, can for example range from statistical analyses, over detection of learners at risk (of failing, falling behind, dropping out), to the prediction of grades, the analysis of social networks and the recommendation of further actions, learning material, etc. Finally, as Clow (2012) stresses, “the cycle is not complete until these metrics are used to drive one or more interventions that have some effect on learners” (p. 135), e.g. the provision of dashboards to support learners and teachers, teachers getting in touch with learners or adapting their teaching or institutions improving their course offerings or structures.

Learning analytics is a multi-disciplinary research field that builds on ideas from and connects to other fields such as learning sciences, computer supported collaborative learning, technology enhanced learning, cyber-learning, learning at scale, and user modelling as Gašević et al. (2015a) have pointed out. It also incorporates ideas and techniques from fields such as process mining, data processing, information retrieval, computer science, information visualisation and psychology. A strongly related field is that of educational data mining. Despite their many commonalities and sometimes even synonymous usage of the terms, the main difference between the two fields is usually attributed to the following: learning analytics focuses more on human interpretation of data and the use of visualisations, while educational data mining focuses more on automated methods (Siemens and Baker, 2012; Baker and Inventado, 2014).

Although research in learning analytics as well as the development and implementation of learning analytics applications in educational institutions has been steadily on the rise during the last years, there is still a gap between the potential of learning analytics identified by research and how much of this potential has been achieved so far (Ferguson et al., 2016a). This is due to several reasons, but one of the main ones is the young age of the field. Becoming a distinct research field only in 2011, the field is simply too young to have produced long-term studies and has had to make do with short-term pilot studies instead. As Ferguson et al. (2016a) point out in their report for the European Commission “for other educational institutions to follow the lead of these early adopters, and to encourage them to build on what they have already achieved, more work is needed on areas related to adoption and implementation” (p. 9). However, they also indicate that a lot of work is being done in supplying tools, data, models and prototypes but little work is done to gather the demands of the involved stakeholders. Additionally, most tools seem to not be 'actionable' enough and do not focus enough on innovative pedagogical processes and practices. A related aspect has been pointed out by the ECAR-ANALYTICS Working Group (2015): "Analytics without action is merely reporting; interventions based on analytics are needed to improve student outcomes" (p. 3).

Over the years, though, the type and scope of learning analytics interventions have widened. While in the early days of learning analytics data was most of the time being analysed on an institutional level or departmental level, learning analytics results are now also being provided to teachers and to students, i.e. the purpose of learning data analysis now ranges from 'plain number crunching' and statistics over detecting students at risk to recommendations of materials or activities as well as to supporting learners and teachers in their learning/teaching processes. Similarly, while in the beginning single pilot studies were being conducted by individual teachers or in connection with researchers, learning analytics applications and initiatives can now also be found on other levels, i.e. from the micro level (individual users and courses), to the meso level (departments, whole institutions) and to the macro level (state and country).

According to Desouza and Smith (2016) "interest in using data more creatively (some might say, more innovatively) as a way to become more precise in how interventions are devised to improve outputs and outcomes is at an all-time high" (p. 12). A very prominent and important issue, however, that cannot be ignored when pursuing that interest is that of ethics and privacy. And thus, this issue is being addressed more and more by the learning analytics community. From codes of practice (Sclater, 2016) to checklists (Drachsler and Greller, 2016) and discussions about the obligation to act (Prinsloo and Slade, 2017), ethical and legal frameworks as well as guidelines and policies are being discussed and set up. Hoel et al. (2017) examine and compare several legal frameworks, stressing the importance and influence the new European General Data Protection Regulation[1] that has been approved in 2016 and will come into effect in 2018 will have on learning analytics research and development. They also point out that "a focus on privacy and data protection also creates the opportunity to achieve the necessary leverage in determining what questions LA should answer" (p. 250). In their discussion whether privacy could be a show-stopper for learning analytics, Drachsler et al. (2016) claim that this will not be the case but that "there is huge economic, social and political momentum behind the big data business model, and this momentum is reflected within the domain of education, for good or for ill" (p.23).

Apart from the fact whether data may be collected and if so, what data is collected and how, the European Commission's ET2020 Working Group on Digital Skills and Competences (2016) has pointed out that attention needs to be paid as to why data is being collected, i.e. "the value or potential value behind data collection" (p. 2). First, it should not be forgotten that learning analytics is about learning

[1] http://www.eugdpr.org

and that learning analytics tools should therefore be grounded in theoretically-established instructional strategies (Gašević et al., 2015b) in order to support the directly involved stakeholders, i.e. the students whose data is being used and the teachers who interact with the students and support them in their learning processes. Second, empirical evidence as to whether learning analytics tools have the desired effect or not is very sparse. Ferguson et al. (2016a) stress, that the "issue with current tools is finding evidence relating to formal validation of tools (e.g. whether the tools fulfil their intended purpose such as having an impact on learning; or making learning more efficient or more effective)" (p. 25) and that "at this stage, there is no overwhelming evidence that learning analytics have fostered more effective and efficient learning processes and organisations" (p. 25). They also specifically point out that the high expectations for learning analytics to improve and innovate learning as well as teaching have not been realised yet. The ECAR-ANALYTICS Working Group (2015) similarly emphasise that "the specifics of the innovations that analytics will drive (or the failures that will occur) are largely speculative" (p. 1) and that further research is needed to inform educational practice.

While there are applications and frameworks available that allow educational institutions to measure or categorise their readiness for learning analytics (e.g. the Learning Analytics Readiness Instrument by Arnold et al. (2014a)) or the maturity of the already implemented learning analytics (e.g. the analytics maturity index by Bichsel (2012)), there are currently no clear indicators for the evaluation of learning analytics tools that actually support the comparison of existing tools. Learners, teachers, courses, programmes, departments or institutions that use or implement learning analytics tools currently have no standardised way of telling whether these tools do what they are meant to do. To be more precise, it is very often not known whether learners and teachers – the stakeholders who are directly involved in the learning process and directly impacted by the use of learning analytics tools – benefit from the implemented learning analytics with regard to any form of indicator. There is thus an urgent need for the quality assessment of learning analytics tools, i.e. for the development of "evaluation checklists for learning analytics tools" (Ferguson et al., 2016a, p. 40).

This thesis addresses the lack of evaluation instruments by creating and validating an evaluation framework for learning analytics that will help standardise the evaluation of learning analytics tools and allow for measuring and comparing the impact of learning analytics on educational practices. Inspired by the System Usability Scale (SUS), a "reliable, low-cost usability scale that can be used for global assessments of system usability" (Brooke, 1996, p. 1), the evaluation framework for learning analytics aims to provide similar facilities for the learning analytics domain. Using the subjective assessments by their users is a quick and simple way to get a general indication of the overall quality of a tool in comparison to other tools or other versions of the same tool as Brooke (1996) points out. The main objectives of the research presented in this thesis therefore are to identify quality indicators for learning analytics, to create an applicable evaluation instrument based on these indicators and to validate the evaluation instrument.

Outline of the thesis

The thesis is structured into three parts that describe the iterative process of creating, applying, evaluating and improving the different versions of the evaluation framework for learning analytics (EFLA). The first part describes the identification of quality indicators for learning analytics as well as the initialisation, first evaluation and first improvement of the EFLA based on input from the learning analytics community as well as related literature; the second part then applies the EFLA to a collaborative learning support widget and describes the subsequent evaluation and improvement; the third part then illustrates the application of the EFLA to widgets of a massive open online course platform and explains the final evaluation and validation process of the framework.

Chapter 1 presents a group concept mapping (GCM) study conducted with experts from the field of learning analytics. After first collecting a list of 103 quality indicators from the learning analytics community, the invited experts sorted and rated the indicators according to their importance and feasibility. Based on their aggregated input, shared patterns are revealed in the collected data using multidimensional scaling and hierarchical clustering. The resulting visualisations are used to interpret the data. The results of the group concept mapping study are then used to construct the dimensions and items of the first version of the evaluation framework for learning analytics (EFLA-1), i.e. an evaluation instrument that aims to standardise the evaluation of learning analytics tools and to provide a mean to capture evidence for the impact of learning analytics on educational practices in a standardised manner. The outcomes of the group concept mapping study are then further extended and contextualised with findings from a focused literature review.

In **Chapter 2** the first version of the evaluation framework is turned into an applicable tool, i.e. a questionnaire. A group of learning analytics experts uses the process of evaluating a collection of learning analytics tools to evaluate the applicability of EFLA-1. Using the quantitative and qualitative results of this evaluation study, useful insights are gained about the characteristics of the evaluation framework that are carried over into the creation of the next framework version. In order to address the requirements established in the evaluation study and to thus improve the framework, the data from the group concept mapping study is reconsidered and complemented by a look at related literature, i.e. other evaluation instruments, frameworks and categorisations, in order to decide on the framework's dimensions and to narrow down the choice of items. For every dimension, further literature is then consulted to motivate and theoretically ground the chosen items. The chapter concludes with the presentation of the second version of the evaluation framework for learning analytics (EFLA-2) that provides a learner and a teacher section, both consisting of four dimensions with three items each.

In order to explore the usage of EFLA-2 when it comes to evaluating learning analytics applications, a learning analytics widget to support collaborative learning is designed. In collaborative learning environments, students work together on assignments in virtual teams and depend on each other's contribution to achieve their learning objectives. Group awareness widgets that visualise information about the different group members based on information collected from the individuals can foster awareness and reflection processes within the group. **Chapter 3** presents a formative data study about the predictive power of several indicators of an awareness widget based on automatically logged user data from an online learning environment. Before evaluating the widget with the EFLA-2, however, this chapter investigates the predictive power of several indicators of the activity widget towards the students' grades at various points in time during the course by instantiating these indicators with data from previous runs of the collaborative online learning course. That is, this chapter analyses the log data from the previous years of the course to explore what the widget indicator scores would have been if the widget had been used in those years. The results of the analysis show that the grades and widget indicator scores are significantly and positively correlated and that some indicators can indeed be used as predictors for the students' grades.

The collaborative learning processes of students in online learning environments can be supported by providing learning analytics-based visualisations that foster awareness and reflection about an individual's as well as the team's behaviour and their learning and collaboration processes. For the empirical study presented in **Chapter 4** an activity widget is implemented into the online learning environment of a live five-months Master course. The predictive power of the widget indicators towards the students' grades is then investigated and compared to the results of the exploratory study presented in Chapter 3. The results of this comparison show that there are indeed predictive relations between the students' actions and their grades and they indicate that some of the observed differences can be attributed to the availability of the widget. Additionally, the implemented learning analytics widget is evaluated using EFLA-2. Students and tutors of the course are asked to fill out the EFLA-2 questionnaire once in the middle of the course and once at the very end. The evaluation results show that the evaluation framework for learning analytics can be used to evaluate a learning analytics application at several points in time and to reflect differences between the two stakeholder groups.

Chapter 5 presents the next iteration of the evaluation and improvement process of the evaluation framework for learning analytics. The results of the widget evaluation using EFLA-2 in Chapter 4 are employed to now evaluate the framework itself. That is, the students' and tutors' answers to the EFLA-2 questionnaire are statistically analysed by conducting principal component analysis as well as reliability analysis in order to identify problematic issues with any of the EFLA-2's items. In addition to this quantitative analysis, qualitative feedback is gathered during a face-to-face full-day experts focus group. All four EFLA-2 dimensions and their items are discussed in detail in order to determine what is needed to improve the framework. Based on the quantitative and qualitative evaluation results, the third version of the evaluation framework for learning analytics (EFLA-3) is constructed. The framework still offers a learner and a teacher section and still consists of four dimensions. Several items, however, are refined and adapted, while others are deleted from the framework, resulting in a total of ten items for EFLA-3.

As the need for more empirical evidences about the effects for learning analytics on the directly involved and impacted stakeholders, i.e. learners and teachers, is increasing, **Chapter 6** reports about the development and implementation of several learning analytics widgets into a massive open online course platform's dashboard. Meant to replace previously used widgets of the platform's dashboard, the new widgets are designed based on input from the literature about learning analytics dashboards for MOOCs. The widgets are implemented according to the xAPI standard for the collection and analysis of activity log data. In a lab study, the old as well as the new versions of the widgets are evaluated by students and teachers using EFLA-3. The evaluation results show that the evaluation framework for learning analytics can be used to measure changes between different versions of widgets as well as differences between the two stakeholder groups. Additionally, this chapter presents the final iteration of the evaluation and improvement process of the evaluation framework for learning analytics. Using the answers to the EFLA-3 questionnaire by the lab study's participants, the framework is evaluated using principal component and reliability analysis in order to determine whether the four current EFLA-3 dimensions validly represent the underlying components and whether the items within the dimensions reliably measure the underlying component. After a first round of analysis, two items are eliminated from the evaluation framework. Also, it is detected that the framework's structure is very likely to be three-dimensional instead of four-dimensional. The second round of analysis confirms this assumption. Based on the analysis results the valid and reliable fourth and final version of the evaluation framework for learning analytics (EFLA-4) is constructed. The framework has a learner and a teacher section and consists of three dimensions with a total of eight items.

The thesis is concluded by a **General Discussion** of the results reported in all studies. Apart from a summary of the findings, general limitations are reviewed and practical implications are discussed.

Part I

That's The Way – Input from the field

Chapter 1

In The Neighbourhood – Or: Quality indicators for learning analytics

The first part of the thesis dives into the research field of learning analytics and involves members of its community in a group concept mapping study (GCM). The chapter proposes the first version of an evaluation framework for learning analytics (EFLA) that aims to standardise the evaluation of learning analytics tools and to provide a mean to capture evidence for the impact of learning analytics on educational practices in a standardised manner. The dimensions of the framework and its items are based on the results of the group concept mapping study, i.e. on input provided by experts in the field of learning analytics. The outcomes of the study are then further extended and contextualised with findings from a focused literature review.

This chapter is based on:

Scheffel, M., Drachsler, H., Stoyanov, S., and Specht, M. (2014). Quality Indicators for Learning Analytics. *Educational Technology & Society, 17*(4):117–132.

1.1 Introduction

In the last few years, the research field of learning analytics (LA) has been growing steadily. According to Siemens (2011) learning analytics is “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”. Building on ideas from process mining, data processing, information retrieval, technology-enhanced learning, educational data mining, and visualisation learning analytics is a multi-disciplinary research field that now forms its own domain. Several resources and organisations are already dealing with the topic in a journal[1], special journal issues (e.g. Rivera-Pelayo et al. (2013)), conferences[2], workshops as well as courses, summer institutes[3], and a society[4] specifically dedicated to learning analytics. There, the research community has worked on the state of the art in learning analytics, its processes, frameworks, definitions, and challenges (see Clow (2012); Drachsler and Greller (2012); Duval (2011); Elias (2011); Ferguson (2012a,b); Greller and Drachsler (2012); Siemens and Baker (2012)).

Making use of learning analytics can give added value to learners as well as educators. Many university courses today consist of a blended approach between classroom lectures and self-regulated learning activities. Learning analytics can help learners to better plan and reflect these activities by becoming aware of their actions and learning processes. According to Endsley (1995, 2000) being aware of one’s own situation is a three level process and a prerequisite for making decisions and effectively performing tasks: the perception of elements in the current situation is followed by the comprehension of the current situation which then leads to the projection of a future status. Once a learner is aware of his situation, he “reflects on the phenomenon before him, and on the prior understandings which have been implicit in his behaviour” (Schön, 1983, p. 68) to then engage in a process of continuous learning. Reflection can promote insight about something that previously went unnoticed (Bolton, 2010) and lead to a change in learning behaviour. Therefore, results of learning analytics can be used to foster awareness and thus reflection (Verpoorten et al., 2011; Verpoorten, 2012; Govaerts et al., 2012) or to give recommendations for further steps in a current learning scenario (Greller and Drachsler, 2012). As Ferguson (2014) explains, learning analytics offers “ways for learners to improve and develop while a course is in progress. These analytics do not focus on things that are easy to measure. Instead, they support the development of crucial skills: reflection, collaboration, linking ideas and writing clearly” (para. 7). Awareness and reflection support for students are consequently highly important aims of learning analytics. The existence and impact of these aims, however, are hard to measure due to the lack of standards that the student support of learning analytics tools can be measured against.

[1] http://learning-analytics.info
[2] https://solaresearch.org/events/lak/
[3] https://solaresearch.org/events/lasi/
[4] https://solaresearch.org

The same applies to educators. In order to support students within a course, teachers should be aware of what the students are doing, how they are interacting with the course material, where comprehension problems arise (cf. Scheffel et al. (2011, 2012)). Especially if the number of students in a course is high and the tasks the students are engaged in are not trivial, teachers need assistance for keeping track of the students' activities, e.g. with the help of activity-based learner-models (Florian et al., 2011). Zinn and Scheuer (2006) conducted a survey among teachers trying to identify requirements for student tracking tools. Among the information deemed mostly important were the students' overall success rate, the mastery level of concepts, skills, methods and competencies as well as the most frequently diagnosed mistakes. Such information is also needed for the evaluation of a course, i.e. didactic concept, materials, contents, tools, and tests. Awareness and reflection support for educators are thus also highly important aims of learning analytics. But as with the learner support, standards that define quality indicators for learning analytics tools are missing.

While the added value of learning analytics for learners and educators is clearly recognised (Long and Siemens, 2011), little research has been done so far to compare the findings of empirical learning analytics studies and their tools as having a desirable effect on learning. We therefore propose to work toward an evaluation instrument that will help standardise the evaluation of learning analytics tools. We provide a first version of the evaluation framework for learning analytics (EFLA) to measure and compare the impact of learning analytics on educational practices.

The EFLA has been developed with experts from the learning analytics domain by using a group concept mapping (GCM) approach. The remaining parts of this chapter are organised in the following way: First, we will present the GCM methodology and provide some demographic description of the participants. Second, we will present and discuss the empirical findings of the study that reflect the learning analytics community's view on such evaluation criteria. Third, we will propose a first version of the evaluation framework for learning analytics. Fourth, we will further extend the findings of the GCM study with a focused literature study of related articles. Finally, we will conclude our results and provide some limitations and potential future research directions toward the application of the evaluation framework for learning analytics.

1.2 Group Concept Mapping

1.2.1 Method

One methodology to identify a group's common understanding of a given issue is group concept mapping. It is a very structured approach that applies quantitative as well as qualitative measures that create a stakeholder-authored visual geography of ideas from a target group, combined with specific analysis and data interpretation methods, to produce maps to guide planning and evaluation efforts on the issues of

Table 1.1 Overview of participants of the group concept mapping study

	started	finished
brainstorming	74	74
demographic questions	33	24
sorting	33	23
rating importance	24	21
rating feasibility	22	20

the group (Kane and Trochim, 2007). Our study makes use of a GCM online tool[5] and consists of three steps for the participants: (1) generation of ideas, i.e. quality indicators of learning analytics, (2) sorting of the collected ideas into clusters, and (3) rating of the ideas according to several values, i.e. importance and feasibility. The individual inputs of the participants are aggregated to reveal shared patterns in the collected data by applying statistical techniques of multidimensional scaling and hierarchical clustering. Visualisations then help to grasp the emerging data structures and to interpret the data. One important aspect of group concept mapping is its bottom-up approach. Instead of presenting a given set of criteria to sort and rate, the community itself generates the ideas that are to be clustered and rated by a group of experts.

1.2.2 Participants

The involvement of participants in our GCM study was twofold (see Table 1.1). The first phase was conducted during the days of the Learning Analytics and Knowledge Conference 2014. Calls for participation were circulated via several channels, e.g. Twitter, project websites, personal contact, email etc., asking people involved and interested in learning analytics to contribute their quality indicators for learning analytics to the brainstorming phase. Participation was accessible via a link and open, i.e. people did not have to register with the GCM tool. In total, 74 people participated in the brainstorming phase.

For the second phase, i.e. sorting and rating of the collected quality indicators, we selected 55 experts from the domain of learning analytics (i.e. they had been involved in the domain for several years, had published about learning analytics-related topics, were from the higher education sector and preferably had a PhD degree) and contacted them personally. Table 1.2 shows a summary of the demographics, the average expert of the study is a researcher at a university with an advanced expertise in learning analytics and has more than six or even more than ten years of work experience.

[5] `https://conceptsystemsglobal.com`

Table 1.2 Answers to demographic questions by participants of phase 2

participant question	option	frequency	%
expertise	novice	0	0.00
	intermediate	6	25.00
	advanced	11	45.83
	expert	7	29.17
		24	100.00
experience	less than 5 years	8	33.33
	6–10 years	5	20.83
	more than 10 years	11	45.83
		24	100.00
involvement	more in research	16	66.67
	more in teaching	1	4.17
	equally in research and teaching	4	16.67
	other	3	12.50
		24	100.00

1.2.3 Procedure

All participants of both phases were informed about the purpose, the procedure, and the time needed to complete the activities. Participants of the first phase were given a link to access the brainstorming section of the GCM tool and asked to generate ideas by completing the following statement: *"One specific quality indicator to evaluate the effects of learning analytics is …"*. Participants had ten days to contribute to the brainstorming. During this first phase the 74 participants generated a total of 92 original ideas. Before releasing the list of statements into the second phase, identical statements were unified and too vague ideas, e.g. "Range of flexibility in moving from one point to another in a theoretical discussion", were taken out by us. Also, those statements that contained more than one idea were split, e.g. "students and teachers change their behaviour in some aspects" was split into one statement for teachers and one for students.

After the cleaning process, the now 103 statements (the full list is given in Appendix A) were randomised and pushed into the sorting and rating phase. Participants first sorted the statements according to their view of the statements' similarity in meaning or theme and were asked to also name the clusters. Dissimilar statements were not to be put into a 'miscellaneous' cluster but rather into their own one-statement-cluster in order to ensure statement similarity within the clusters. Then, participants rated all quality indicators on a scale of 1 to 7 according to their importance and feasibility, with 1 being of lowest and 7 being of highest importance/feasibility. Participants had two weeks to complete the sorting and rating activities.

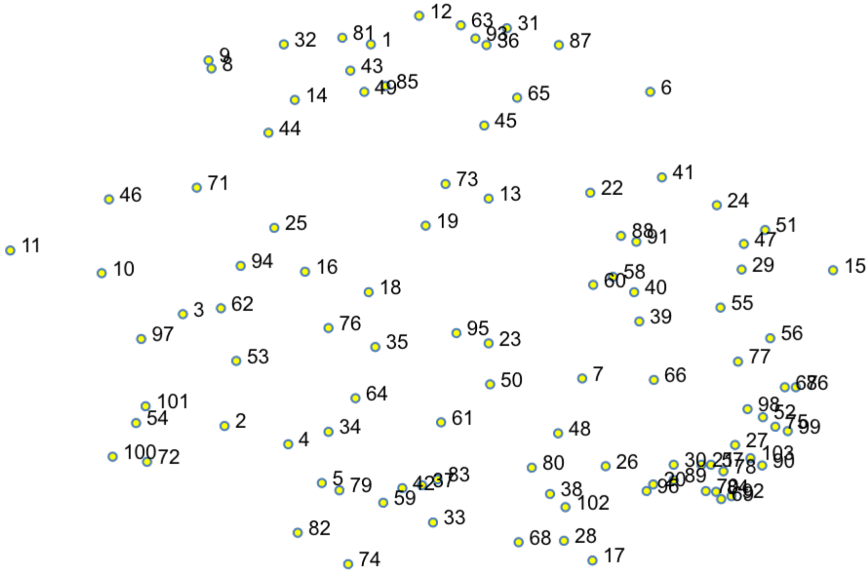


Figure 1.1 Point map of the 103 quality indicators

1.3 Results

1.3.1 Point Map

The GCM tool offers a number of automated analyses of the collected data: multidimensional scaling and hierarchical clustering for the sorting data and mean, standard deviation and correlation for the rating data. Figure 1.1 shows a point map of the 103 collected ideas, i.e. the outcome of the multidimensional scaling analysis. The multidimensional scaling analysis assigns a so-called bridging value between 0 and 1 to each statement. Statements with low bridging values have been grouped very close together with other statements around it, e.g. statements 98, 52, 75, 99 on the lower right side of Figure 1.1 all deal with some form of student motivation and can be considered quite coherent. Statements with higher bridging values can also be grouped together but the surrounding statements are then further apart, e.g. statements 95, 23, 50, 61 about teacher motivation, engagement and feedback. Thus, statements that are close to one another in the map are also close to one another in meaning and have thus been clustered together by the experts.

1.3.2 From Point Map to Cluster Map

In some areas of the map it is quite easy to detect groups by simply looking at the point map. In other areas, however, it is more difficult to decide where group boundaries could be. The hierarchical clustering analysis of the GCM tool offers

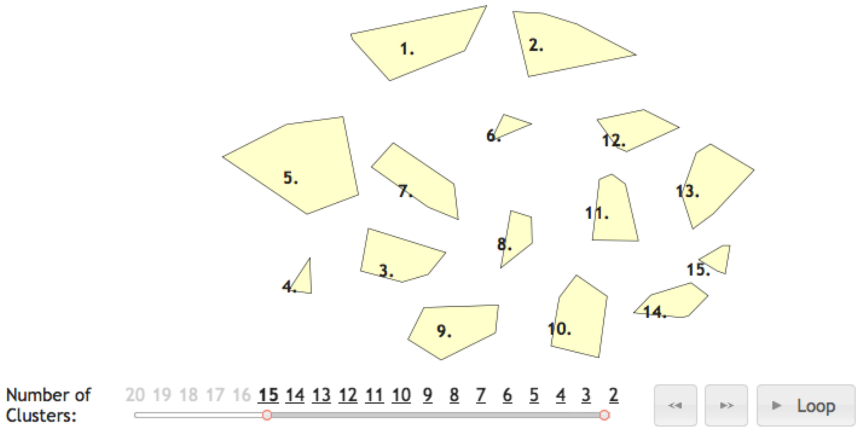


Figure 1.2 Replay map showing 15 clusters

several solutions to a given point map. We used a cluster replay map, starting at 15 clusters and working down to two (see Figure 1.2). For each cluster-merging step we carefully looked at the statements of the clusters that were to be combined to check whether a merging made sense. The solution that seemed best to be representing the collected data and the purpose of the study was the one with eight clusters.

After deciding on the number of clusters to work with, meaningful labels needed to be constructed for the clusters. The system automatically suggests a list of labels per cluster. Another way of finding appropriate labels is to look at the bridging values of the statements within a cluster. The lower the bridging values are, the better do those statements define the cluster. A third way to find meaningful cluster labels is to find the overarching theme of a cluster by looking at all statements of a cluster. We combined all three methods to define the labels of the 8-cluster solution (see Figure 1.3): *1. Data: open access*, *2. Data: privacy*, *3. Acceptance & uptake*, *4. Learning outcome*, *5. Teacher awareness*, *6. Learning performance*, *7. Learning support*, and *8. Student awareness*. A list of all clusters, the statements they contain and the statements' bridging values is given in Appendix B.

1.3.3 Cluster Descriptions

Cluster 1 *Data: open access* contains eleven statements with bridging values ranging from 0.06 to 0.60. Most statements deal with aspects of openness and transparency of the used data as well as the used algorithms, e.g. "that data are open access", "portability of the collected data".

Cluster 2 *Data: privacy* is about exactly that: privacy, control of data, and transparency of data access. There are eight statements in the cluster with bridging values ranging from 0.10 to 0.72. Representative statements are "that privacy is ensured", "if learners can influence which data are provided".

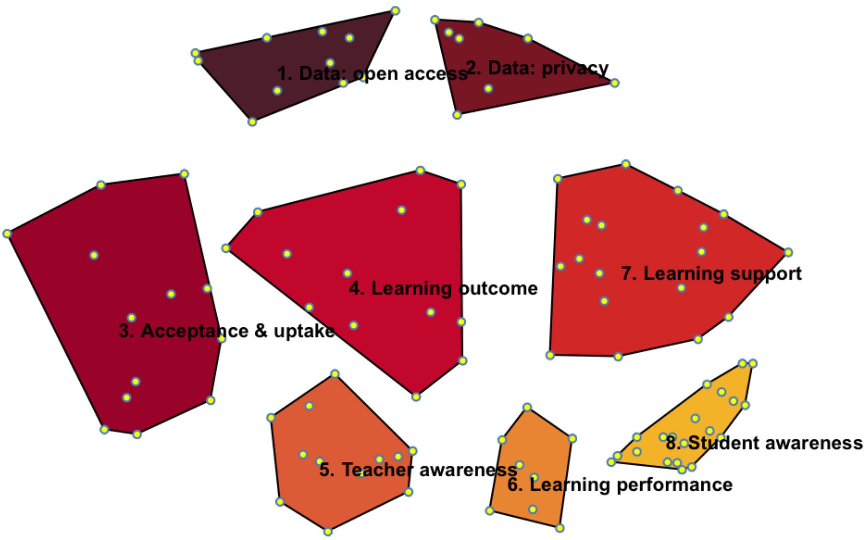


Figure 1.3 Cluster map with labels

Cluster 3 about *Acceptance & uptake* contains 13 statements and is very diverse as can be seen from the bridging value range from 0.66 to 1.00. The cluster describes aspects of acceptance of learning analytics and its results by different stakeholders but also the comparability of methods or the context and objectives dependence of learning analytics. An example statement is "that administrators invest in scaling successful tools across their programming".

Cluster 4 *Learning outcome* is also somewhat diverse with a bridging value range from 0.19 to 0.87. It contains 13 statements that deal with comparability of learning analytics results, teacher motivation, result accuracy and feedback for teachers, e.g. "if teachers are able to gain new insights using the given LA methods", "that LA results are compared with other (traditional) measures".

Cluster 5 *Teacher awareness* consists of twelve statements with bridging values from 0.18 to 0.73. Most statements are connected to teachers changing their course material or their teaching behaviour in response to learning analytics results about their students: "that teachers change their behaviour in some aspects", "that teachers react in a more personalized way to how their students are dealing with learning material".

Cluster 6 *Learning performance* is one of the smallest clusters as it consists of only eight statements. The bridging value range is relatively small, i.e. 0.11 to 0.59. Statements in this cluster are about student performance, learning and achievement improvement. Representative statements are "that change in workplace learning is measurable", "the extent to which the achievement of learning objectives can be demonstrated".

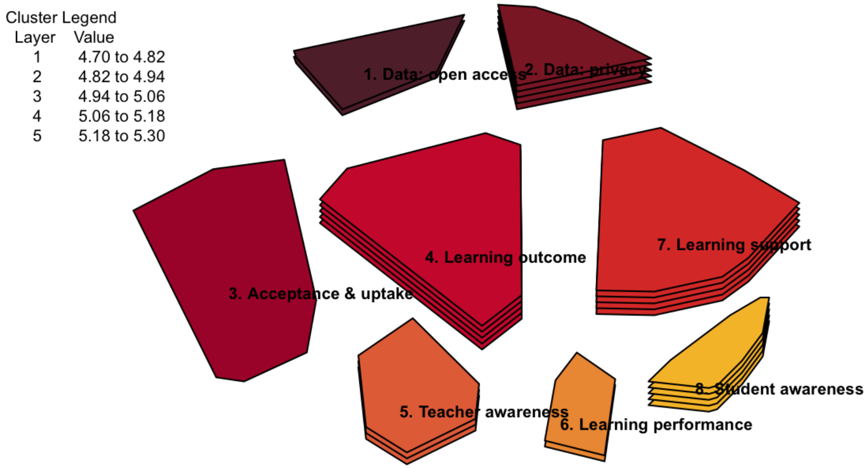


Figure 1.4 Rating map on *importance*

Cluster 7 *Learning support* is a very stable but also rather large cluster with 18 statements. Its bridging values range from 0.14 to 0.76. Statements in this cluster are often formulated generally and deal with support for teachers as well as for students, e.g. “an early detection of students at risk”, “the ability to explain what could help to further improve”, “that students regularly utilize the tools provided”.

Cluster 8 about *Student awareness* is the largest and most coherent cluster. It contains 20 statements and its bridging value range is from 0.00 to 0.43. The cluster is also very stable and consistent. All statements are related to students, their achievement, success, self-regulation, awareness, learning behaviour and motivation, e.g. “that students become more self-regulated in their learning processes”, “that students are more aware of their learning progress”.

1.3.4 Rating Maps

Once the cluster map is settled upon, the experts’ ratings of the quality indicators can be included in the calculation as well. Two aspects were given to the experts to be rated on a scale from one to seven (one for a low, seven for a high rating): *importance* and *feasibility*. While the former refers to the priority or importance of an item in relation to the evaluation of effects of learning analytics, the latter indicates the perceived ease of applicability of an item. The GCM tool automatically applies the experts’ ratings to the cluster map, indicating the importance or feasibility by layering the clusters. The system always divides the ratings into five layers based on the average ratings provided by the participants for the rating maps. The anchors for the map legend are based on the high and low average ratings across all of the participants. One layer indicates a low rating, whereas five layers indicate a high rating of the respective aspect.

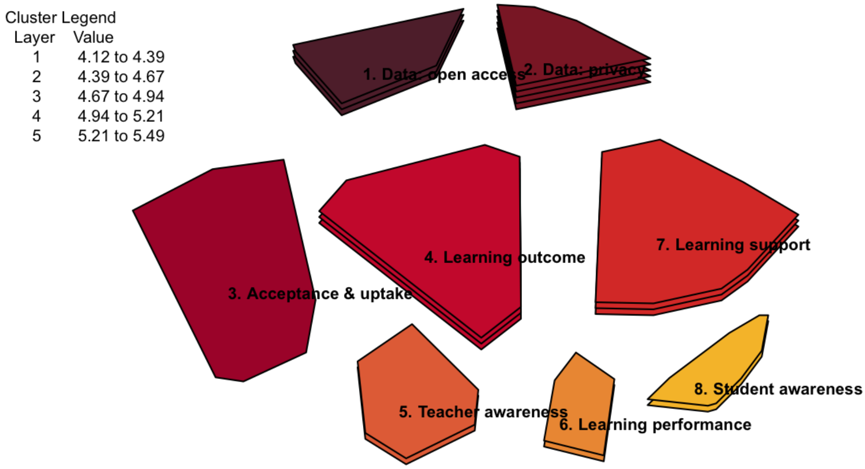


Figure 1.5 Rating map on *feasibility*

Figure 1.4 shows the rating map according to the *importance* aspect. Clusters *Data: privacy*, *Learning outcome*, *Learning support*, and *Student awareness* each received very high importance ratings as they all have five layers. *Teacher awareness* has three layers, while *Data: open access* and *Learning performance* have two layers each and *Acceptance & uptake*, i.e. the least coherent cluster, has only one layer.

Looking at the *feasibility*-rating map (see Figure 1.5) one can see a change in the rating behaviour of the experts. Although the *Data: privacy* cluster also gets five layers and is thus deemed highly feasible by the experts, the other three very important clusters have been rated less feasible: *Learning outcome* and *Learning support* only receive an intermediate level of feasibility with three layers each. *Student awareness*, a highly important cluster, receives a low feasibility rating with two layers only. *Teacher awareness* also drops down to two layers. The cluster dealing with *Acceptance & uptake* was seen as neither important nor feasible by the experts. The only cluster that receives more layers in the *feasibility*-rating map than in the *importance*-rating map is *Data: open access* and is thus deemed more feasible than it is important.

A ladder graph (see Figure 1.6) offers a form of visualisation that is well suited to compare the clusters' ratings according to *importance* and *feasibility*. The rating values are based on a cluster's average rating. A Pearson product-moment correlation coefficient ($r = 0.65$) indicates a strong positive relationship between the two aspects of *importance* and *feasibility*. For both ratings, the *Data: privacy* cluster receives the highest values while *Acceptance & uptake* receives the lowest. As was already observable from the rating maps, the three clusters about *Learning outcome*, *Learning support* and *Student awareness* have all been rated as very important but as much less feasible.

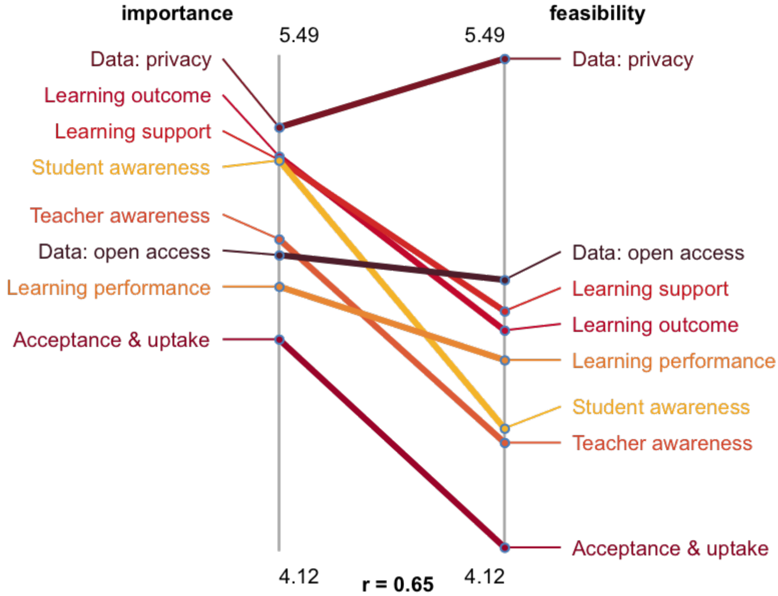


Figure 1.6 Ladder graph of the rating values for the clusters

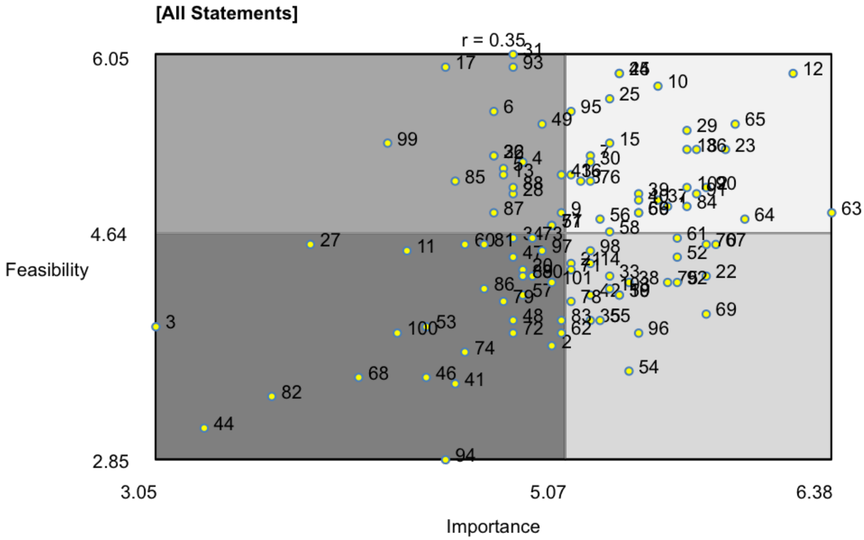


Figure 1.7 Go-zone graph of all 103 statements

A third visualisation the GCM tool offers for the rating aspects are so-called go-zones, i.e. bivariate graphs that allow to explore the statements in relation to their ratings more deeply. A Go-zone graph maps each statement onto a space between x- and y-axis based on the mean values of the two rating aspects of *importance* and *feasibility*. Go-zone graphs can be created for all statements together or for individual clusters. Figure 1.7 shows the go-zone graph for all 103 statements. The graphs for the individual clusters can be found in Appendix C. Go-zone graphs are very supportive for the selection of suitable indicators for the framework as they highlight those statements with a good balance of *importance* and *feasibility*. When deciding on the individual indicators for the evaluation framework it can also be sensible to choose statements from the only feasible or only important quadrant if they are close enough to the upper right quadrant and support a given dimension.

1.4 Discussion

1.4.1 Constructing the Framework

Looking at the clusters in Figure 1.3, their coherence is also observable visually. One can see that the four most coherent clusters (*Data: open access*, *Data: privacy*, *Learning performance* and *Student awareness*), i.e. the ones with smaller bridging values, are the smaller ones in relation to area size and that the three least coherent clusters (*Acceptance & uptake*, *Learning outcome* and *Learning support*), i.e. the ones with higher bridging values, are much larger in area size. The two most stable clusters are the ones about *Learning support* and *Student awareness*, i.e. they both remained stable until the five-cluster solution while the others merged. This implies a fairly high agreement between the experts' sorting and the system's multidimensional scaling and hierarchical clustering. We therefore take cluster coherence and stability to be a first indication of relevance when trying to settle on the dimensions of the evaluation framework.

Also very interesting conclusions can be drawn when comparing the two rating maps (see Figure 1.4 and Figure 1.5) with one another. The *Acceptance & uptake* cluster received low ratings for *importance* as well as for *feasibility*. The experts' low rating is also supported by the cluster's coherence. With an average bridging value of 0.86 and individual statement bridging values spanning from 0.66 up to 1.00, the cluster contains a rather diverse collection of statements. While constructing the framework to evaluate effects of learning analytics we therefore focus on all other clusters first in order to find suitable dimensions and items before taking this cluster into account as the statements it contains are too incoherent, too vague, unimportant and unfeasible as a group.

This leaves us with a slightly different but nonetheless very interesting cluster landscape: The two clusters in the North (1, 2) both deal with data, access, methods, algorithms, transparency and privacy, i.e. with technical issues, while the clusters in the South (5, 6, 8) deal with awareness, reflection, performance and behavioural change of students and teachers, i.e. with human issues. The 'technical North' (*Data: open access* and *Data: privacy*) and the 'human South' (*Teacher awareness*, *Learning*

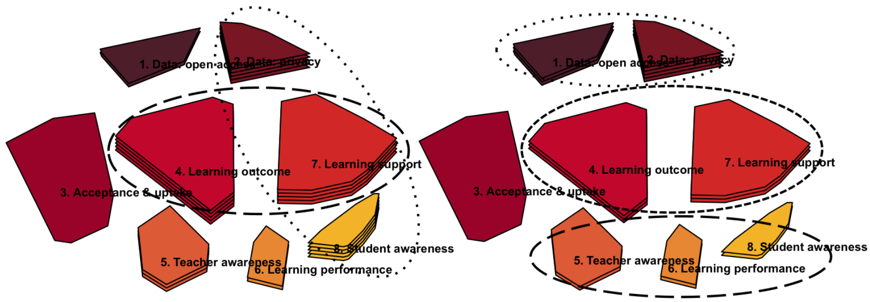


Figure 1.8 Comparison of the rating maps of *importance* (left) and *feasibility* (right)

performance and *Student awareness*) are bridged by a wide layer of learning-related clusters (*Learning outcome* and *Learning support*). Apart from the North-South view, one can also look at the map with an East-West perspective: The three Eastern clusters (*Data: privacy*, *Learning support* and *Student awareness*) are more concerned with issues during the learning process while the Western clusters (*Data: open access*, *Learning outcomes*, *Teacher awareness* and *Learning performance*) are slightly more concerned with issues of learning output and results. This division is of course not to be seen strictly, but these groupings clearly show a thematic tendency. As for the construction of the framework, we conclude that the aspects of technology, stakeholders (humans), learning processes and learning outcomes should all be reflected in the dimensions.

Taking the two rating aspects *importance* and *feasibility* into account, we get two different versions of the landscape described above. The *importance* map on the left side of Figure 1.8 clearly shows that the learning-related middle layer, i.e. the clusters about *Learning outcome* and *Learning support* within the dashed line, is deemed highly important by the learning analytics experts. But all Eastern clusters, i.e. the ones about *Data: privacy*, *Learning support* and *Student awareness* within the dotted line, also receive five layers of importance. Generally one can thus say that the focus of *importance* is on the learning process-related clusters. For the *feasibility* map on the right side of Figure 1.8 the landscape shifts. Now there is a clear North-South divide: The technically-oriented clusters in the North (dotted circle) are deemed most feasible by the experts, followed by the learning-related layer in the middle (short dashed circle) and concluded by the human-related clusters in the South (long dashed circle). This again supports the construction of the framework's dimensions according to the data-learning support & process-stakeholder view.

Looking at the ladder graph in Figure 1.6 allows a closer look at the differences in average ratings for the different clusters. Especially the drop in *feasibility* compared to *importance* for a number of clusters is quite obvious. The most striking drop is that of the *Student awareness* cluster. The experts think student awareness to be quite an important aspect to take into consideration when evaluating effects of learning analytics theoretically but deem it difficult to apply in real world settings. This can be explained with the fact that many teachers, and thus very likely also the learning analytics experts involved in this study, do not think students to be capable enough of judging their own learning processes and progresses as it has been identified by Drachsler and Greller (2012).

When deciding upon the dimensions of the framework it is important to find a good trade-off between the *importance* ratings and the *feasibility* ratings. Due to the high importance of the clusters about *Data: privacy*, *Learning outcome*, *Learning support* and *Student awareness* it seems to be sensible to use them as a basis for the dimensions of the framework. The feasibility ratings of the clusters can then be used to associate the remaining clusters with these four dimensions: The *Data: privacy* cluster is by far the most feasible one, followed by the *Data: open access* cluster. The two can thus be combined into one data dimension. The next two clusters on the *feasibility* scale are *Learning support* and *Learning outcome*, two of the dimension candidates that stay on their own due to their high importance rating. As the latter cluster is followed by the *Learning performance* cluster and as they both deal with learning results and effects, it makes sense to construct a combined dimension from them. The next cluster on the *feasibility* scale is *Student awareness*, closely followed by *Teacher awareness*. Both of them are "human clusters" and concerned with awareness, reflection and behavioural change. They can therefore be combined into one dimension even though they address different stakeholders.

1.4.2 Outline of the Framework

From the results of the GCM study we can identify four topic areas that can be turned into dimensions for the framework: the first deals with anything related to data, algorithms, transparency and privacy. It is based on the clusters *Data: privacy* and *Data: open access*. For the sake of simplicity the dimension is called *Data Aspects*. It contains the items *Transparency*, *Data Standards*, *Data Ownership*, and *Privacy*. The second topic area concerns support for students and teachers during the learning process, i.e. while using learning analytics tools. It is entirely based on the *Learning Support* cluster and also takes over this name. The items of this dimension are *Perceived Usefulness*, *Recommendation*, *Activity Classification*, and *Detection of Students at Risk*. The third topic area deals with the results at the end of the learning process, i.e. any issues of output, consequence, performance, outcome etc. In this case, however, it is not primarily to be seen in relation to individual student performance, e.g. their grades, but refers to the learning analytics tools' results and outcomes. It is comprised of the two clusters *Learning outcome* and *Learning performance*. The dimension is named *Learning Measures and Output* and contains the items *Comparability*, *Effectiveness*, *Efficiency*, and *Helpfulness*. The fourth topic area contains the items *Awareness*, *Reflection*, *Motivation*, and *Behavioural Change* of students and educators during the learning processes, i.e. it is about the educational aims identified at the beginning of this article. This dimension is called *Objectives*.

Most statements related to stakeholders are about learners and teachers and hardly about institutions. This is partly due to the fact that we did not take the cluster

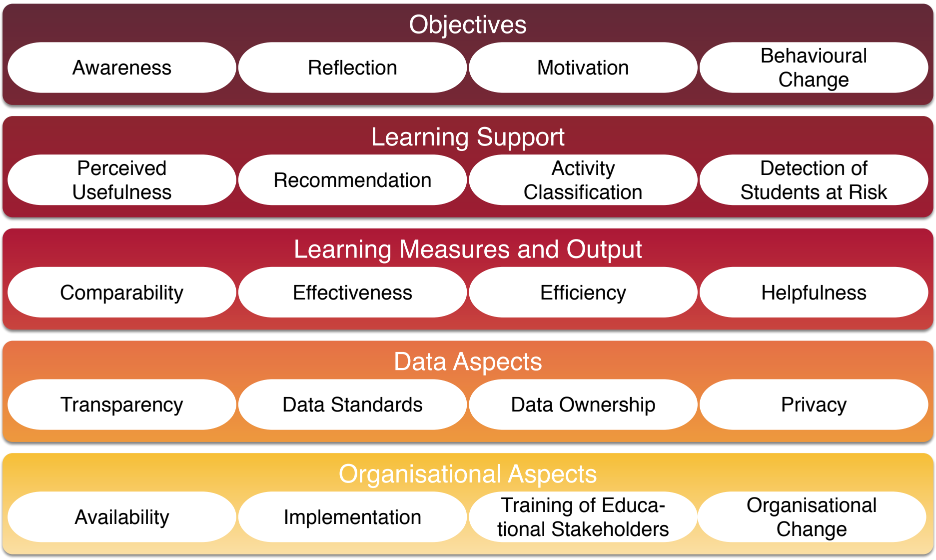


Figure 1.9 First version of the evaluation framework for learning analytics (EFLA-1)

Acceptance & uptake that contains some statements about this into account right away. It is also due to such statements being spread over all clusters. As we consider indicators of organisational issues to be an important aspect when considering the evaluation of learning analytics tools (cf. (Arnold et al., 2014a)), we decided to add a fifth dimension to the framework: *Organisational Aspects* containing the items *Availability*, *Implementation*, *Training of Educational Stakeholders*, and *Organisational Change*.

The dimensions' items are based on a review of the statements in the go-zone graphs of each cluster. In most cases the statements related to these items can be found in the upper right quadrant of the go-zone graphs. In some cases statements from the only feasible or only important quadrant were chosen as well if they are close to the important and feasible quadrant. The statements chosen for each dimension are then combined and turned into slightly shorter, more general statements that clearly represent an indicator for a given dimension. Figure 1.9 shows the first version of the evaluation framework for learning analytics (EFLA-1), i.e. five dimensions with four items each.

1.5 Literature Supporting the Framework Dimensions

In this section we present a focused literature review to further extend the GCM study with the latest insights from the learning analytics community. It is structured according to the different layers of the framework. Although the three dimensions

Objectives, *Learning Support*, and *Learning Measures and Output* are clearly separable from one another in regards to their items and purposes, it is often difficult to exclusively attribute findings from the literature to one of the dimensions only. Literature concerning these three dimensions has therefore been combined into one section.

1.5.1 Objectives, Learning Support, and Learning Measures and Output

Some works on awareness (Endsley, 1995, 2000; Charlton, 2000) and reflection (Schön, 1983; Bolton, 2010) have already been mentioned in the introduction of this chapter. They all deal with these educational and pedagogical concepts in general and are not directly attached to the domain of learning analytics. Their findings, however, matter to this domain. Work by McAlpine and Weston (2000) also deals with reflection as a general concept in educational settings. They argue that "reflection is not an end in itself, but a mechanism for improving teaching and hence maximizing learning" (p. 382).

Studies in the related domain of technology-enhanced learning reveal several aspects that can be used for outcome measurement of recommender systems (Drachsler et al., 2009) but could also be used for the analysis of other educational technologies and learning analytics tools. The first measurement category is a technical one with the parameters of accuracy, coverage and performance. The second measurement category covers educational aspects and involves the parameters of effectiveness, efficiency, satisfaction and drop-out rate. Social network measures form the third category with parameters of variety, centrality, closeness and cohesion.

Clow (2012) points out that "learning analytics should generate metrics that relate to what is valued in the learning process. If the final assessment rewards undesired behaviour, improving the control system to more effectively optimise the results will make the learning worse" (p.137). Clow therefore identifies three strategies by which the effectiveness of learning analytics can be improved: (1) enhancement of the speed of response, e.g. real-time feedback rather than summarising feedback, (2) enhancement of the scale of response, e.g. feedback to more than one stakeholder, and (3) improvement of the quality of an intervention, e.g. testing of the intervention or participation of more stakeholders.

Course Signals is an early intervention solution for collegiate faculty (Arnold and Pistilli, 2012) and serves as an example tool of implemented learning analytics. With this tool, teachers can provide feedback to students about their performance and predicted progress. The feedback is comprised of a personalised email and a progress visualisation, i.e. a traffic light signal. Courses that used the tool showed a strong increase in positive grades and at the same time a decrease in negative grades and withdrawals. Both with teachers and with students, Course Signals received positive overall experiences although teachers approached it with more caution than students.

This caution can be set in relation to the findings of a survey among teachers and researchers of learning analytics. The study revealed that "trust in learning analytics algorithms is not well developed" yet (Drachsler and Greller, 2012, p. 127). Many educators hesitate to take the calculations of algorithms about learning and educational effects as valid while at the same time they hope to gain new insights from those analytics results. The study also showed that for many participants the application of learning analytics cannot provide a more objective assessment than they could do on their own and that a proper assessment of a learner's state of knowledge is not possible.

A combination of learning analytics and action research to support teachers in educational settings is presented by Dyckhoff et al. (2013). They describe possible effects of learning analytics on teaching and investigate how this could be evaluated. For them, learning analytics tools should be useful to achieve the set goals in a given context. Their findings show that in many cases learning analytics tools do not yet answer all of the questions that teachers have in regard to their educational setting. This especially concerns qualitative analysis as well as data correlation from different sources. Quantitative results, however, are often easily available. Among others, the authors relate these shortcomings to an insufficient involvement of teachers in the design process of learning analytics tools and the lack of appropriate, diverse data sources, e.g. student profile data, and mobile data. Dyckhoff et al. (2013) conclude that "there is a necessity for creation of evaluation tools to measure the impact and effects of learning analytics on the learning process" (p. 227) and for mechanisms to support and reassure awareness and reflection, as well as to improve teaching processes.

An example of work dealing with the design of pedagogical interventions is that of Wise (2014). The author presents four principles of pedagogical learning analytics intervention design: (1) integration, (2) agency, (3) reference frame, and (4) dialogue. Teachers and course developers can build upon these principles in order to support students in productively making use of learning analytics. Wise also describes three core processes that students should be engaged in: (1) grounding, (2) goal-setting, and (3) reflection. The principles together with the core processes form a model of pedagogical learning analytics intervention design.

1.5.2 Data Aspects

One important aspect when dealing with data-related issues is the availability of data sets and data standards. While a few years ago, open access to data sets was hardly constituted (Drachsler et al., 2010), the years since then have shown an immense rise in the availability of and open access to data sets for the technology-enhanced learning, learning analytics and educational data mining domains. Verbert et al. (2012) provide an overview of existing datasets and analyse them along the dimensions of their framework for the analysis of educational datasets. There are three dimensions: (1) dataset properties, (2) data properties, and (3) learning analytics objectives. Several initiatives are now offering access to educational data sets such as the LinkedUp Project[6] with its LinkedUp Dataset and LinkedUp Data Challenge, the LAK Dataset (Taibi and Dietze, 2013), the DataHub[7] and the PSLC DataShop (Koedinger et al., 2010).

A number of legal, risk and ethical issues that should be taken into account when implementing learning analytics at educational institutions in the United Kingdom is presented by Kay et al. (2012). They describe that these institutions have to find a balanced way to assure educational benefits, that they are under as much competitive pressure as organisations in the consumer world and that they need to satisfy the expectations of the now arising born digital generations of learners. The authors suggest four principles that provide good practice when tackling the above-mentioned conflicts: (1) clarity, (2) comfort and care, (3) choice and consent, and (4) consequence and complaint.

Willis et al. (2013) apply The Potter Box, i.e. an ethical model in business communications, to learning analytics. They conclude that institutions will have to "balance faculty expectations, various federal privacy laws, and the institution's own philosophy of student development. It is therefore critical that institutions understand the dynamic nature of academic success and retention, provide an environment for open dialogue, and develop practices and policies to address these issues" (Conclusion section, para. 2).

During the EDUCAUSE IPAS Summit in 2013 participants were asked to discuss issues associated with managing risk in student success systems and to identify opportunities for the development of such systems (EDUCAUSE, 2014). More specifically, the discussions focused on three aspects: (1) the identification of internal and external drivers that encourage the implementation of learning analytics, (2) the identification of institutional risks, documentation of effective practices and review of existing and new solutions, and (3) the development of strategies that already take risk issues into account during the design of learning analytics processes. The authors conclude that existing and new data sources have to be integrated in a better way and that educational institutions should know exactly which data they collect for what purpose and who has access to that data. Institutions should also address the movement of students and their data between institutions and should not misuse the collected data to predetermine a student's success.

An analysis of privacy and ethical issues specific to the context of learning analytics and its related research as well as guidelines about how to comply with common privacy principles are presented by Pardo and Siemens (2014). These principles are conceived from the review of learning analytics proposals, government frameworks and regulatory directives and allow educational institutions to assess their current level of compliance in order to then possibly improve their privacy-related matters. The principles are: (1) transparency, (2) student control over data, (3) right of access / security, and (4) accountability and assessment.

[6] `https://linkedup-project.eu`
[7] `https://datahub.io`

Also relevant for the *Data Aspects* dimension is the methodology based on value-sensitive design that incorporates ethical and legal considerations and requirements throughout the research and development cycle of technology as Friedman (1997) explains. Value-sensitive design is the idea that ethical analysis and reflection needs to take place when and where it can make a difference for the design and governance of technology: starting early on in the design and development process, and close to where the technology is being shaped and designed. Ethical considerations concern first of all the privacy of individuals taking part in the system. A high degree of configurability, the provision of meaningful default options that relate to a privacy-by-default approach, combined with informative explanations given to users are some of the ingredients that will allow the achievement of the notion of informed consent (van den Hoven, 2008).

1.5.3 Organisational Aspects

In the 21st century, more and more higher education organisations apply learning analytics to optimise student success. According to Norris and Baer (2013) such intelligent investments from the organisations have a strong and justifiable return on investment: the implementation of enhanced analytics is to be seen as critical for student success on the one hand and achieving institutional effectiveness on the other as without it, organisations cannot meet the current gold standard for institutional leadership. Norris and Baer conducted a survey among institutional practitioners and vendors about the building capacity in analytics to improve student success and how they determine the state of practice and gaps between needs and solutions. They interviewed 40 leading institutions from the American higher education sector as well as 20 technology vendors and came up with a framework for optimising student success through analytics that contains seven elements: (1) manage the student pipeline, (2) eliminate impediments to retention and student success, (3) utilise dynamic, predictive analytics to respond to at-risk behaviour, (4) evolve learner relationship management systems, (5) create personalised learning environments / learning analytics, (6) engage in large-scale data mining, and (7) extend student success to include learning, workforce, and life success.

In their discussion paper, Siemens et al. (2013) present a national learning analytics strategy to the Australian Government after undertaking a four step process: First, they evaluated the benefits of analytics in other sectors than education, then had a closer look at the data collection policies on provincial, territory, state and national level, followed by a review of universities around the world that are already developing analytics strategies, and finally, they inspected the role corporate partners can play in helping universities achieve analytics competence. Their five final suggestions are: (1) Australian higher education leaders should coordinate a high level learning analytics task force with a variety of stakeholders, (2) existing national data and analytics strategies should be leveraged, (3) guidelines for privacy and ethics should be established, (4) a coordinated leadership program should be set up, and (5) open and shared analytics curricula should be developed with the learning analytics community. Although their paper focuses on the learning analytics situation in Australia, the findings can be applied to other countries as well.

Arnold et al. (2014a) tackle the readiness of institutions to implement learning analytics. Instead of only looking at the maturity of an institution's already implemented learning analytics solution, the authors try to investigate how institutions that do not apply any analytics yet can become mature to do so. Their Learning Analytics Readiness Instrument (LARI) survey was conducted at nine higher education institutions and focuses on five readiness components for learning analytics implementations: (1) ability, (2) data, (3) culture and process, (4) governance and infrastructure, and (5) overall readiness perception.

With the help of learning analytics educational institutions are able to tune or correct the inner workings of their programs. Méndez et al. (2014) present five techniques that allow institutions to gain such insights: (1) difficulty estimation, (2) dependence estimation, (3) curriculum coherence, (4) dropout and enrolling paths, and (5) a load/performance graph. For their example analysis the authors used data from 2543 undergraduate computer science students at the ESPOL University in Ecuador spanning from 1978 until 2012. With their large study the authors want to show how simple analytics can be used to re-design whole program curricula.

Finally, in their panel discussion at LAK 2014, Arnold et al. (2014b) argue that "in order to truly transform education, learning analytics must scale and become institutionalized at multiple levels throughout an educational system" (p. 257). During the discussion, panel participants focused on five areas related to the adoption of learning analytics: (1) technology infrastructure, analytics tools and applications, (2) policies, processes, practices and workflows, (3) values and skills, (4) culture and behaviour, and (5) leadership. From the discussed case studies the authors conclude that institutions have to put effort and intention into planning the implementation and adoption of learning analytics. They suggest using existing research and theory as a foundation when beginning to build new theories and research about system level thinking.

1.6 Conclusion

This chapter proposed the first version of the evaluation framework for learning analytics (EFLA-1) to help standardise the evaluation of learning analytics tools. The work was motivated by the lack of evaluation standards that define indicators of learning analytics tools. After introducing the objectives of learning analytics, we presented a GCM study with experts from the learning analytics domain to identify a list of indicators. With the help of a number of analysis steps within the GCM tool we first created a point map of the statements that we then turned into a cluster map including cluster labels. The experts' ratings on importance and feasibility of the statements allowed us to further narrow down the list of possible dimensions as well as indicators. After taking the rating maps, the ladder graph and the go-zone graphs into account, we were able to propose a first version of the evaluation framework for learning analytics (see Figure 1.9) with the following five dimensions and their items:

- ***Objectives*** (*Awareness*, *Reflection*, *Motivation*, *Behavioural Change*)
- ***Learning Support*** (*Perceived Usefulness*, *Recommendation*, *Activity Classification*, *Detection of Students at Risk*)
- ***Learning Measures and Output*** (*Comparability*, *Effectiveness*, *Efficiency*, *Helpfulness*)
- ***Data Aspects*** (*Transparency*, *Data Standards*, *Data Ownership*, *Privacy*)
- ***Organisational Aspects*** (*Availability*, *Implementation*, *Training of Educational Stakeholders*, *Organisational Change*)

In order to extend the found dimensions we conducted a focused literature review to show their usage within the community so far.

Limitations of our current approach are related to the participants of our GCM study: Most participants work at a university and are more research- than practice-oriented. It would be interesting to see whether and how the framework and its indicators would change if (high) school teachers and/or more practice-oriented university staff were involved in the process.

For our future research we aim to transfer the findings into a concrete evaluation instrument, i.e. a questionnaire, that allows learning analytics stakeholders to evaluate any given learning analytics tool. The initial application of the evaluation instrument will be done within the context of higher education institutions. A good mean for this approach is the Learning Analytics Community Exchange project[8] that focuses on the exchange of best practises and the collection of evidence in the field of learning analytics.

[8] `http://www.laceproject.eu`

Chapter 2

Back In The Crowd – Or: Developing an evaluation framework for learning analytics

This chapter presents results from the continuous process of developing the evaluation framework for learning analytics (EFLA). Building on the previous group concept mapping study, this chapter presents how the EFLA-1 is turned into an applicable tool, i.e. a questionnaire, which is then used by a group of learning analytics experts in order to evaluate the framework. Using the quantitative and qualitative results of this evaluation study, useful insights are gained about the characteristics of an evaluation framework that are carried over into the creation of the next framework version. For this the data from the group concept mapping study presented in Chapter 1 is reconsidered under the light of the results of the evaluation study and combined with a look at related literature. Additionally, further literature is consulted to motivate and theoretically ground the items of the improved evaluation framework. Finally, this chapter concludes with the presentation of the second version of the evaluation framework for learning analytics (EFLA-2).

This chapter is based on:

Scheffel, M., Drachlser, H., and Specht, M. (2015). Developing an evaluation framework of quality indicators for learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, LAK '15, pages 16–20, New York, NY, USA. ACM.

and

Scheffel, M., Drachsler, H., and Specht, M. (in preparation for submission to LAK18). *Eeny meeny miny moe, catch all items by the toe*.

2.1 Introduction

Over the years that learning analytics (LA) have become more and more prominent, the number of tools and applications using such techniques as well as publications about them has grown rapidly. And although the added value of learning analytics for learners as well as for educators has clearly been recognised (Long and Siemens, 2011), research on the comparability of empirical LA studies is sparse. The comparison of learning analytics approaches, i.e. their measures, algorithms, results, effects, etc., is hardly possible due to the lack of a comprehensive knowledge base about what makes a good, effective, efficient, useful learning analytics tool in a given situation.

We therefore developed the Evaluation Framework for Learning Analytics (EFLA) to help standardise the evaluation of learning analytics tools (Scheffel et al., 2014)[1]. The framework comprises five dimensions (*Objectives*, *Learning Support*, *Learning Measures and Output*, *Data Aspects*, and *Organisational Aspects*) with four items each. In order to ensure an organically grown and accepted evaluation framework, stakeholders active in the domain of learning analytics have been involved in the development process of the first EFLA using a group concept mapping (GCM) approach.

The aim of this chapter is two-fold. We first present an evaluation study (Section 2.2) in order to find out whether the EFLA-1 is applicable to evaluate learning analytics tools or whether it needs to be further adapted, changed, restructured or defined differently for another evaluation cycle. The section about this study is structured as follows: we first present the methodology to evaluate the framework by applying it to a number of LA tools, followed by the presentation of quantitative as well as qualitative study results. We then revisit the first framework version taking the previous results into account and present ways to work towards an improvement of the framework for the next evaluation cycle. After taking stock we present a follow-up study (Section 2.3) about the construction process of the EFLA-2 for which the data from the GCM (Chapter 1) is reconsidered under the light of the results of the first evaluation study and combined with a look at related literature. Additionally, further literature is consulted to motivate and theoretically ground the items of the improved evaluation framework. Finally, this chapter concludes with the presentation of the second version of the evaluation framework for learning analytics (EFLA-2).

2.2 Framework Evaluation Study

2.2.1 Method

For the evaluation of the first framework version two things had to be done: on the one hand the EFLA-1 needed to be turned into an applicable tool itself and on

[1] This publication is included as **Chapter 1** in this thesis.

the other hand a collection of learning analytics tools to validate the framework against had to be compiled. As a first step, the EFLA's dimensions and items were therefore transformed into a questionnaire using Google Forms[2]. For every item the questionnaire asked (1) whether that item was present in/supported by a tool or not or whether it was not applicable, (2) in what way that item was present in/supported by the tool, and (3) how difficult or easy (on a scale of 1 (very difficult) to 5 (very easy)) it was to judge that item. At the end of each dimension section participants were offered an open text box asking for any additional comments.

To find suitable learning analytics tool candidates the submissions to the previous Learning Analytics and Knowledge conferences as well as a number of existing tools from previous project partners were browsed. Eight prominent learning analytics tools were then selected to be used for the evaluation of the EFLA-1:

- Blackboard Learn 9.1 Retention Centre[3]
- CourseSignals[4] (Arnold and Pistilli, 2012; Arnold, 2010)
- EnquiryBlogger[5] (Ferguson et al., 2011; Buckingham Shum et al., 2012a)
- the LeMo project[6] (Elkina et al., 2013)
- SNAPP[7] (Baron and Jayaprakash, 2014)
- StepUp! (Santos et al., 2012, 2013)
- Student Activity Meter (Govaerts et al., 2011, 2012)
- Student Explorer (Aguilar et al., 2014; Lonn and Teasley, 2014)

The study was conducted with members from the LACE project[8] consortium and its associated partners. Each of the eight participants evaluated two of the eight tools, which in turn meant that each of the eight tools was evaluated twice.

Due to the nature of the study, i.e. the evaluation of the EFLA-1, outcomes dealing with individual tools are not addressed. Instead the focus is entirely on the setup and applicability of the EFLA's dimensions and items.

2.2.2 Quantitative Results

To get an overview of the results for all items Table 2.1 shows how many *yes*, *no* and *not applicable* every item received. The highest scoring instance for *yes*, *no* and *not*

[2] http://forms.google.com
[3] https://help.blackboard.com/en-us/Learn/9.1_2014_04/Instructor/130_Student_Performance
[4] http://www.itap.purdue.edu/learning/tools/signals/ and http://www.itap.purdue.edu/studio/signals/
[5] http://learningemergence.net/tools/enquiryblogger/
[6] http://www.lemo-projekt.de
[7] http://www.snappvis.org
[8] http://www.laceproject.eu

Table 2.1 Presence (yes/no) or non-applicability of items in a tool

		yes	no	not applicable
Objectives	awareness	***15***	-	1
	reflection	12	2	2
	motivation	9	4	3
	behavioural change	14	1	1
Learning Support	perceived usefulness	14	-	2
	recommendation	8	6	2
	activity classification	6	8	2
	det. of students at risk	12	3	1
Learning Measures and Output	comparability	12	1	3
	effectiveness	9	1	6
	efficiency	4	4	***8***
	helpfulness	14	-	2
Data Aspects	transparency	9	5	2
	data standards	5	6	5
	data ownership	1	***10***	5
	privacy	9	2	5
Organisational Aspects	availability	7	3	6
	implementation	6	3	7
	training of stakeholders	7	1	***8***
	organisational change	8	5	3

applicable are highlighted. Table 2.2 summarises the rating values of all items and also lists their average rating. The highest average rating is achieved by the item *awareness*, i.e. 4.3, while the lowest average is achieved by *efficiency*, i.e. 2.6. These two items are also the ones with the lowest (*awareness*) and highest (*efficiency*) non-applicability.

The data shows that the items of the first dimension, i.e. *Objectives*, are often present in/supported by the tools analysed. Also, the amount of non-applicability of these items is rather low compared to that of other dimensions. The item of *awareness* has the highest score of *yes*, followed closely by that of *behavioural change*. Non-applicability of items is quite low in this dimension which in reverse means that they are applicable and thus suitable items when evaluating learning analytics tools. *Motivation* seems to be the most controversial item as it has the most diverse results. Looking a the ratings for the *Objectives* dimension this view is supported as most study participants found it easy or very easy to judge the items of this dimension.

The non-applicability of the items in the dimension *Learning Support* is similarly low as that of the *Objectives* dimension. Although they are applicable, however, they are not present in/supported by the tools as often as the items of the first dimension. Especially *recommendation* and *activity classification* seem not to be as common in

Table 2.2 1(very difficult)-to-5(very easy) scale ratings plus average rating for all items

		1	2	3	4	5	avg.
Objectives	awareness	-	1	1	7	7	***4.3***
	reflection	1	-	2	6	7	4.1
	motivation	1	3	3	4	5	3.6
	behavioural change	-	3	5	5	3	3.5
Learning Support	perceived usefulness	2	-	1	7	6	3.9
	recommendation	1	1	3	4	7	3.9
	activity classification	4	3	3	1	5	3.0
	det. of students at risk	-	1	3	6	6	4.1
Learning Measures and Output	comparability	-	6	2	5	3	3.3
	effectiveness	2	5	4	4	1	2.8
	efficiency	4	3	5	3	1	***2.6***
	helpfulness	2	4	2	5	3	3.2
Data Aspects	transparency	-	4	6	5	1	3.2
	data standards	3	2	2	5	4	3.3
	data ownership	3	3	3	6	1	2.9
	privacy	-	3	2	8	3	3.7
Organisational Aspects	availability	2	1	1	3	9	4.0
	implementation	2	1	2	2	9	3.9
	training of stakeholders	2	-	1	8	5	3.9
	organisational change	2	-	1	12	1	3.6

learning analytics tools. The ratings for the items in the *Learning Support* dimension are not as tendentious as the previous ones. There are still many easy and very easy ratings. However, the number of difficult and very difficult ratings is notably higher. Especially *activity classification* was deemed a difficult to evaluate item by the study participants.

Looking at the ratings for the items in the *Learning Measures and Output* dimension we can see that they are almost evenly spread over the scale. No clear tendency of either difficulty or ease can be identified. Also, the non-applicability of the items is quite a bit higher than that in the first two dimensions. In 50% of the cases *efficiency* was not applicable while *effectiveness* was not applicable in 38% of the analysed cases. All items in this dimension, however, have rather low (or even none) *no* values. It thus seems that items here tend to be either present in/supported by the learning analytics tools or not applicable.

The items with the most *no* values are those of the *Data Aspects* dimension, i.e. they are often not present in/supported by the analysed LA tools. Non-applicability is on a medium level of about a third for this dimension while the *yes* values vary from low to medium levels. Study participants tended to be rather positively confident when rating the items of this dimension. Although there are hardly any very easy ratings, the number of easy ratings is quite high.

The most clear and obvious rating tendency was given to the items of the *Organisational Aspects* dimension. In three quarters of the cases the items have either been rated as easy or very easy to judge by the study participants. The non-applicability of the items is the highest for this dimension while *yes* and *no* values vary.

2.2.3 Qualitative Results

Apart from collecting quantitative feedback about the items, study participants were also offered the opportunity to describe the application of each item and to add comments.

Generally, participants thought that it was rather easy to judge the items of the *Objectives* dimension. The resources they used to evaluate the tool often provided information about whether it supported *awareness*, *reflection*, *motivation* and *behavioural change*. One issue raised by participants was the distinction between a tool intending to foster something and actually being successful in doing so. Based on the fact that in many cases only the actual user of a tool can assess whether *awareness*, *reflection*, *motivation* or *behavioural change* was fostered, they suggest to ask whether a tool intends to do something when evaluating it. Another issue raised was that the main user type of a tool should be identified before evaluating it as some tools might cater to learners, other to teachers, etc. A third issue mentioned by the participants was that of direct or indirect fostering (or better the intention to do so) of the items.

For the items of the *Learning Support* dimension participants also stressed that taking the user type into account when evaluating a tool is important. Participants also mentioned that there are two types of items in this dimension. While *usefulness* can be deemed an intended goal of a tool, the items of *recommendation*, *activity classification* and *detection of students at risk* are features / functionalities of a tool. Although both types of items are valid to be used to evaluate a learning analytics tool, an evaluation framework should benefit from using only one type of item per dimension. It was also noted that for some items it might not suffice to say whether a tool does something or not in order for it to be deemed a good tool, e.g. too many recommendations might be worse than no recommendations. The item that caused most trouble to the study participants is that of *activity classification*. Participants found it rather difficult to judge this item as they did not fully understand what it meant while participants of the GCM study most likely had a clear concept in mind, i.e. that learning analytics tools “know” what their users are doing automatically. It was therefore suggested to rephrase or redefine the item.

The dimension *Learning Measures and Output* was an overall difficult one to judge for the participants. They not only had difficulties judging some of the items but the dimension title increased this difficulty even more. They were unsure whether to relate an item to the measures or the output of a tool, to the processes or the tool itself and thus suggested to define a better, clearer name and concept for this dimension. The item *comparability* was quite difficult for participants to apply as they were not completely sure about what was to be comparable. In the GCM study the learning analytics experts had identified comparability in relation to the measures and outcomes of an analytics tool, e.g. that effects of one tool could be compared to those of another tool. From the responses of this study's participants, however, it is clear that some participants assumed the comparability to be for users within one tool. This misunderstanding clearly needs to be addressed by a better definition of the item and possibly a rephrasing. For the items *effectiveness* and *efficiency* it was suggested to distinguish between the intention of a tool and the fulfilment of that intention. Also, participants would have liked to see clear definitions in order to better distinguish them from another. They also suggested to clearly indicate the type of user of a tool, too, when applying these items. The same applies to the item *helpfulness*. They also suggested to clearly distinguish this item from the one about *perceived usefulness* by giving a clear definition to both.

In the GCM study the learning analytics experts had identified the items dealing with *Data Aspects* as the most important and as the most feasible ones. This time, however, the items of this dimension were often either not supported by a tool or not applicable. The main reason given for either saying *no* or *not applicable* was that they had not used the tool themselves but had to rely on the resources describing the tools. It was thus suggested to add an *I don't know*-option. Here, the item *data ownership* was deemed the most difficult to rate. Some participants were not able to fully grasp and apply the concept to a given tool and therefore suggested a more detailed definition of the item. Again, they would have liked to see the type of user in focus mentioned when doing the evaluation. For the items *transparency* and *privacy* the issue of differentiation and a clearer definition was raised. It was also mentioned that in the case of *transparency*, two types could be present in a tool: a tool supports transparency if users know what data about them is collected and stored but also if one user can see information about other users.

The dimension on *Organisational Aspects* was by far the easiest to rate for the participants. It is also the one with the most *not applicable* values. Many participants reported that this was due to many of the tools being prototypical implementations that had only been used within one course or as a small test bed study. Another reason given was the lack of information provided by the resources used for the evaluation about anything related to *Organisational Aspects* and not being able to use the tool. The difference between the items of *availability* and *implementation* was not clear to a number of participants. They thus suggested to either define the items more clearly or merge them into one.

2.2.4 Discussion

The results of this framework evaluation study allow us to identify several issues with the EFLA-1 that need to be addressed in order to work towards an improved EFLA-2 for the next evaluation cycle. The issues identified can be divided into the following categories: (1) *concept definitions*, (2) *differentiations*, (3) *framework structure*, and

(4) *questionnaire adaption*.

The first category, *concept definitions*, relates to any case where it was expressed that either a dimension or an item needs to be rephrased or defined more clearly in order to be properly applied to a tool evaluation. One dimension and three items where this is the case were particularly mentioned: *Learning Measures and Output*, *activity classification*, *comparability*, and *data ownership*. Renaming, and thus redefining, a whole dimension also influences how the items of that dimension are interpreted. When constructing the EFLA-2, this will have to be taken into account.

The issues of the second category, *differentiations*, are closely related to those of the first category. Participants identified some items, or better pairs of items, that needed to be defined more clearly and supported by some distinct example so as to be able to properly distinguish between them. Otherwise users of the EFLA might misunderstand them and thus distort the results of a tool evaluation. The items mentioned by the study participants are *usefulness* vs. *helpfulness*, *effectiveness* vs. *efficiency*, *transparency* vs. *privacy*, and *availability* vs. *implementation*.

The third category, *framework structure*, deals with the issue of inter-dimensional heterogeneity vs inter-dimensional homogeneity of item types. It was suggested to ensure that the types of items within one dimension are the same in order to improve the applicability of the whole dimension. Generally, items should tend to be concept rather than feature driven. Participants identified this issue in the dimension *Learning Support* but all other dimensions should be inspected as well so as to avoid this issue from appearing again in the next evaluation cycle.

The fourth category, *questionnaire adaption*, comprises issues that need to be addressed when setting up the second version of the EFLA's questionnaire or better the next practical, applicable and executable version of the EFLA. Several aspects were noted that would highly improve the applicability of the items. For many items the answers would differ depending on the user type addressed. This should thus be clarified for each questionnaire and lead to specific instances of the EFLA for different stakeholders. Questions for the items should best ask about the intention of a tool as this is something that can be answered much more easily than a tool's actual impact on a user. This is especially true if the evaluator has no access to the tool but has to work with descriptive resources. The third issue related to *questionnaire adaption* is the possible addition of answer options. Several participants of the evaluation study remarked that they would have liked to see an *I don't know*-option or a *too much*-option as information for some items might be too sparse.

An issue that is not related to any of the categories and that cannot be improved by us is the sparsity of information provided in the resources about learning analytics tools. While addressing the issues mentioned above will make it easier for externals to evaluate a tool, the most complete evaluations will be those of the actual users or creators of a tool. In those cases where users or creators apply the EFLA to their own tool, however, the results might be biased which has to be taken into account as well.

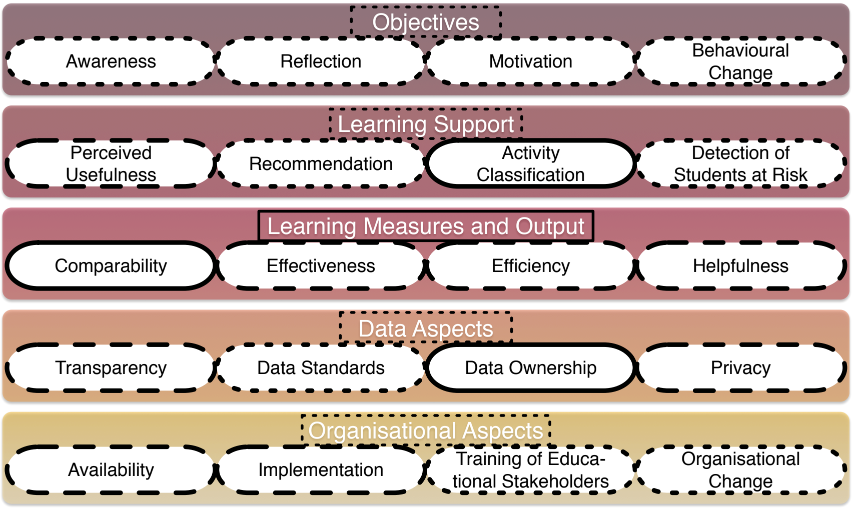


Figure 2.1 Items and dimensions with category 1 issues (solid outline), category 2 issues (dashed outline) and no issues (dotted outline)

2.2.5 Taking Stock

This section built on the findings of a group concept mapping study that empirically identified dimensions and items for an evaluation framework for learning analytics. We conducted an evaluation study with members of the LACE project to apply the EFLA-1 to a number of tools in order to evaluate it. With the feedback from the participants we were able to identify problematic issues and have collected suggestions how to overcome the issues and improve the framework.

Figure 2.1 shows which dimensions and items have been identified with category 1 issues of *concept definitions* (solid), category 2 issues of *differentiations* (dashed) or no issues (dotted). The issues related to category 3 (*framework structure*) and category 4 (*questionnaire adaption*) refer to the EFLA as a whole and are thus not marked. The outcomes of the evaluation study will be carefully analysed and taken into account when creating the EFLA-2. Apart from the theoretical framework set up, the structure of the related evaluation instrument will also be improved as different stakeholders might require different versions of the instrument. The improved EFLA as well as its implementation will then form the basis of another evaluation cycle.

2.3 Framework Construction Study

Simply addressing the re-wording and re-defining issues mentioned by the learning analytics experts in the evaluation study of the EFLA-1 is not enough to create a new version of the framework. According to the experts, in order for the framework to be used for the comparison of learning analytics tools with one another and to also provide a quick and easy way to evaluate tools in a standardised way that the LA community can work with, the following requirements need to be met in the next framework version:

(R1) the framework and its questionnaire need to be more condense;

(R2) dimension titles and item names need to be clear and easy to understand;

(R3) there need to be different questionnaires for the different user types;

(R4) the questionnaire needs to be answered by those that actually use the LA tools;

(R5) users have to be able to relate to the items and to provide information about them;

(R6) the items must be motivated, i.e. concept-driven and not feature driven.

During the development process of the second version of the Evaluation Framework for Learning Analytics we therefore did several things: To condense the framework and reduce the number of dimensions and items we on the one hand looked at the GCM data again as this was based on input from the LA community. On the other hand we also had a look at related literature to find other categorisations and classifications in the field of learning analytics as well as other evaluation frameworks, scorecards or indices in the educational field. Then, once the dimensions were set and a range of possible items named, we again turned to the literature to decide which items to include and how to motivate them. Based on the results of these steps the EFLA-2 was developed.

2.3.1 Results of the GCM Data Review

The first version of the framework is entirely based on the data from the group concept mapping study. During that first analysis we worked with a replay map ranging from 15 to two clusters and finally settled on eight. To the eight-cluster map we then applied the ratings for importance and feasibility as given by the learning analytics experts. During the framework-creation process we combined some clusters while discarding others and finally ended up with five dimensions and 20 indicators. One of the points of criticism in the evaluation of the EFLA-1 is that five sections with a total of 20 indicators, and thus 20 items in a questionnaire, was too cumbersome and could thus be too time consuming to motivate people to use it. Keeping this in mind, we therefore approached the GCM data from the lower side of the replay map, i.e. we only looked at two, three and four clusters in order to narrow down the number of dimensions and items to meet R1. For the items we also looked at the importance ratings of statements in the two-, three- and four-cluster solutions as these are the ones that matter most to the LA community.

The two-cluster map consists of one very large (cluster 1) containing 65 statements and one medium sized cluster (cluster 2) containing 38 statements (see Figure 2.2 left). Looking at the statements in cluster 2 shows that most of them are concerned

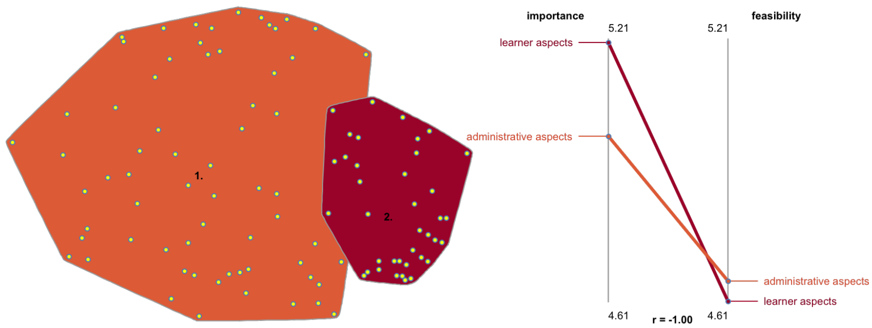


Figure 2.2 Point map and ladder graph for the 2-cluster solution

with issues related to learners. We labelled the cluster *learner aspects*. While cluster 2 is a clearly distinct cluster with statements all about the learner, cluster 1 has no such focus. It is thus hard to come up with a label that covers everything in this cluster. One way to describe it would be to say 'everything not directly related to learners'. The diversity of cluster 1 is reflected in its statements and its spread-out size. Looking at the individual statements in more detail, though, reveals that most of them can either be attributed to teachers or to an organisation. We therefore labelled this cluster *administrative aspects* where the term 'administration' encompasses the teacher perspective as well as the organisational perspective.

The ladder graph of the ratings shows that the statements in cluster 2 are rated more important than those in cluster 1 (see Figure 2.2 right). The ratings for feasibility show the opposite effect, i.e. the statements in cluster 1 are rated more feasible than those in cluster 2. The graph also shows that the average ratings for feasibility are much closer together than those for importance. On the other hand, both clusters in general got higher importance ratings than feasibility ratings. The correlation is given as -1.

In the three-cluster solution the previous cluster 1 *administrative aspects* is split into two clusters (cluster 1 and cluster 2) while the previous cluster 2 *learner aspects* stays the same and is now cluster 3 (see Figure 2.3 left). The new cluster 1 is the smallest of the clusters. It contains 19 statements. Looking at the individual statements in this cluster also shows that most of them deal with the transparency, privacy or accessibility of data. We therefore labelled this cluster *data aspects*. The largest cluster in this point map now is cluster 2. It contains 46 statements. While some of the statements in this cluster are related to institutional issues, general educational aspects or general users of learning analytics tools, most of them are in one way or another about teachers and their issues and behaviours. We therefore labelled the cluster *teacher aspects*. The cluster about learners is now cluster 3 and contains the same statements as before. We thus kept the label *learner aspects*.

The ladder graph (see Figure 2.3 right) clearly shows that the statements in cluster 3 (*learner aspects*) have been rated more important than those in cluster 1 (*data*

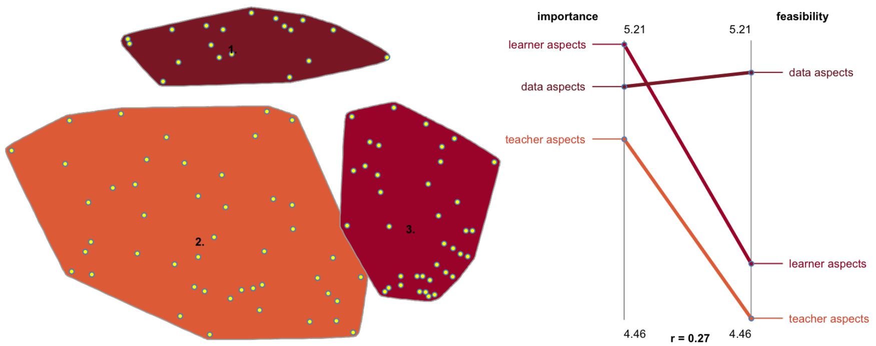


Figure 2.3 Point map and ladder graph for the 3-cluster solution

aspects) which in turn are deemed more important than those in cluster 2 (*teacher aspects*). The feasibility side of the graph shows a different picture. Here the *data aspects* cluster is deemed most feasible, followed by the *learner aspects* cluster and then the *teacher aspects* cluster. The graph also reveals that the importance ratings are closer to one another for the three clusters than the feasibility ratings. The correlation is given as +0.27.

In the 4-cluster solution the previous cluster 1 *data aspects* and the previous cluster 3 *learner aspects* stay the same and are now clusters 1 and 4 respectively (see Figure 2.4 left). As the clusters contain the same statements as before, we kept the labels of those clusters. The previous cluster 2 *teacher aspects* is split into two new clusters (now clusters 2 and 3). The larger of the two new clusters is cluster 2 containing 29 statements. Cluster 3 consists of 17 statements and is thus the smallest of all four. Looking at cluster 2 reveals quite a diversity of topics even after the split. Although a number of statements still deal with issues related to teachers, their proportion is not as high any more in this cluster. Many statements are about users in general or the tool itself and its functions. We therefore labelled this cluster *impact & integration*. In cluster 3, however, most statements still deal with teachers and their work, either implicitly or explicitly which is why we took over the label *teacher aspects* for this cluster.

The ladder graph (see Figure 2.4 right) shows that the *learner aspects* cluster is deemed most important, followed by the *data aspects* cluster and the *impact & integration* cluster. The lowest importance is given to those statements from the *teacher aspects* cluster. Again, looking at the clusters' feasibility shows a different picture: here the *data aspects* get the highest ratings by far. The two clusters *impact & integration* and *teacher aspects* have very low feasibility ratings, while *learner aspects* are rated slightly higher but still low. The large gap in the feasibility ratings is quite prominent. The correlation is given as +0.45.

From the three presented cluster solutions, the one with four clusters provides the best split of the statement landscape as each cluster has a distinct enough theme

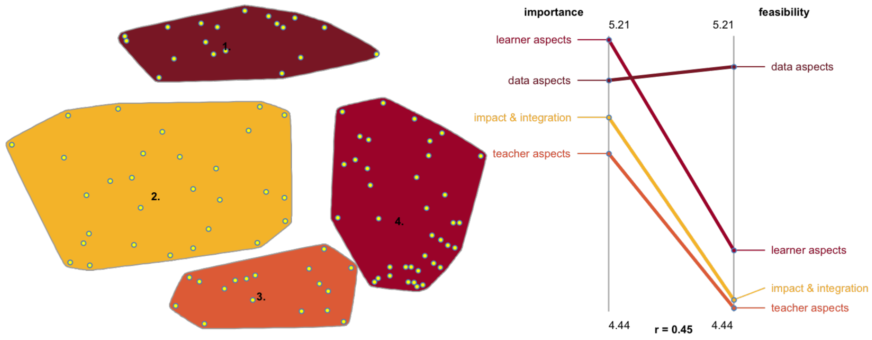


Figure 2.4 Point map and ladder graph for the 4-cluster solution

focus for a dimension and offers a wide enough range of statements to choose from for the items. The following four dimensions will thus be used as the tentative structure of the new framework version while reviewing relevant literature for this study: (1) *data aspects*, (2) *impact & integration*, (3) *teacher aspects*, and (4) *learner aspects*. A list of the four clusters, the statements they contain and the statements' bridging values in given in Appendix D.

2.3.2 Results of the Literature Review

One of the more well-known instruments to evaluate a system (or application or tool) is the System Usability Scale (SUS) (Brooke, 1996). It offers a quick way to measure people's perceived usability of a system by having them rate ten statements. Brooke's work also influenced the creation of ISO 9241, part 11 (ISO9241-11:1998, E), i.e. a set of requirements for office work with visual display terminals with regards to effectiveness, efficiency and satisfaction. Another instrument often used to evaluate a system is the Technology Acceptance Model (TAM) (Davis, 1989). It covers several aspects: (1) perceived ease of use, (2) perceived usefulness, (3) attitude towards using, (4) behavioural intention, and (5) actual system use. Although the SUS as well as the TAM can of course also be used to evaluate learning analytics applications, they do not provide any insights specifically related to learning and teaching processes but to generally using a system only. The EFLA is closing this gap.

Ferguson (2012a) gives a very detailed overview of how the research field of learning analytics came to be. She also gives a differentiation from educational data mining and academic analytics. Four drivers of learning analytics are named: big data, online learning, political concerns and the question of who benefits. Also, four challenges for the development of the field are given: (1) build strong connections with learning sciences, (2) develop methods of working with a wide range of datasets in order to optimise learning environments, (3) focus on perspectives of learners, (4) develop and apply a clear set of ethical guidelines. Several aspects of this categorisation of challenges is interesting with regards to the creation of the EFLA-2: that the use and application of learning analytics should be connected to the learning sciences, i.e. that it needs to be motivated, also applies to the evaluation of learning analytics, i.e. the items of the EFLA-2 questionnaire need to be motivated. The focus on the learner as well as on ethical guidelines are both already reflected by our dimensions *data aspects* and *learner aspects*.

In their generic framework for learning analytics Greller and Drachsler (2012) point out six dimensions that should form the foundations of every learning analytics tool: (1) stakeholders, (2) internal limitations, (3) external limitations, (4) instruments, (5) data, and (6) objectives. The framework is meant to pose as a guide when developing and implementing learning analytics applications. The three dimensions that are most relevant for the creation of the EFLA-2 are the stakeholders, data and the objectives. Similarly, Chatti et al. (2012) created a reference model for learning analytics that consists of four dimensions: (1) what: data environments, (2) who: stakeholders, (3) why: objectives, and (4) how: techniques. The EFLA-2 can build on all those dimensions and evaluate the outcomes. Once a learning analytics application is in place, Verbert et al. (2013) distinguish four stages in the usage of such applications: (1) awareness, (2) reflection, (3) sensemaking, and (4) impact. This is to say that collected data can be visualised to make users aware of and reflect about something, to then gain new insights and to finally react to these insights. All of those aspects are important for the creation of the EFLA-2 as they are what needs to be evaluated.

Papamitsiou and Economides (2014) conducted a literature review of empirical evidence in learning analytics and educational data mining. Of the 209 articles that passed their initial inclusion criteria, 40 were considered as 'key studies' and classified according to several categories. The one of most interest to our purpose of improving the evaluation framework for learning analytics is their classification according to the research objectives: (1) student / student behaviour modelling, (2) prediction of performance, (3) increase (self-) reflection / (self-) awareness, (4) prediction of dropout & retention, (5) improve assessment & feedback services, and (6) recommendation of resources. That there should be a focus on the learners and their behaviour is, again, reflected in our dimension *learner aspects*. Awareness and reflection are two aspects that appear in both clusters related to the LA tool users, i.e. *learner aspects* and *teacher aspects*, and are thus of high importance for the EFLA-2. The prediction of dropout is also something that is of interest for both of those stakeholder groups and that should be kept in mind when deciding on the items of the EFLA-2.

Based on the paper by Papamitsiou and Economides as well as the first version of the EFLA, the LACE project settled on four so called propositions for their Evidence Hub (Clow et al., 2014): (1) LA improve learning outcomes, (2) LA improve learning support and teaching, including retention, completion and progression, (3) LA are taken up and used widely, including deployment at scale, and (4) LA are used in an ethical way. The interesting aspect here is that both learning and teaching are being focused on. Learning analytics tools can thus be useful for learners and for teachers and both points of view need to be taken into account when evaluating a tool as it is reflected by our dimensions *teacher aspects* and *learner aspects*. Ethical usage of learning analytics is also stressed again which is reflected by our dimension *data aspects*.

In the related field of technology enhanced learning, Law and Wild (2015) created an evaluation framework for personal learning environments based on the TOPS model. While there are four perspectives in the model, the framework in the end consists of three dimensions: (1) technological, (2) organisational/social, and (3) psycho-pedagogical. Each dimension groups several constructs which each have their own method of evaluation (e.g. questionnaire, focus group, user test, observation, etc.). Their framework is thus not a concise and concrete tool, but a collection of methods to be used iteratively while developing and implementing a personal learning environment. The interesting aspect for the EFLA-2 here is the iterative usage of the framework while developing and improving learning analytics applications.

Shelton (2010) conducted a six round Delphi study to create a quality scorecard for the administration of online education programs. Basing the study on existing quality indicators, the scorecard in the end consisted of nine categories containing a total of 70 quality indicators: (1) institutional support, (2) technology support, (3) course development and instructional design, (4) course structure, (5) teaching and learning, (6) social and student engagement, (7) faculty support, (8) student support, and (9) evaluation and assessment. For the scoring method it was decided that each indicator can reach up to 3 possible points (0-not observed, 1-insufficient, 2-moderate use, 3-completely meets criteria). Shelton also suggests a six-point scale to then categorise the number of achieved points: perfect, exemplary, acceptable, marginal, inadequate, unacceptable. An updated version was published in 2015 (Shelton and Pedersen, 2015). The scorecard can be bought for use online. In relation to the EFLA, the interesting aspect here is the similar set-up of basing the framework on the input of experts from the field (whether it be via a delphi study or a group concept mapping study). Also, the idea of using a questionnaire-based scorecard as an applicable tool reflects our own ideas for the EFLA. Shelton's scorecard, however, is rather long as it consists of 70 quality indicators. With the EFLA we aim at creating an easy-to-use way to evaluate learning analytics tools that does not take much time from the involved stakeholders to answer.

One goal of the LinkedUp project (Drachsler et al., 2013, 2014) was to create an evaluation framework for linked and open data tools. The project set up a competition of three cycles, each time developing the framework further. The important lessons learned from their evaluation framework creation process are: (1) start with definition of evaluation dimensions (and indicators), (2) test their understandability, (3) do not use a 'not applicable'-option, (4) less (indicators) are more (preferable), (5) unification of the scale, (6) weighting of specific dimensions/indicators. As we did for the EFLA-1, the LinkedUp project also used a group concept mapping study to gather input from field experts. The important aspects from the LinkedUp experience that impact the EFLA-2 are the suggestion to not use a 'not applicable'-option, to make sure that the framework and its questionnaire are as short as possible and to use the same scale for each question.

In 2012, Bichsel (2012) from the Educause Center for Applied Research conducted an analytics survey that resulted in the development of the analytics maturity index. The index consists of six dimensions: (1) data/reporting/tools, (2) governance/infrastructure, (3) investment, (4) expertise, (5) culture, and (6) process. The tool is not geared towards students and only indirectly at teaching staff but focuses on an institution's administration. The tool measures how mature an institution's current implementation of analytics is. The major difference to the EFLA here is the intended user of the index. The EFLA is clearly geared towards the direct users of learning analytics tools. The maturity index, however, evaluates how mature the overall integration of learning analytics is on an institutional level and not the learning analytics tool itself.

With the Learning Analytics Readiness Instrument (LARI) Arnold et al. (2014a) address the institutions' need for reflection on how ready they are to implement learning analytics solutions. The tool can be used 'in the formative stages of planning'. The work is based on the above mentioned ECAR analytics maturity index. The authors list five learning analytics readiness factors: (1) ability, (2) data, (3) culture & process, (4) governance & infrastructure, and (5) overall readiness perception. LARI, like the ECAR maturity index, is meant to be used by the administrative level of an institution. It can be situated a step before the application of the EFLA, i.e. as a precursor.

In 2016 the Australian Government Office for Learning and Teaching published a report on the current status of learning analytics within higher education in Australia (Colvin et al., 2016). The report presents two studies: the first one consists of interviews with senior leaders in higher education about learning analytics implementations at their institutions while the second study investigates the factors that make learning analytics uptake sustainable via a group concept mapping approach. From the interviews the authors were able to mark four variable categories: (1) concept, (2) readiness, (3) implementation, and (4) context. They also identified two types of institutions: type 1 mainly uses LA for retention purposes while type 2 uses LA to inform student learning (of which retention is seen to be a part). The group concept mapping study resulted in seven dimensions: (1) strategy: whole-of-organisation view, (2) compatibility with existing values/practices/systems, (3) data platform: standards and governance, (4) data use: accessible, transparent, valid/reliable, (5) actionable tools with an evidential base, (6) conditions for educator adoption, and (7) supporting student empowerment. Again, it is interesting to see that the authors of this publication relied on the input from their research field by using a group concept mapping study to develop their framework. While this framework, however, deals with the advancement and uptake of learning analytics, some of the different dimensions are quite closely related to those that we have so far settled on for the EFLA, e.g. *data aspects*, *teacher aspects* and *learner aspects*.

Based on the literature review, the following insights are taken into account for the development of the EFLA-2: data, impact, teachers and learners are all very important aspects in the field of learning analytics and are thus good dimensions for an evaluation framework. Also, with regards to its content as well as its future users, the framework will focus on learners and teachers and their benefits (e.g. awareness and reflection) in learning analytics and not on the administrative or organisational level. Finally, the same scale will be used for all items (without using a 'not applicable'-option) and the number of items will be as reduced as possible.

2.3.3 Discussion of Dimensions and Motivation of Items

Based on the input from the GCM review as well as the literature review, the first rough outline of the EFLA-2 so far consists of the four dimensions *data aspects*, *impact & integration*, *teacher aspects* and *learner aspects*. To make the dimension titles even easier to grasp, they are shortened to *Data*, *Impact*, *Teachers* and *Learners*. The aim is to have about three to four items per dimension to keep the framework neat and to the point and to use the same Likert scale for each item in the questionnaire. To decide on the items' focus, several sources are taken into account: the requirements for the EFLA-2 resulting from the evaluation of the EFLA-1, the statements from the GCM with an above average rating in each dimension and the insights from the literature review. Based on this, the dimensions and the items can then be motivated and grounded by a further look into the literature related to those focus topics.

Considering the requirement of having different questionnaires for the different user types and taking into account our focus on learners and teachers as the important stakeholders, two sections of the EFLA-2 need to be developed in parallel, one for each stakeholder group. For the structure of the framework, this means that the learner section will consist of the dimensions *Data*, *Impact* and *Learner* and the teacher section will consist of the dimensions *Data*, *Impact* and *Teachers*. To ensure consistency of the evaluated aspects across user types, the items for the two sections will be kept as alike as possible and will only differ to accommodate the users' different angles on the same matter. The go-zone graphs for the four clusters can be found in Appendix E.

Looking at all statements with an importance rating above average for the *Data* cluster reveals the following main topics: transparency, control, accessibility and privacy. In the *Impact* cluster there are more statements than in the *Data* cluster. It is thus a little bit more difficult to narrow down the topics but the main ones are: training for LA tool usage, motivation/engagement, adaptation, and effectiveness. As the *Teachers* cluster is quite small, there are not so many statements in the rated-important-above-average group. The overarching main themes, however, are reflection, taking actions and behavioural change. Implicitly included in this topic triple is awareness as this is what fosters reflection which in turn can foster change. For the *Learners* cluster there are more statements to choose from again. The main topics are: effect on outcomes, motivation/engagement, and at-risk detection, awareness and change (which implies reflection).

The *Data* Dimension

Taking all previous discoveries into account, the first dimension, *Data*, needs to be supported by items that reflect the ethical and transparent handling of data within a learning analytics tool. For Slade and Prinsloo (2013) learners are to be seen as the primary beneficiaries of learning analytics and they suggest an ethical framework for higher education institutions to consist of six principles: (1) learning analytics as moral practice, (2) students as agents, (3) student identity and performance are temporal dynamic constructs, (4) student success is a complex and multidimensional phenomenon, (5) transparency, and (6) higher education cannot afford to not use data.

Pardo and Siemens (2014) analyse the privacy and ethical issues that are specific to the learning analytics context. The principles they present are gathered from the review of LA proposals, government frameworks and regulatory directives and allow educational institutions to assess their current level of compliance in order to then possibly improve their privacy-related matters. The principles are: (1) transparency, (2) student control over data, (3) right of access / security, and (4) accountability and assessment.

The LACE project has organised a series of workshops about ethics and privacy for learning analytics (EP4LA). In their editorial for a special issue on the topic Ferguson et al. (2016b) provide a list of learning analytics challenges with ethical dimensions. The challenges that are of most interest for the creation of items for the EFLA-2 taking into account previously mentioned references are: (1) use data to benefit learner, (2) ensure results are comprehendible to end users, (3) ensure that data collection, usage, and involvement of third parties are transparent, and (4) consider how, and to whom, data will be accessible.

Further resources that deal with ethics and privacy in the field of learning analytics and stress their importance are JISC's code of practice for learning analytics (Sclater and Bailey, 2015; Sclater, 2016), SURF's report about learning analytics under the Dutch data protection act (Engelfriet et al., 2015) and the DELICATE checklist by Drachsler and Greller (2016).

Taking all of these aspects of ethics and privacy into account, the main theme for the *Data* dimension is transparency on three levels: for a learning analytics tool to be evaluated positively, it needs to be clear (i.e. transparent) what data is being collected for the tool, access to that data needs to be available (i.e. transparent), and the tool's output needs to be presented in an understandable (i.e. transparent) way.

The *Impact* Dimension

Based on the previous discoveries, the second dimension, *Impact*, needs to be supported by items that illustrate the effect, i.e. the impact, of a learning analytics tool for its users.

Studies in the related domain of technology-enhanced learning reveal several aspects that can be used for outcome measurements of recommender systems but also for the analysis of other educational technologies and LA tools (Drachsler et al., 2009). The measurement category that is of most interest to the creation of the EFLA-2 is the second one about educational aspects that involves the parameters of effectiveness, efficiency, satisfaction and drop-out rate. Related to these categories, the usability of a product, and thus the success of a product, is highest when a user can use it to achieve set goals with effectiveness, efficiency and satisfaction (ISO9241-11:1998, E). Touré-Tillery and Fishbach (2014) list several behavioural measures when it comes to outcome-focused motivation: (1) higher speed on goal-related tasks, (2) higher speed when moving from one goal-related task to the next, (3) higher accuracy, (4) higher amount of work done and (5) higher level of achievement. Effectiveness and efficiency are thus two aspects that are of high importance when it comes to identifying the impact a learning analytics tool has on learners and teachers. In their review of learning analytics dashboards, Verbert et al. (2014) list effectiveness, efficiency, usefulness and usability as the factors of dashboards that have been evaluated. For the issues covered by effectiveness measurement they name better engagement, higher grades/better results, lower retention rates and improved self-assessment.

Identifying at-risk students, predicting success or drop-out, constructing early warning systems, all of these activities mainly aim at maintaining or even increasing retention rates. On the macro and especially the meso level, i.e. the regional/national/international and the institutional levels respectively, reducing dropout rates and increasing retention are usually some of the main reasons for educational institutions to implement learning analytics instruments (Greller and Drachsler, 2012; Colvin et al., 2016). The interest in these topics has been steady over the years even though the aspect of individual student and teacher support has become recognised as well (see e.g. especially the proceedings of the LAK conferences in 2012, 2015 and 2016 (Buckingham Shum et al., 2012b; Blikstein et al., 2015; Dawson et al., 2016)). On the micro level, i.e. the individual user's level, learning analytics is meant to make learning and teaching processes more effective and efficient (Arroway et al., 2016), however, little evidence has been gathered so far as to what impact learning analytics can and does have on these processes (Ferguson et al., 2016a).

Taking these aspects of the impact of learning analytics on the different users into account, three themes can be distilled: for a learning analytics tool to be evaluated positively, it needs to be able to detect whether a student is falling behind, it needs to make learning more efficient and it needs to make learning more effective.

The *Teacher* and *Learner* Dimensions

Considering the previous discoveries, the dimensions *Teachers* and *Learners* both need to be supported by items that represent those aspects of a learning analytics tool that support and improve learning and teaching processes. For an improvement to happen, users need to change or adapt their behaviour which presupposes reflection on previous behaviour which in turn presupposes awareness of previous behaviour which in turn presupposes some form of trigger, e.g. the provision of feedback (Hattie and Timperley, 2007; Mory, 2004) via a learning analytics tool. According to Butler and Winne (1995) feedback can serve five functions: (1) it can confirm previous beliefs, (2) it can add information, (3) it can replace or overwrite earlier beliefs, (4) it can fine-tune beliefs, and (5) it can restructure the current domain scheme.

By providing feedback and then possibly triggering a cycle of awareness, reflection, behavioural change, further action and in turn feedback, learning analytics applications are thus an important tool to support self-regulation processes (for both learners and teachers). There are several definitions of what self-regulation, and self-regulated learning in particular, entails. Puustinen and Pulkkinen (2001) provide a review and comparison of the most commonly referred to models of self-regulated learning, i.e. those by Boekaerts and Niemivirta (2000), Borkowski et al. (2000), Pintrich (2000), Winne and Hadwin (1998), and Zimmerman (2000). Despite their differences in background, definition, components and type of research, one aspect that all models have in common according to Puustinen and Pulkkinen is their cyclic three-phase nature. All models consist of a preparatory phase, a performance phase and an appraisal phase. The latter phase is where learning analytics applications can play a role by providing feedback about the users' performance (read: actions) and triggering some form of re-action to start preparing the next action and therefore starting the cycle again.

Two aspects that a learning analytics tool thus affects directly are awareness and reflection, while behavioural change is affected indirectly (see also the learning analytics process model by Verbert et al. (2013)). As this applies to teachers as well as to learners and due to their role as an immediate connection between user and tool, we conclude to use *Awareness* and *Reflection* as dimension titles for the frameworks of both stakeholders.

The dimension *Awareness* thus needs to be supported by items that expose the different steps of becoming aware of one's status. In a series of publications that spans decades Endsley (1995) has developed a theoretical model of situation awareness that she later reviewed critically (Endsley, 2000). According to Endsley situation awareness consists of three hierarchical phases: (1) the perception of the elements in the environment, (2) the comprehension of the current situation, and (3) the projection of a future status. She also states, that in dynamic systems the process of becoming aware of one's situation has to be seen as a separate and different process than the ones of making decisions and actually performing actions. This supports our decision to include *Awareness* and *Reflection* as separate dimensions in the Evaluation Framework for Learning Analytics.

Once aware of their situation, people can reflect about it and deduce possible next steps and thus engage in a process of continuous learning (Schön, 1983). Kember et al. (2000) have developed a questionnaire to measure the level of reflective thinking within a course and the items in their Reflection dimension are of high relevance when it comes to establishing the *Reflection* dimension of the EFLA-2: (1) I sometimes question the way others do something and try to think of a better way, (2) I like to think over what I have been doing and consider alternative ways of doing it, (3) I often reflect on my actions to see whether I could have improved on what I did, and (4) I often re-appraise my experience so I can learn from it and improve for my next performance. With the help of reflection, insight about until then unnoticed issues can thus be promoted and a change in learning or teaching behaviour can be achieved (Bolton, 2010).

Taking the above mentioned aspects of awareness into account, the three hierarchical phases from Endsley's situation awareness model can be used to inform the items of the *Awareness* dimension: a learning analytics tool can be evaluated positively if it makes users aware of the current situation, if users comprehend the situation and if they can project a future situation based on their current experience. Taking the aspects about reflection into account, the three main themes that can be distilled for the *Reflection* dimension and that should be covered by a learning analytics tool for it to be evaluated positively are: it makes the users reflect about their activities, it makes them reflect about alternative activities and by this supports behavioural change.

2.3.4 Outline of the Framework

The dimensions as well as the items had to undergo changes in order to create an improved framework. From the results of the review of the group concept mapping study in combination with the results from the review of related literature, four dimensions are identified for the new framework. They are ordered to reflect the procedural steps of a learning analytics application: *Data*, *Awareness*, *Reflection*, and *Impact*. Additionally, the framework is split into two parallel parts: one for students and one for teachers. For each dimension, several core themes are identified and all items are motivated and theoretically grounded. Next to the one-word dimension titles, items are now formulated as statements. All items are available for both stakeholder sections, i.e. they are adapted to the students' or teachers' points of view.

The *Data* dimension now consists of three items: (1) knowing what data is being collected, (2) having access to one's own / one's students' data, and (3) understanding the presented results. The three items of the *Awareness* dimension are (1) being aware of one's own / one's students' current learning status, (2) comprehending one's own / one's students' current learning status, and (3) being able to project one's own / one's students' future learning status. The *Reflection* dimension also consists of three items. They deal with (1) reflecting on learning/teaching activities, (2) reflecting on alternative learning/teaching activities, and (3) knowing when to change one's learning/teaching behaviour. Finally, the *Impact* dimension concludes the framework with three items about (1) detecting of one's own / one's students' falling behind, (2) being able to study more efficiently, and (3) being able to study

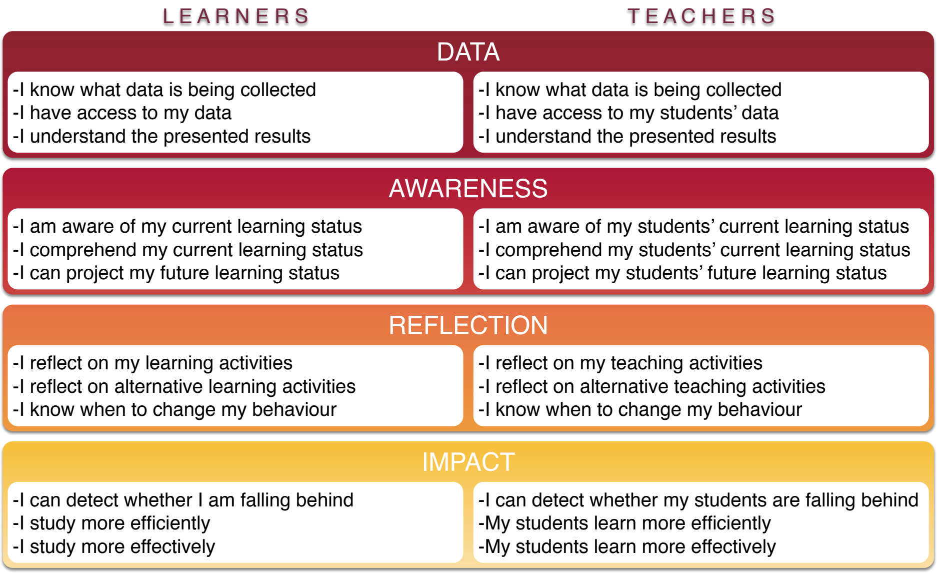


Figure 2.5 Second version of the evaluation framework for learning analytics (EFLA-2)

more effectively. Figure 2.5 shows the second version of the evaluation framework for learning analytics (EFLA-2), i.e. four dimensions with three items each.

With regards to the requirements set by the evaluation study, the second version of the evaluation framework for learning analytics meets them all. There are now less dimensions and less items (requirement 1); the dimension and item names are clear and easy to understand (requirement 2); there are two sections of the framework, one for students and one for teachers (requirement 3); the framework questionnaire is meant to be answered by those two stakeholder groups (requirement 4) that are actively using the tools (requirement 5); and finally, all items are motivated, i.e. concept-driven, and theoretically-grounded (requirement 6).

2.4 Conclusion

This chapter presented the evaluation of the first and the creation of the second version of the evaluation framework for learning analytics. In a next step, the second evaluation framework for learning analytics will be turned into a concrete evaluation instrument as well. The questionnaire will be distributed among students and tutors of a collaborative learning online course. A learning analytics widget about the student's actions in the course environment has been developed specifically for this course. By letting the students and tutors of the course use the EFLA-2 to evaluate the widget, we are able to do two things: we can evaluate our widget (which we will do at two points in time) and we can evaluate the EFLA-2 based on the questionnaire's quantitative results as well as qualitative results directly gathered from students and tutors.

Part II

New Coat Of Paint – Input from the users

Chapter 3

Picture In A Frame – Or: Widget, widget on the wall, am I performing well at all?

In order to explore the usage of the EFLA when it comes to evaluating learning analytics applications, we wanted to apply it to an application we created ourselves. We thus designed a learning analytics widget to support collaborative learning processes. Before evaluating the widget with the EFLA, however, we first investigated the predictive power of several indicators of the activity widget towards the students' grades by instantiating these indicators with data from previous runs of a collaborative online learning course. That is, although the activity widget had not been put in place yet, in this chapter we analyse the log data from these previous years to explore what the widget indicator scores would have been if the widget had been used in those years. This allowed us to better prepare the implementation of the widget into the course's environment which is presented in the next chapter.

This chapter is based on:

Scheffel, M., Drachsler, H., de Kraker, J., Kreijns, K., Slootmaker, A., and Specht, M. (2017). Widget, widget on the wall, am I performing well at all? *IEEE Transactions on Learning Technologies, 10*(1):42–52.

3.1 Introduction

Already from the early days of online education and e-learning, collaborative learning has been one of the prominent pedagogical approaches. Synchronous and asynchronous communication technologies are employed to enable collaborative learning in small, virtual teams of students. However, mediating all communication, coordination and collaboration through online tools appears to result in suboptimal support of, in particular, the social interaction and the group dynamics among team members (Kreijns et al., 2003). This can lower feelings of social presence (Kreijns et al., 2014) and can hamper cognitive processes. One solution is to provide group awareness to students as this might alleviate the problems encountered (Kirschner et al., 2015), i.e. to provide explicit information on the activity of group members and to stimulate awareness, reflection and social interaction. Very often, this information is based on data collected via questionnaires or similar forms filled in by the group members themselves (Phielix et al., 2011) which can be time consuming, tedious and disruptive. This process, however, can be automated by including learning analytics based on interaction data automatically collected within the learning environment. While measurements based on behavioural data are not a one-to-one replacement for measurements based on subjective experience, i.e. proximal variables have indeed more predictive power than distal variables (Fishbein and Ajzen, 2010), learning analytics based on activity data can be used as an additional indication towards group activities that is non-disruptive and covers the whole student population of a course.

Learning analytics (LA) is “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” as defined in the call for papers of the first international conference on learning analytics and knowledge (LAK) 2011[1] and subsequently taken up by (Long and Siemens, 2011). The field has been growing steadily over the last few years as can be seen by the rising numbers of publications as well as events dedicated to learning analytics (Gašević et al., 2015a).

While the term learning analytics may evoke an impression of a field mainly geared towards computing and analysing the collected data to improve outcome, it is indeed about more than that, i.e. a holistic view on the different processes involved in the support and improvement of learning and teaching (Gašević et al., 2015b). The generic framework for learning analytics (Greller and Drachsler, 2012) also shows that the variety of issues in this field is quite diverse, i.e. it covers aspects from stakeholders, objectives, data and technologies to competences and constraints. It is thus important to not simply reduce learning analytics to plain ‘number crunching’ on an institutional level but to purposefully support the immediately involved stakeholders, i.e. teachers and learners. As Ferguson (2014) explains, learning analytics offers “ways for learners to improve and develop while a course is in progress. These analytics do not focus on things that are easy to measure. Instead, they support the

[1] `https://tekri.athabascau.ca/analytics/call-papers`

development of crucial skills: reflection, collaboration, linking ideas and writing clearly" (para. 7).

A learning analytics widget can provide feedback by visualising the learners' activities within a learning environment and can thus support awareness and reflection processes. It allows learners as well as teachers to see the learners' current situation and to adapt their behaviour, e.g. learners could decide to participate more while teachers could decide to get in touch with a specific student. Being able to not only project an immediate future status but to also relate the visualised information to a learner's overall outcome of the course could increase the usefulness of such a widget especially with regards to self-regulation as well as collaborative learning.

3.1.1 Related Work

This section reviews related research about the purpose and impact of learning analytics widgets and dashboards as well as research about the predictive power of students' behaviour during a course. The literature presented can roughly be divided into two sections: the theoretical perspective and the practical perspective.

On the theoretical side there are the two crucial aspects of 'awareness' and 'reflection' that need to be taken into account when dealing with learning analytics dashboards and widgets. The reflection on presented analytics results is not possible without awareness which in turn depends on some form of feedback to the user (Hattie and Timperley, 2007; Mory, 2004). According to Endsley (1995, 2000) being aware of one's own situation is a three level process and a prerequisite for making decisions and effectively performing tasks: the perception of elements in the current situation is followed by the comprehension of the current situation which then leads to the projection of a future status. Once a learner is aware of his situation, he "reflects on the phenomenon before him, and on the prior understandings which have been implicit in his behaviour" (Schön, 1983, p. 68) to then engage in a process of continuous learning. Reflection can promote insight about something that previously went unnoticed (Bolton, 2010) and lead to a change in learning or teaching behaviour. Verbert et al. (2013) emphasise the importance of these aspects in their process model for learning analytics applications: it consists of the four stages awareness, reflection, sensemaking, and impact.

Awareness, however, is not the only aspect that influences the process of feedback, reflection and behavioural change, i.e. of self-efficacy and self-directed learning (Butler and Winne, 1995). Winne (1995) describes self-regulated learning as "principally comprised of knowledge, beliefs, and learned skills, [...] malleable in response to environmental influences" (p. 186) and as something that learners inherently do. Zimmerman (1995) adds to this that self-regulated learning is indeed about more than knowledge and skills and that personal influences such as emotions, one's behaviour and one's social environment play an important role. Learners thus have different ways to construct knowledge on the basis of the information given to them when learning in a self-regulated way (Winne, 2006) and can act and react in different ways.

On the practical side there have been various studies about the positive or negative effect of different behaviour during a course on study outcomes. For face-to-face classes in college, for example, Credé et al. (2010) have shown in a meta-analytic review that there is a correlation between class attendance and class grades and that class attendance is a better predictor than other known predictors of performance. Bennett and Yalams (2013), too, report that attendance and participation in class are positively and significantly related with performance, with attendance achieving better results than participation. Whether class attendance can be an indicative predictor for student performance was also tested and confirmed in a study by Stewart et al. (2011). In an undergraduate statistics course Latif and Miles (2013) explored the impact of class attendance on course outcomes as well and found that the impact was a significant and positive one after controlling for factors related to ability and effort. Louis et al. (2016) also conducted studies to investigate whether class attendance in face-to-face classes is significantly and positively related to the students' performance and found that in undergraduate psychology courses this was indeed the case. Thus, being present in a course can be seen as an important predictor for study success.

What has been confirmed in face-to-face classes has also been observed in online and distance education as shown by Korkofingas and Macri (2013). The researchers revealed that the more time students spent online and are 'present' on the course's website the better their assessed performance was. Macfadyen and Dawson (2010) on the other hand found that time online only weakly correlated with course outcomes while the contribution to discussion forums received significant results. While the recent findings of Strang (2016) suggest that course logins are significant in predicting student online learning outcomes, Tempelaar et al. (2015) on the other hand investigated the predictive power of learning dispositions, formative assessment results and log data, and showed that computer-supported formative assessment during a course was a better predictor than the collected LMS data. The effects of different types of behaviour and activities in online classes are thus still under discussion and are most likely strongly context-dependent.

As part of this discussion about effectiveness and contextuality there are some recent studies that try to go further and investigate the impact of learning analytics dashboards on aspects such as motivation and individual goal attainment of learners. Lonn et al. (2015) investigated the effect of a learning analytics dashboard on the motivation of students that are in danger of failing. Their findings show that student goal perceptions and formative performance results need to be carefully considered in the application of learning analytics dashboards as the results can significantly affect the interpretation of the students' own academic chances. Beheshitha et al. (2016) also focused on investigating the effect of visualisations on different factors of learning. They showed that depending on the data used for the visualisations, positive as well as negative results can be found for the impact of visualisations on the learning progress and suggest a structured methodology for those types of studies. Khan and Pardo (2016), too, identified the need to present different kinds of dashboards and widgets depending on the information needs of the learners as well as the learning activity to make them effective.

All three studies thus emphasise the need to carefully embed dashboards into instructional designs and to try to take the learners' personal preferences into account. A good learning analytics system therefore seems to need either good moderation or different analytics visualisations depending on the learners' different goals and performances to increase their motivation.

3.1.2 Our Approach

Taking all this into account, we have designed a widget based on learning analytics within the learning environment of the European Virtual Seminar on Sustainable Development (EVS), a joint course of about ten European universities that is coordinated by the Open University of the Netherlands. The widget provides several types of feedback based on data automatically collected in the EVS platform, visualised in radar charts and bar charts. Its aim is to make students aware of their own platform activity relative to that of the group and of differences in activity between the group members. The widget also aims at fostering reflection about how their behaviour influences their future status, i.e. in relation to their position within the group and in relation to their course outcome.

To achieve these goals, however, and before offering the learning analytics widget in a live run of the course, we report in this article the results of a formative data study measuring whether the widget indicators validly reflect the individual students' grades given by the tutors. That is, the purpose of this study is to find out whether and if so how the different widget indicators relate to the grades given by the tutors. Thus, before deploying the widget in a live run of the course, we tested whether the information visualised in the widget is indeed valid and reliable in terms of outcome reflection and how it can be interpreted. We therefore wanted to know: *How do the widget indicators correlate with the tutor gradings and can they validly reflect them?* To answer this question, we computed the widget indicator scores with data from four previous runs of EVS and analysed how the tutor gradings of the individual students in those years correlated with the scores generated for the widget indicators with the aim to establish the reflective, i.e. predictive, validity of the widget indicator scores for the students' grades. The analysis was done for the whole run of the course as well as for individual months in order to obtain results for different levels of granularity and for different points in time.

We analysed the data with the following research questions in mind:

(RQ1) Do the widget indicator scores correlate with the tutor gradings of individual students?

(RQ2) Are the scores of some widget indicators better predictors for the students' individual grades than others?

(RQ3) Do certain points in time produce indicator scores that are better grade predictors than others?

Based on these questions the following hypotheses were thus tested in the experiment:

(H1) There is a significant positive correlation between tutor gradings of individual students and the widget indicator scores.

(H2) The scores of the widget indicator 'presence' are better predictors for the students' individual grades than those of the widget indicators 'initiative' and 'responsiveness'.

(H3) The widget indicator scores produced in the second half of the course are better predictors than those of the first half.

The next section describes the course as well as the widget in more detail and elaborates on our method of a two-step analysis, i.e. correlation analysis to uncover potential relationships between tutor grades and widget indicator scores followed by structural equation modelling to determine the strength of the relationships as well as the fit on the data. After the presentation of the analysis results, the discussion section sets the results in relation to the hypotheses and addresses some limitations of our study. The final section concludes the article.

3.2 Method

3.2.1 Participants and Materials

The EVS Course

The European Virtual Seminar on Sustainable Development (EVS)[2] is a joint, web-based Master-level course offered by a partnership of about ten universities (regular as well as distance) from across Europe. The aim of EVS is to foster competences for sustainable development through collaborative learning in virtual, international, multidisciplinary student teams. Here, we provide a brief description of the characteristics of EVS, relevant to the study presented in this article. An extensive description of EVS is provided in (de Kraker and Cörvers, 2014).

Each year, EVS runs from 1 November till 1 April of the next year. During these five months, students from different countries and disciplines work together in teams of four to seven persons on sustainability issues, such as waste management, nature conservation, and climate adaptation. The students from the regular universities are usually between 20 and 25 years of age while the students from the distance universities are usually between 30 and 50 years old. In each run, there are about nine teams in EVS, working on different topics. Coached by a tutor and guided

[2] `http://www.ou.nl/evs`

Table 3.1 Aspects of the individual grades for students within EVS

	grade	aspects covered by grade
T1	planning & progress	planning a realistic own workload dealing with deadlines and agreements flexibility in making appointments/agreements/planning
T2	contribution to team	ability to change roles and responsibilities dealing with feedback from the group taking initiative, helping the group to progress productivity and quality of contributions
T3	support	being supportive (offering support and help others) encourage the learning of the other members giving feedback / reviewing contributions of others
T4	individual-overall	overall grade (average of the three sub-grades)

by an issue expert, the teams conduct a small-scale research project, mostly using secondary data that can be accessed through the internet.

The final grade of the students consists of a combination of grades for several aspects of the course. 50% of the final grade is based on the grade for the quality of the research report a team produces, assessed by the expert; 20% of the final grade is based on the grade for the quality of the group collaboration process, assessed by the tutor; and 30% of the final grade is based on a grade for the individual student's contribution to this collaboration process, also assessed by the tutor. The individual student's contribution grade, i.e. the 'individual-overall' grade (T4 in Table 3.1) is determined by taking the average of the three sub-grades 'T1 planning & progress', 'T2 contribution to team', and 'T3 support'. Each of them covers a range of aspects in the students' contributions (see Table 3.1).

The grades for the report and for the group collaboration process are strongly correlated, and the more team members have low grades for their individual contribution, the lower the grade for the group collaboration will be (de Kraker and Cörvers, 2014). A high level of participation of individual team members is thus important for a good collaboration process in the team, which in turn translates in high-quality group products. In our experience, a common cause of poor group performance in EVS are large differences in individual contributions between the team members, which often results in gradual demotivation of the more active students or an increasing frequency of open conflicts. Visualisation of the individual students' activity could thus help to detect and openly discuss such differences at an early stage, which may prevent conflicts and have a positive effect on team performance and group atmosphere.

The Elgg-based platform[3] used by EVS since 2011 automatically collects and generates data on student activity, which can be used to feed a learning analytics widget

[3] Elgg is a leading open source social networking engine, see: `https://elgg.org/`

Table 3.2 Calculation of the five widget indicator scores

	widget indicator	calculation of the widget indicator scores
W1	initiative	# of posts (discussion, blog, files, pages)
W2	responsiveness	# of comments to posts (discussion, blog, files, pages)
W3	presence	# of page views (on EVS platform)
W4	connectedness	# of contacts made
W5	productivity	(W1 initiative + W2 responsiveness) / W3 presence

that gives the students visual feedback on their own activity and how this compares to their team members and team average.

The Widget

While this section describes the widget we developed and its indicators and functionalities, it is important to emphasise that for the current study we did not test the widget with real users in a live run, but rather tested the reflective and thus predictive validity of the widget indicators (see Table 3.2) with data gathered in previous course runs. Nevertheless, it is important to present the widget and its functionalities here to provide the reader with the idea behind the developed tool and how it can be applied.

The widget, available for download under GNU GPL version 2 (Slootmaker et al., 2015), can be embedded within an Elgg environment as a plugin to make students aware of and reflect on their activity level within the environment relative to other group members and the group average. The widget contains information about the users' platform activities with two subsections, i.e. the cumulative view and the periodic view.

Platform activity is expressed in five widget indicators: 'W1 initiative', 'W2 responsiveness', 'W3 presence', 'W4 connectedness', and 'W5 productivity'. The widget indicator scores are automatically calculated from activity data recorded by the EVS platform (see Table 3.2). The students' activity is visualised in a radar chart, with five axes for the five widget indicators. When hovering with the mouse over the labels of the axes, the definition of the widget indicator is displayed. When pointing with the cursor at the dots in the chart, the corresponding widget indicator score is displayed.

The 'Cumulative activity' radar chart (see Figure 3.1) presents the widget indicator scores for the whole run of EVS, i.e. from the beginning of the course until the current date. In this and all other charts, orange is used for a user's own scores ('Me'), and blue for the group average ('Group'). The scores in the radar chart are scaled from 0 to 10. For each widget indicator, the group member with the highest activity gets a score of 10 and the scores of the other members are scaled accordingly. The colour coding also applies to the 'My activity' bar chart. The orange bar shows a user's average activity, i.e. average of the widget indicators 'W1 initiative', 'W2

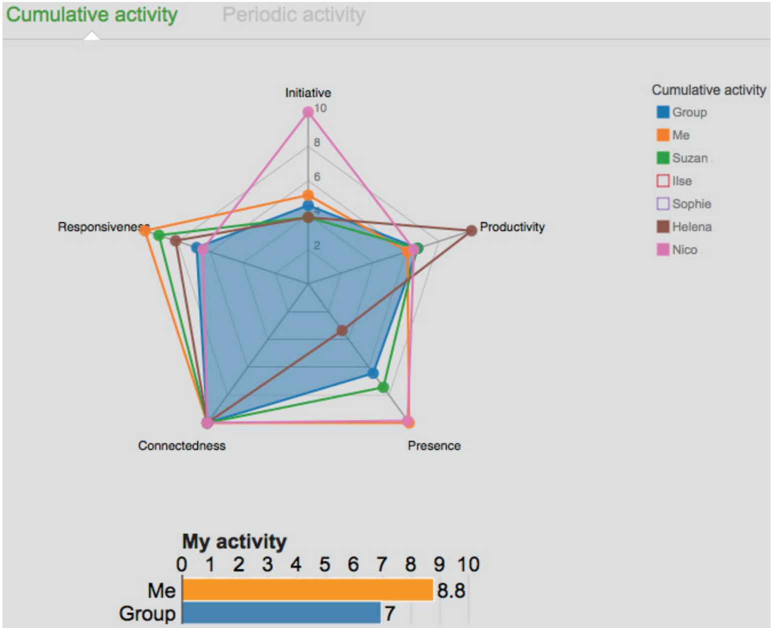


Figure 3.1 Cumulative view of the widget

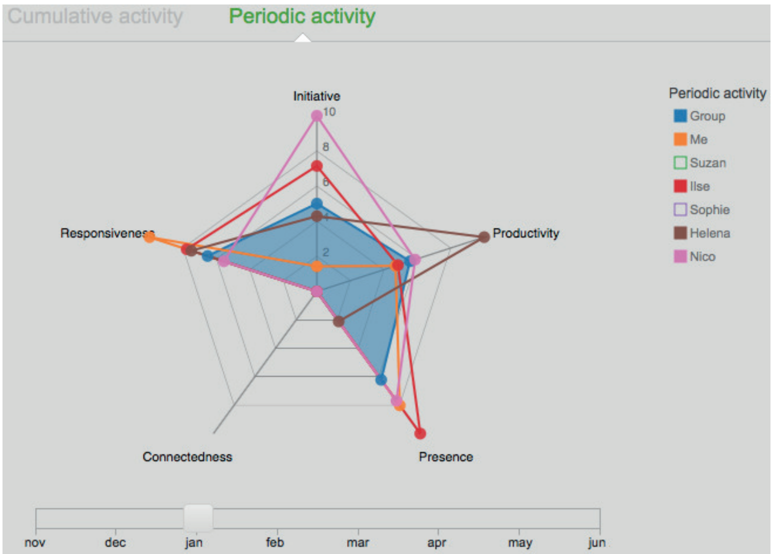


Figure 3.2 Periodic view of the widget

responsiveness', 'W3 presence' and 'W4 connectedness', compared to the average of the entire group (blue bar). The 'Periodic activity' radar chart presents the widget indicator scores per month (see Figure 3.2). Users can choose the specific month with a slider below the chart.

In order to facilitate group performance by enabling co- and self-regulation processes, the widget indicator scores of the individual members of a group are visualised. As explained by Drachsler and Greller (2016) in their article about privacy and ethics in learning analytics, this information can be classified as 'privacy-sensitive' information that needs to be handled according to the DELICATE checklist as not all students of a group might agree to share this 'privacy-sensitive' information within the group. To deal with these privacy and ethical issues, the process suggested by the DELICATE checklist was followed. When the widget will be used in a live run of an EVS course, a widget manual explaining the intentions behind the learning analytics widget, making clear what data is being collected, how it is presented in the widget, and what students can do to protect their privacy will be provided to all EVS users.

Catering to this last point, a Reciprocal Privacy Model (RPM) is implemented into the widget. The RPM enables students to decide how they would like to share their activity data. A target student can only see the individual performance of other students in his team if he also agrees to share his own data with the rest of the team. If a student disagrees with sharing his data, he will only see his own performance in comparison to the group average value in the radar chart of the widget. When he agrees to sharing his own activity data, he will also see the data shared by other members of the team. The RPM model is a very innovative approach that empowers the students to decide with whom and on which level they want to share their data.

3.2.2 Procedure

As explained, data from the previous four runs of EVS were used in order to obtain those years' widget indicator scores for widget indicators 'W1 initiative', 'W2 responsiveness' and 'W3 presence'[4]. The widget indicator scores for 'W4 connectedness' and 'W5 productivity' were not included in the analysis. 'W4 connectedness" was excluded as it turned out that the number of contacts students made (similar to 'friending' in informal social networks) varied strongly and irregularly between EVS runs and teams within the same run. The course manual advised students to make other students contacts, in particular their team members, as this allows them to receive notifications about their platform activities. However, it seems that the number of contacts students in EVS made primarily depended on whether or not the tutor of a group emphasised the need of this feature, rather than the internal motivation of the students to improve communication. 'W5 productivity' was excluded as it represents a combination of three other widget indicators and is thus not an independent variable. Gender, age and nationality of the students were not taken into account in the analysis. Table 3.3 shows the descriptive statistics of the

[4] Unfortunately, the 'W3 presence' scores for the EVS run of 2011-2012 were not available.

three widget indicators for all years pooled with all months combined as well as all individual months.

In a first step, the scores of the three widget indicators (W1, W2, W3) for the four runs were correlated with the students' four individual grades (T1, T2, T3, T4) as given by their tutors. As the data from the widget indicators consist of count variables and thus have a Poisson distribution, Spearman's rank correlation was used, i.e. all widget scores as well as all grades were ranked with 1 being assigned to the highest ranking scores and grades and ties being assigned an average rank. Due to the ranking, differences in grading style between tutors as well as differences in units and scales were thus corrected for. Spearman's rank correlation coefficient was calculated to determine the strength of association between ranked grades and widget indicator scores as well as the significance level. The correlation coefficients were calculated for all runs pooled for the entire length of a run and for individual months.

In order to not only learn something about the strength of association but also about predictive relations between widget indicator scores and grades, more advanced statistical analysis on the data is necessary. For analyses such as structural equation modelling, however, the data needs to be normally distributed. With the data from the widget indicators having a Poisson distribution, this is thus theoretically not possible. However, if the collected count variable data are nearly normally distributed, i.e. if their mean value is far enough from 0, such analyses can be done[5]. As this is the case for most of the means of the widget indicator data (see Table 3.3), we assumed them to be nearly normally distributed and thus, as the second step of our analysis, also conducted structural equation modelling between the three widget indicators (W1, W2, W3) and the students' four individual grades (T1, T2, T3, T4) on the basis of an entire run as well as the individual months for all years pooled.

[5] The mean should be > 10 to be far enough from 0 according to `https://www.umass.edu/wsp/resources/poisson/` and `https://www.umass.edu/wsp/resources/poisson/poisson1.html` and `https://www.umass.edu/wsp/resources/poisson/poisson2.html`.

Table 3.3 Descriptive statistics of the widget indicators ‘W1 initiative’, ‘W2 responsiveness’ and ‘W3 presence’: all runs pooled, activity measured over the entire length of a run as well as per month

	N	Range	Min	Max	Mean		Std. Dev.	Variance	Skewness		Kurtosis	
	Stat.	Stat.	Stat.	Stat.	Stat.	Std. Err.	Stat.	Stat.	Stat.	Std. Err.	Stat.	Std. Err.
W1 all months	172	124	0	124	17,30	1.323	17.346	300.888	3.182	.185	15.507	.368
W2 all months	172	217	6	223	59.86	2.800	36.721	1348.448	1.414	.185	2.543	.368
W3 all months	134	5239	240	5479	1291.88	85.539	990.186	980468.452	2.109	.209	4.780	.416
W1 month1	172	40	0	40	3.53	.393	5.151	26.531	3.513	.185	17.521	.368
W2 month1	172	89	1	90	17.00	.922	12.097	146.327	2.328	.185	8.898	.368
W3 month1	134	2409	35	2444	378.13	30.597	354.187	125448.583	2.967	.209	11.496	.416
W1 month2	172	74	0	74	3.44	.506	6.637	44.049	7.496	.185	75.248	.368
W2 month2	172	54	0	54	11.30	.648	8.500	72.245	1.672	.185	4.064	.368
W3 month2	134	1392	21	1413	227.98	18.692	216.371	46816.443	2.514	.209	8.909	.416
W1 month3	172	15	0	15	2.03	.196	2.574	6.625	2.118	.185	5.776	.368
W2 month3	172	37	0	37	7.93	.469	6.153	37.855	1.750	.185	4.819	.368
W3 month3	134	1093	6	1099	177.86	13.640	157.898	24931.671	2.411	.209	9.298	.416
W1 month4	168	60	0	60	3.82	.450	5.833	34.028	5.825	.187	51.470	.373
W2 month4	172	50	0	50	11.20	.775	10.169	103.399	1.668	.185	2.696	.368
W3 month4	134	1294	6	1300	243.81	21.002	243.113	59103.777	2.364	.209	6.251	.416
W1 month5	172	38	0	38	4.56	.372	4.875	23.769	2.478	.185	12.034	.368
W2 month5	166	58	0	58	12.89	.841	10.841	117.520	1.654	.188	3.561	.375
W3 month5	134	1216	18	1234	264.10	20.734	240.019	57609.186	1.919	.209	4.195	.416

Table 3.4 Spearman correlation coefficients of the association between individual grades and widget indicator scores: all runs pooled, activity measured over the entire length of a run

			W1 initiative	**W2 respon-siveness**	**W3 presence**
T1	planning &progress	Corr Coeff	.267**	.338**	.084
		N	172	172	134
T2	contribution to team	Corr Coeff	.316**	.415**	.192*
		N	172	172	134
T3	support	Corr Coeff	.299**	.414**	.216*
		N	172	172	134
T4	individual overall	Corr Coeff	.313**	.414**	.182*
		N	172	172	134

**. significant at the 0.01 level (2-tailed).
*. significant at the 0.05 level (2-tailed).

3.3 Results

3.3.1 Correlations

The correlation calculations were conducted using IBM's SPSS Statistics 23. The results of Spearman's rank correlation for all runs pooled (see Table 3.4) show that when student activity is measured during the entire length of the course run, all four tutor-based grades (T1-T4) are significantly and positively correlated with all widget indicators (W1-W3) except for the T1/W3 combination. For the widget indicators, the highest correlation coefficients are obtained for the indicator 'W2 responsiveness' and the lowest for the 'W3 presence' indicator.

This holds true for all grades. For the grade 'T1 planning & progress' the correlation coefficient obtained with the 'W2 responsiveness' indicator is .338, for the grade 'T2 contribution to team' it is .415 and for the grade 'T3 support' it is .414. The 'T2 contribution to team'/ 'W2 responsiveness' combination is the highest scoring grade-widget indicator combination but with a correlation coefficient of .415 the 'T3 support'/ 'W2 responsiveness' combination as well as the 'T4 individual-overall'/ 'W2 responsiveness' combinations are almost as high.

When the Spearman rank correlation coefficients for all runs pooled are calculated per month instead of over the entire length of a run, there are again many grade-widget indicator combinations that are significantly positively correlated (see Table 3.5). All four grades correlate best with the 'W2 responsiveness' indicator in month1 or month2 (i.e. November and December). The 'W3 presence' indicator, again, has the lowest correlation coefficients. While the coefficients for the 'W2 responsiveness' indicator are almost all highest in month2, the 'W1 initiative' indicator

Table 3.5 Spearman correlation coefficients of the association between individual grades and widget indicator scores: all runs pooled, activity measured per month

			W1 initiative					W2 responsiveness					W3 presence				
			m1	m2	m3	m4	m5	m1	m2	m3	m4	m5	m1	m2	m3	m4	m5
T1	planning & progress	Corr Coeff	.347**	.252**	.170*	.189*	.032	.363**	.327**	.195**	.272**	.138	.074	.161	.067	.093	-.056
		N	172	172	172	168	172	172	172	172	172	166	134	134	134	134	134
T2	contribution to team	Corr Coeff	.369**	.324**	.218**	.224**	.050	.378**	.393**	.295**	.353**	.189*	.159	.232**	.150	.193*	.005
		N	172	172	172	168	172	172	172	172	172	166	134	134	134	134	134
T3	support	Corr Coeff	.361**	.284**	.205**	.166*	.064	.356**	.399**	.275**	.330**	.241**	.166	.227**	.179*	.195*	.072
		N	172	172	172	168	172	172	172	172	172	166	134	134	134	134	134
T4	individual overall	Corr Coeff	.380**	.306**	.209**	.207**	.046	.386**	.403**	.272**	.338**	.197*	.151	.232**	.145	.176*	.011
		N	172	172	172	168	172	172	172	172	172	166	134	134	134	134	134

**. significant at the 0.01 level (2-tailed). *. significant at the 0.05 level (2-tailed).

has the highest correlation coefficients in month1. The 'W3 presence' indicator only has a few significant correlations. The highest of these are received in month2. The grade 'T1 planning & progress' never significantly correlates with the 'W3 presence' indicator. The lowest correlation coefficients for all three widget indicators are obtained in month5 with only the 'W2 responsiveness' indicator obtaining significant correlations at all. The 'W3 presence' indicator score of month5 even receives a negative correlation coefficient with the grade 'T1 planning & progress', albeit a non-significant one.

Looking at the correlations from the perspective of the different grades, it shows that the 'T1 planning & progress' grade correlates best with the 'W2 responsiveness' indicator in month1 (.363), and the 'T2 contribution to team' grade correlates best with the 'W2 responsiveness' indicator in month2 (.393), as do the 'T3 support' grade and the 'T4 individual-overall' grade (.399 and .403 respectively).

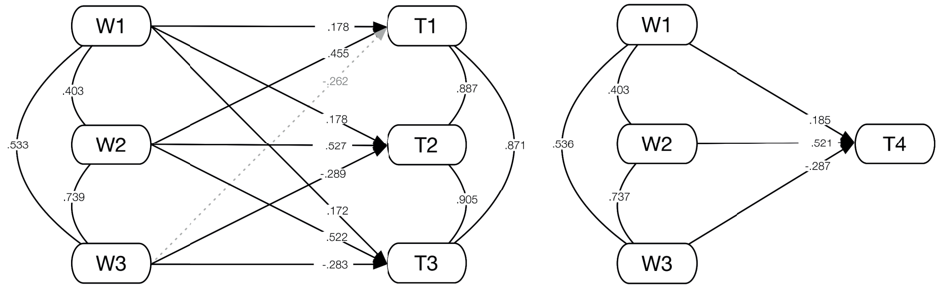
3.3.2 Structural Equation Modelling

Regression analyses using structural equation modelling were performed in Mplus 7. The regressions performed pertained to two situations: in the first one the three grades 'T1 planning & progress', 'T2 contribution to team' and 'T3 support' functioned as the dependent variables, while in the second one grade 'T4 individual-overall' was the only dependent variable. This was done due to T4 being a combination of the other three grades. All calculations were done with all years pooled for the whole run of the course as well as for the individual months.

Different fit indices have been calculated for the different analyses: the Comparative Fit Index (CFI) (Hoyle, 1995; Marsh et al., 1996), the Tucker-Lewis Index (TLI), and the Root Mean Squared Error of Approximation (RMSEA) as well as the Standardised Root Mean Square Residual (SRMR) (Browne and Cudeck, 1989). In order to have a moderate to good model fit these indices should satisfy the following conditions: CFI $\geq .90$; TLI $\geq .90$; RMSEA $\leq .80$; and SRMR $\leq .08$. The model we entered was fully saturated, i.e. all relationships were considered, and all CFIs and TLIs were therefore equal to 1.0 and all RMSEAs and SRMRs were equal to 0.0.

Figures 3.3a and 3.3b depict the results of the two regression analysis situations mentioned above for the entire length of the run. Conducting the structural equation modelling for the entire length of the run and the three grades 'T1 planning & progress', 'T2 contribution to team' and 'T3 support' shows that except for the 'T1 planning & progress'/ 'W3 presence' combination all three widget indicator scores can be used as predictors for the grades (see Table 3.6). The strongest predictive relations are achieved with the 'W2 responsiveness' indicator (all of them are above .455). The relations between the 'W3 presence' indicator and the grades are negative but stronger than the positive relations between the 'W1 initiative' indicator and the grades (the former are around -.285 while the latter are around .175).

Conducting the structural equation modelling for grade 'T4 individual-overall' results in very similar standardised path coefficients (β weights). The strongest predictor for



(a) **(b)**

Figure 3.3 Graphs of the structural equation modelling with standardised path coefficients (β weights) for grades T1, T2 and T3 (left 3.3a) and for grade T4 (right 3.3b) with all widget indicator scores: all runs pooled, activity measured for the entire length of the run

Table 3.6 Standardised path coefficients (β) and their significances from the structural equation modelling with the individual grades as dependent and the widget indicator scores as independent variables: all runs pooled, activity measured over the entire length of a run

			W1 initiative	**W2 respon siveness**	**W3 presence**
T1	planning	β	.178*	.455**	-.262
	& progress	Sig.	.045	.000	.059
T2	contribution	β	.178*	.527**	-.289*
	to team	Sig.	.040	.000	.032
T3	support	β	.172*	.522**	-.283*
		Sig.	.048	.000	.040
T4	individual	β	.185*	.521**	-.287*
	overall	Sig.	.034	.000	.035

**. significant at the 0.01 level (2-tailed).
*. significant at the 0.05 level (2-tailed).

the grade is the 'W2 responsiveness' indicator while 'W3 presence' shows a negative predictive relation. All three widget indicators obtain significant relations.

Looking at the standardised path coefficients of the structural equation modelling for the different months (see Table 3.7) shows that the 'W2 responsiveness' receives a positive and significant relation with all grades in all months, i.e. it can be used as a predictor for the three grades. The 'W3 presence' indicator always obtains a negative relation with the grades which is significant only in month1 (-.333). For indicator 'W1 initiative' the relations are positive and significant in month1 and month3 only.

Table 3.7 Standardised path coefficients (β) and their significances from the structural equation modelling with the individual grades as dependent and the widget indicator scores as independent variables: all runs pooled, activity measured per month

			month 1			month 2			month 3			month 4			month 5		
			W1	**W2**	**W3**	**W1**	**W2**	**W3**	**W1**	**W2**	**W3**	**W1**	**W2**	**W3**	**W1**	**W2**	**W3**
T1	planning &progress	β	.314**	.434**	-.333*	.050	.385**	-.149	.180*	.247*	-.113	.117	.313**	-.145	-.014	.249*	-.060
		Sig.	.000	.000	.023	.630	.000	.326	.025	.018	.368	.227	.004	.318	.877	.040	.661
T2	contribution to team	β	.295**	.372**	-.223	.033	.423**	-.136	.185*	.338**	-.142	.140	.382**	-.186	-.032	.355**	-.140
		Sig.	.000	.002	.127	.746	.000	.359	.018	.001	.241	.146	.000	.187	.724	.003	.295
T3	support	β	.297**	.371**	-.235	.024	.424**	-.140	.163*	.300**	-.097	.126	.374**	-.173	-.020	.357**	-.115
		Sig.	.001	.002	.114	.812	.000	.349	.040	.003	.433	.188	.000	.227	.821	.002	.395
T4	individual overall	β	.306**	.398**	-.254	.036	.425**	-.141	.185*	.309**	-.123	.134	.371**	-.175	-.020	.339**	-.114
		Sig.	.000	.001	.088	.724	.000	.345	.019	.002	.320	.165	.000	.219	.826	.004	.399

**. significant at the 0.01 level (2-tailed). *. significant at the 0.05 level (2-tailed).

In month5 all grade/ 'W1 initiative' combinations have a negative relation but are not significant.

From the perspective of the grades, the highest positive predictive relation for 'T1 planning & progress' is achieved with the 'W2 responsiveness' score in month1 (.434) while the strongest negative predictive relation is received with the 'W3 presence' score in month1 (-.333). The best positive predictive relation for grades 'T2 contribution to team', 'T3 support' and 'T4 individual-overall' is obtained with the widget indicator score 'W2 responsiveness' in month2 (.423 and .424 and .425). There are no significant negative predictive relations for these grades in the individual months.

3.4 Discussion

When all runs are pooled and the activity is calculated over the whole run of the course, the Spearman correlation results show that the scores of all three widget indicators significantly and positively correlate with all four grades except the 'T1 planning & progress'/ 'W3 presence' combination whose relation is not significant. Hypothesis 1 (There is a significant positive correlation between tutor gradings of individual students and the widget indicator scores) can thus be accepted.

Adding to this, the results of the structural equation modelling shows that there is indeed a positive and significant predictive relation between the widget indicators 'W1 initiative' and 'W2 responsiveness' and all four grades while the widget indicator 'W3 presence' is in a significant but negative relation with the grades 'T2 contribution to team', 'T3 support' and 'T4 individual-overall', i.e. the widget indicators in those cases can be seen as predictors for the grades. The individual grades of the students as given by the tutors are mostly defined in qualitative terms (see Table 3.1). However, the analysis results between the purely quantitative widget indicator scores and these individual grades suggest that posting more while having lower presence scores tends to lead to better course grades, i.e. the more productive students (see our definition of the 'W5 productivity' indicator in Table 3.2) seem to be the better performers.

In particular, scores of the 'W2 responsiveness' indicator, i.e. the number of response posts made on the platform, correlate well with the different individual grades. This holds true for the calculations of the whole run as well as for the individual months. This suggests that it provides a reliable indication of students' individual performance. As the correlation between the scores of the widget indicator 'W3 presence' and the four grades tends to be lowest (but still significant) for the whole run as well as the individual months and as – except for the T1 / W3 combination – the 'W3 presence' indicator scores have no significant predictive relation with any of the grades, Hypothesis 2 (The scores of the widget indicator 'presence' are better predictors for the students' individual grades than those of the widget indicators 'initiative' and 'responsiveness'), is rejected.

This is interesting as a number of related works reported that class attendance or time online can be used as predictors for the course outcome. Also, one would intuitively assume that those students that are most interested in and motivated for the course are also those that show a high presence on the platform and thus receive the better grades. However, this does not seem to be the case here. The 'W3 presence' indicator scores therefore are not a very good a predictor for the students' individual grades. Our results thus correspond with those from Macfadyen and Dawson (2010) who reported that contribution to discussions, i.e. posting something, received better correlation results with students' outcome than time online.

The positive and significant Spearman correlation results as well as the positive and significant regression analysis results between the score of the widget indicator 'W2 responsiveness' and the individual grades especially in month2 could be explained by the observation that in the first months of the course, the students almost exclusively use the EVS platform, whereas after these months the students increasingly move to other means of communication, outside the EVS platform, notably Skype and Google Docs. As a consequence, a large part of the students' activity in these later months is not measured by the learning analytics widget. Based on the widget data alone, hypothesis 3 (The widget indicator scores produced in the second half of the course are better predictors than those of the first half) thus has to be rejected.

Again, this finding is interesting as we had originally thought that the last few months of the course would render better results than the first few as the most part of the group work in EVS is done towards the end of the course. The change to other means of communication over the time span of an EVS run, however, seems to have more impact than foreseen. The increased use of these other tools in the later months does, however, not necessarily mean that the students made fewer posts on the EVS platform (overall, the number of initiative posts increased towards the end, while the number of response posts decreased; presence also slightly decreased towards the end). It does, however, mean that there was a relative shift, i.e. the share of communication and collaboration decreased relative to the share outside the platform, and that there was a qualitative shift, i.e. the platform was still used for communication but much less for collaboration on joint products. The expected increase of activity thus did happen but not on the EVS platform and could thus not be captured by the widget.

Pertaining to the discussion about the effectiveness of learning analytics visualisations, our study contributes to it as we provide evidence for the effectiveness of dashboards for reflection and awareness of pure online collaborative learning processes. We investigated the predictive power of the indicators from our widget and were able to show that the final grades and widget indicator scores are significantly and positively correlated. This overall positive result provides a useful empirical basis for the development of instructional designs and activities within the EVS online course. As the EVS students do not meet face-to-face, we are confident that the widget, once it is implemented in a live run of the course, will support reflection and awareness of the collaborative learning processes, will provide valuable feedback to the learners on different activities of collaborative learning, and will contribute to an adjustment of the learning design of the course.

There are several aspects that have to be kept in mind when looking at the results of our analyses. First of all, as mentioned earlier, analysing distal data such as activity logs from a learning environment can never be used as a one-to-one replacement for proximal data such as questionnaires or interviews. However, we support the view that the use of learning analytics can contribute to and enrich reflection and awareness processes for learners as well as teachers especially due to its non-disruptiveness and its taking into account of the full student cohort at the same time.

Another limitation of our study is that although we do look at behavioural data, we do not examine learning as a process itself. Neither do we explore whether any learning actually took place (for the purposes of our study we assume that a student's grade is an indicator of knowledge level) nor do we actually observe learning where and how it takes place, e.g. in the form of brain activity and modifications. Bio-psychological and educational neuroscience research is of huge importance for discovering the phenomenon of learning. On many levels, however, the brain and its ways of working are still a mystery (Bruyckere et al., 2015; Martynoga, 2015). And although the recent year has seen learning analytics researchers contributing to this field by combining log data with data from biophysical sensors (e.g. (Pijeira-Díaz et al., 2016)), addressing and taking into account these issues is out of the scope of this thesis.

One of the biggest risks associated with this type of awareness and reflection support widget, or better, with this type of visualised information as we describe here is that students could use it 'strategically', e.g. by posting many short, largely irrelevant messages to improve their scores. Beheshitha et al. (2016) report that showing students the top contributors of their course often resulted in more postings but not necessarily in ones with higher quality. As we did not use the widget in a live run of a course for this study, we did not have to take this risk into account yet. However, once the widget will be used, the best way to deal with such risks is to properly embed it into the instructional design of the course and to explain its aim and function to students and tutors. This might help to overcome issues like students 'playing the system' and tutors only using the widget indicators scores for grading. In addition, it may be useful to introduce a weighted form of scoring in the widget, e.g. by taking the length of posted comments into account, and to control for achievement goal orientations (Beheshitha et al., 2016; Lonn et al., 2015).

Relating to the usage of the widget in a live run of the course, it will also be interesting to observe if and how the students will make use of the privacy option offered by the reciprocal privacy model implemented into the widget. Theoretically, if many or even all students within a group choose not to share their data, the widget's intention to support awareness and reflection of collaborative learning processes would be seriously interfered with or even prevented. A further risk is thus that by providing the students with privacy mechanisms, the likelihood of the widget being able to be the supportive tool it is meant to be decreases.

3.5 Conclusion

This chapter presented a formative study about the reflective and thus predictive power of widget indicators of a learning analytics-based awareness widget towards students' grades. The results of our analysis show that the grades and widget scores are indeed significantly and positively correlated, with some widget indicators being valid reflectors, i.e. predictors, of the grades. On the basis of the results presented and discussed above, we suggest several guidelines concerning the interpretation of this learning analytics widget's visualisations in a live run of the course.

The scores of the widget indicator 'W3 presence' are not to be seen as a valid reflector for the final tutor-based grades of an individual student as they tend to have non-significant and negative predictive relations with all grades. They can, however, be useful to make students within a team aware of their group's dynamics.

The 'W2 responsiveness' indicator scores provide a good indication of an individual student's contribution to the group work and can thus be used as a basis for group reflection. Due to the significant and positive correlations and predictive relations of this widget indicator with all grades in the first few months, it can be used as a reflector for the students' final individual grades, under the condition of unchanged behaviour.

Taking the results from this analysis into account, the learning analytics widget is being integrated into the course platform for tutors and students in future live runs of EVS. Its impact on group awareness processes will be analysed with quantitative and qualitative measures such as the evaluation framework for learning analytics (Scheffel et al., 2014) and face-to-face experts workshops.

Chapter 4

Straight To The Top – Or: Widget, widget as you lead, I am performing well indeed!

After conducting the exploratory study presented in the previous chapter where we investigated the predictive power of several indicators of the activity widget towards the students' grades, this chapter presents how the learning analytics widget was implemented and used in the environment of an online course. The same analyses as in the exploratory study are run again on this year's data to compare the results. They show that there are indeed predictive relations between the students' actions and their grades and they indicate that some differences in results can be attributed to the availability of the widget. Additionally, the widget is evaluated using the EFLA-2 to show that the framework can be used to evaluate a learning analytics application at several points in time and to reflect differences between the two stakeholder groups.

This chapter is based on:

Scheffel, M., Drachsler, H., Kreijns, K., de Kraker, J., and Specht, M. (2017). Widget, widget as you lead, I am performing well indeed!: Using results from an exploratory offline study to inform an empirical online study about a learning analytics widget in a collaborative learning environment. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge*, LAK '17, pages 289–298, New York, NY, USA. ACM.

4.1 Introduction

One way to support the collaborative learning processes of student teams in virtual learning environments is to provide explicit information to the students about the activities of the group members and to stimulate awareness, reflection and social interaction (Kirschner et al., 2015). Although using behavioural data automatically collected from the learning environment is not to be seen as a one-to-one replacement for using subjective data collected via questionnaires or interviews (Fishbein and Ajzen, 2010), making use of learning analytics based on interaction data does have the advantage of being non-disruptive and covering the whole student population of a course. A learning analytics widget in a computer-supported collaborative learning environment can thus provide feedback (Hattie and Timperley, 2007; Mory, 2004) to students as well as teachers by visualising the students' activities within the virtual learning environment in order to facilitate awareness and reflection (Phielix et al., 2011).

Endsley (1995, 2000) describes being aware of one's own situation as a three level process: (i) perceiving the elements in the current situation, (ii) comprehending the situation, and (iii) projecting what a future status could look like. Once awareness of the situation is established, a user can reflect on it in relation to his behaviour (Schön, 1983) and can subsequently adapt or even change his behaviour if necessary. According to McAlpine and Weston (2000) reflection is to be seen as a mechanism that can improve teaching and thus maximise learning and not as an end in itself. Reflection processes and behavioural change are, however, not only influenced by awareness (Butler and Winne, 1995). Whenever someone engages in self-regulated learning, they bring their own knowledge, beliefs and skills into the process (Winne, 1995). Additionally, emotions, the social environment as well as one's own behaviour play a role (Zimmerman, 1995). The way in which someone acts and reacts in a given situation thus depends on the different ways they have constructed their current knowledge (Winne, 2006).

The relevance of these aspects has been emphasised by Verbert et al. (2013) in their process model for learning analytics applications that consists of four stages: awareness, reflection, sensemaking, and impact. As the discussion about the effect of learning analytics and the need for empirical studies has increased (Siemens et al., 2013; Ferguson and Clow, 2016), a number of recent studies have investigated the impact of learning analytics dashboards on different aspects, e.g. individual goal attainment and motivation. Lonn et al. (2015) investigated whether the motivation of students in a summer bridge program, i.e. students among the at-risk population in postsecondary education, was affected by the use of learning analytics. Their findings suggest that being exposed to a learning analytics application displaying their academic performance can negatively predict the change of mastery orientation, i.e. it decreases, and can thus affect a student's interpretation of their data and their success. The authors stress that student goal perception and formative performance thus need to be carefully considered when designing learning analytics interventions.

Beheshitha et al. (2016) also examined the effect of learning analytics visualisations. Their experiment took place in a blended course setting where each student was randomly assigned to one of three available visualisations. The results revealed that the visualisations had different, i.e. positive or negative, effects on the quality and quantity of forum posts by the students that depended on the students' achievement goal orientation. These authors stress that it is important to consider individual differences such as achievement goal orientation in the design process of learning analytics visualisations. A third study by Khan and Pardo (2016) showed that students use learning analytics differently, i.e. depending on their information need or the learning activity or phase. All three of these studies clearly emphasise that for learning analytics visualisations to have a positive effect, they need to be embedded into the instructional design and that the students' personal preferences, e.g. goal attainment or motivation, need to be considered.

In order to add further results to the collection of empirical data studies, we have designed a learning analytics widget called 'activity widget' and implemented it into the learning environment of the European Virtual Seminar (EVS), an online course where geographically dispersed students work together on different topics in small teams. Based on data automatically collected in the EVS platform, the activity widget is made up of several radar and bar charts. The aim is to make students aware of the platform activity of their team in relation to their own activity level. Apart from making students aware, the activity widget also aims to foster reflection about how their behaviour can influence their position in the team and their course outcome.

4.1.1 Exploratory Offline Study

In a previous exploratory data study (Scheffel et al., 2017a)[1] – referred to as 'exploratory study' throughout this chapter – we investigated the predictive power of several indicators of the activity widget towards the students' grades by instantiating these indicators with data from the four previous runs (2011-2012, 2012-2013, 2013-2014 and 2014-2015) of the European Virtual Seminar on Sustainable Development (EVS). That is, although the activity widget had not been used in those years, we analysed the log data from these years to explore what the widget indicator scores would have been if the widget had been used in those years. We tested whether the students' activity scores of the previous runs correlated with the tutor gradings and whether they validly reflected them. We did so for the whole run of the courses as well as for individual months.

More specifically, in the exploratory study we wanted to know (1) whether the widget indicator scores correlate with the tutor gradings of individual students at all, (2) whether the scores of some widget indicators are better predictors for the students' individual grades and (3) whether certain points in time produce indicator scores that are better grade predictors than others. We hypothesised that significant positive correlations exist between the widget indicators and the grades, that the

[1] This publication is included as **Chapter 3** in this thesis.

widget indicator 'presence' (see explanation below) is a better predictor than the other ones and that the widget indicator scores produced in the second half of the course are better predictors towards the grades than those in the first half of the course.

The results of the correlation analysis and the structural equation modelling (SEM) of the exploratory study showed that most of the indicators indeed significantly and positively correlated with the grades and that they can be used as predictors. The scores of the 'presence' indicator, however, did not turn out to be better predictors for the grades, neither for the whole run nor for the individual months. Instead, the 'responsiveness' indicator achieved the best results. Looking at the individual months, the analysis showed that the months in the first half of the course yielded better correlation and SEM results than those in the second half. This unexpected outcome was due to an unforeseen large usage of communication tools outside of the course's learning environment. For detailed results and their discussion please refer to chapter 3.

4.1.2 Approach

Keeping these results in mind, we implemented the activity widget into the learning environment of EVS and made it available to students and tutors in the 2015-2016 run of the course. In this current study (referred to as 'online study' throughout this chapter) we investigate whether using the activity widget live in a run of the course yields similar or different correlations between the widget indicator scores and the grades and whether the regression analyses performed in SEM shows approximately the same path-coefficients when compared to the exploratory study. The same set of analyses as used in the exploratory study is therefore applied to the data from the 2015-2016 run. The research questions that guided the correlation and regression analyses in our online study are:

RQ-A1: With the activity widget in use, do widget indicator scores again correlate significantly and positively with the tutors' gradings of individual students?

RQ-A2: With the activity widget in use, are the scores of the responsiveness indicator again better predictors for the students' individual grades than those of the others?

RQ-A3: With the activity widget in use, are the widget indicator scores produced in the first half of the course again better predictors than those produced in the second half?

As the activity widget aims at making students aware of their own activities relative to those of their fellow students as well as fostering reflection about how their behaviour influences their position within the team and the team's collaboration processes, we were interested in the users' experience with the widget during the 2015-2016 run. We therefore evaluated the activity widget using the second version of the evaluation framework for learning analytics (EFLA-2) questionnaire twice:

the first evaluation was conducted in the middle of the course and the second one at the end. Using the EFLA-2 allowed us to take the students' as well as the tutors' points of view into account and to compare the two user groups with one another. The research questions that guided the widget evaluation are:

RQ-B1: Is there a difference in widget evaluation results between the mid-course questionnaire and the end-course questionnaire?

RQ-B2: Is there a difference in widget evaluation results between students and tutors?

The next section describes the course, the activity widget and the evaluation questionnaire in more detail and also elaborates on the method of analysis. After that, we present the results of our online study followed by a discussion and the conclusions.

4.2 Method

4.2.1 Participants and Materials

The EVS Course

Coordinated by the Open University of the Netherlands, the European Virtual Seminar on Sustainable Development (EVS) is a web-based Master course jointly offered by approximately ten different universities in Europe each year, some of which are campus universities while others are distance education institutions. An extensive description of EVS[2] and its aims is provided in (de Kraker and Cörvers, 2014).

EVS runs for five months (November 1 till April 1) every year. During that time students work together on sustainability issues in teams of four to seven, with about six to nine teams every year. Ages range between 20 and 25 years for the students from the regular universities and between 30 and 50 years for those from the distance universities. Every team is coached by a tutor and guided by an expert on the team's topic.

The students' final grade for the course can range from 0 to 10 and is comprised of several components: 50% are based on the grade for a team's research report which is given by the expert; 20% are based on the grade for a team's collaboration process which is given by the tutor; 30% are based on the grade for the individual student's contribution which is also given by the tutor. This last grade is called the 'individual-overall' grade (T4) and is divided into three subgrades: 'T1 planning & progress', 'T2 contribution to team' and 'T3 support'. These four grades evaluating an individual student's contribution are the ones used in our analyses. Table 3.1 on page 65 explains the different aspects covered by these grades.

[2] http://www.ou.nl/evs

Since the run of 2011-2012 EVS has been using an Elgg-based[3] platform which automatically collects and generates data on the students' activities on the platform. This data forms the input to our awareness widget.

The Activity Widget

We developed the widget as an Elgg environment plug-in. It can be downloaded under the GNU GPL version 2 (Slootmaker et al., 2015). The widget is meant to make students aware of their activities on the platform in relation to those of their team members and to then reflect on this information. It also allows the tutors to become aware of the different activity levels of the students in their team. There are five indicators representing different types of activities on the platform: 'W1 initiative', 'W2 responsiveness', 'W3 presence, 'W4 connectedness' and 'W5 productivity'. Table 3.2 on page 66 explains how the scores of the different widget indicators are calculated.

There are two different views available in the activity widget: one showing the widget indicator scores for the whole run of EVS (see Figure 3.1, page 67) and one showing them per month (see Figure 3.2, page 67). The widget indicator scores are automatically calculated from the data recorded in the EVS platform and are scaled from 0 to 10. The team member with the highest activity gets a score of 10 for that widget indicator and the scores of the other team members are then scaled in relation to that. In both views, the team average scores are shown in blue while the current user's scores are shown in orange.

As showing a student's widget indicator scores to the other team members is a privacy-sensitive issue, we followed the process suggested by the DELICATE checklist by Drachsler and Greller (2016) and created a manual explaining the widget's intentions and functionalities. It was distributed to all EVS users making clear what data is collected, how it is visualised and how they can protect their privacy. Implemented within the widget is a Reciprocal Privacy Model (RPM) that allows students to decide whether their team members can see their widget indicator scores or not. Those students that share their data get to see the data from those who also decided to share theirs. Those students that do not want to share their data do not get to see their team members' data. The team average is visible to all students all the time.

The Evaluation Framework for Learning Analytics

The added value of providing learning analytics to students and teachers has clearly been recognised in many educational institutions. While new widgets and dashboards are continuously being developed and implemented, their evaluation has not been standardised yet. We thus developed the Evaluation Framework for Learning Analytics (EFLA)[4] that can be used to evaluate learning analytics tools according to several aspects.

[3] `https://elgg.org/`
[4] `http://www.laceproject.eu/evaluation-framework-for-la/`

Table 4.1 Dimensions and items of the learner and the teacher section of the second version of the evaluation framework for learning analytics (EFLA-2)

		Learners	Teachers
Data	**D1:**	I know what data is being collected.	I know what data is being collected.
	D2:	I have access to my data.	I have access to my students' data.
	D3:	I understand the presented results.	I understand the presented results.
Awareness	**A1:**	I am aware of my current learning status.	I am aware of my students' current learning status.
	A2:	I comprehend my current learning status.	I comprehend my students' current learning status.
	A3:	I can project my future learning status.	I can project my students' future learning status.
Reflection	**R1:**	I reflect on my learning activities.	I reflect on my teaching activities.
	R2:	I reflect on alternative learning activities.	I reflect on alternative teaching activities.
	R3:	I know when to change my behaviour.	I know when to change my behaviour.
Impact	**I1:**	I can detect whether I am falling behind.	I can detect whether my students are falling behind.
	I2:	I study more efficiently.	My students learn more efficiently.
	I3:	I study more effectively.	My students learn more effectively.

The first version of the EFLA was developed with experts from the learning analytics community using a group concept mapping study (Scheffel et al., 2014)[5]. It consisted of five criteria ('Objectives', 'Learning Support', 'Learning Measures and Output', 'Data Aspects' and 'Organisational Aspects') with four items each. In a follow-up study (Scheffel et al., 2015)[6], this first version of the EFLA was evaluated by a small group of learning analytics experts. Based on the results of this evaluation combined with a revisit of the original group concept mapping data as well as a thorough look at related literature, a second version of the framework (EFLA-2) was developed. This version is split in two sections, one for learners and one for teachers, that both consist of four criteria ('Data Aspects', 'Awareness', 'Reflection' and 'Impact') with three items each. Table 4.1 shows the twelve items of the learner as well as the teacher part of the framework. This version was turned into an applicable tool, i.e. a questionnaire for students and teachers, and then used to evaluate the activity widget in EVS.

4.2.2 Procedure

Correlation and Regression Analyses

As in our exploratory study, we used the scores of the widget indicators 'W1 initiative', 'W2 responsiveness' and 'W3 presence'[7] for our analysis. The other two widget indicators 'W4 connectedness' and 'W5 productivity' were excluded again for the same reasons as in the previous study (see Section 3.2.2 on page 68).

We first conducted a t-test to see whether the difference between the widget indicator scores from the online study and those from the exploratory study were significant or not. Then, the scores of the three widget indicators (W1, W2, W3) were correlated with the students' four individual grades given by the tutors (T1, T2, T3, T4) using

[5] This publication is included as **Chapter 1** in this thesis.
[6] This publication is included as **Chapter 2** in this thesis.
[7] For the EVS run of 2011-2012 the 'W3 presence' scores were unfortunately not available.

Spearman's rank correlation. The ranking corrects for differences in scales and units as well as for differences in grading style of the tutors.

We also applied structural equation modelling in order to determine predictive relations between the widget indicators and the grades. Although the data follows a Poisson distribution because the widget indicators consist of count variables, we could assume a normal distribution because most count variable data had a nearly normal distribution and a mean value far enough from 0[8]. We were thus able to do the regression analysis.

Spearman's rank correlations and the t-test were calculated using IBM's SPSS Statistics 23 while the regression analyses were performed in Mplus 7. All calculations were done for the entire length of the run as well as for the individual months.

Widget Evaluation

At the beginning of the course in the fall of 2015, all EVS users received a course manual that included information about the activity widget, i.e. its intentions and functionalities. Two weeks into the course a discussion thread was opened in EVS offering students the opportunity to ask questions about the widget and to comment on it. The discussion thread was kept open and active throughout the course's runtime.

In order to apply the EFLA to the activity widget in EVS, it was turned into a questionnaire. Using online forms, we created a section for each criterion and its three indicators. Every indicator could be rated on a scale from 0 to 6. At the end of the questionnaire, open ended comment boxes were provided for each section asking the users whether they had any comments about this section. Two separate questionnaires were created: one for the students and one for the tutors of EVS.

About halfway through the course, on January 12, 2016, students as well as tutors were sent an invitation to participate in the evaluation of the widget by answering the EFLA questionnaire. They were given ten days to answer. Shortly before the end of the course, on March 18, 2016, students and tutors were invited to participate in a second evaluation round of the widget by answering the EFLA questionnaire again. They were given a week to answer.

4.3 Results

4.3.1 Correlation and Regression Analyses

Looking at the average number of actions per student during the different months gives us a first impression of the students' behaviour of the online study in comparison to the data from the exploratory study. Figure 4.1 shows a student's average number

[8] The mean should be > 10 to be far enough from 0 according to `www.umass.edu/wsp/resources/poisson/` and `www.umass.edu/wsp/resources/poisson/poisson1.html` and `www.umass.edu/wsp/resources/poisson/poisson2.html`.

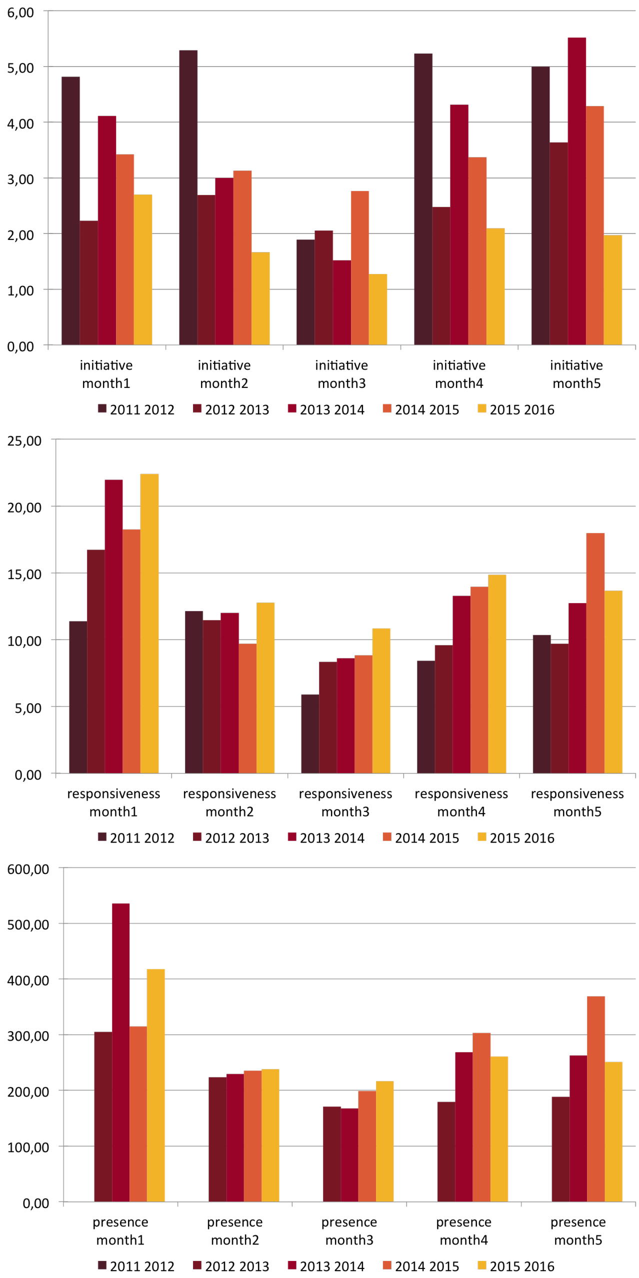


Figure 4.1 A student's average number of actions for the three widget indicators per month for five different years

of initiative and responsiveness posts as well as the presence counts per month for the four years of the exploratory study (2011/12, 2012/13, 2013/14, 2014/15) and the year of the online study (2015-2016) where the activity widget was in use. While the number of initiative posts clearly varies a lot between the years, the number of responsiveness posts and presence counts are much closer together. The most striking difference between the years is that the course run with the widget (2015-2016) has the fewest initiative posts (highly significant, $P < 0.000$) and the most responsiveness posts (marginally significant, $P = 0.053$).

The regression analyses were done in two sets: one had T1, T2 and T3 as the dependent variable while the other had T4 as the dependent variable due to T4 being a combination of the other grades. All Root Mean Squared Errors of Approximation and all Standardised Root Mean Squared Residuals were equal to 0.0 while all Tucker-Lewis Indices and all Comparative Fit Indices were equal to 1.0 except for the CFI of the analysis in month3 between the three indicators and grade T4 which was equal to 0.0.

In the exploratory study, all grade-indicator combinations except the one between 'T1 planning & progress'/'W3 presence' yielded significant and positive correlations when measuring the students' activity over the entire length of the run. In the online study, however, 'W2 responsiveness' is the only widget indicator that positively and significantly correlates with the four grades (see Table 4.2[9]). All grade-W2 correlations are significant at the 0.01 level and higher than .500. That is, there are less significant correlations in the online study than in the exploratory study but those that are significant are stronger.

Table 4.2 Spearman correlation coefficients and standardised path coefficients (β) for individual grades and widget indicator scores based on the entire length of the run from the online study in 2015-2016, n=33

			correlations coefficients				**standardised path coefficients**		
			W1	**W2**	**W3**		**W1**	**W2**	**W3**
T1	planning & progress	Corr.	.234	.508**	.281	β	.190	.366	-.091
		Sig.	.189	.003	.113	Sig.	.452	.063	.702
T2	contribution to team	Corr.	.285	.518**	.168	β	.299	.500**	-.323
		Sig.	.108	.002	.351	Sig.	.200	.005	.142
T3	support	Corr.	.266	.512**	.231	β	.214	.404*	-.148
		Sig.	.135	.002	.197	Sig.	.389	.036	.530
T4	individual overall	Corr.	.285	.527**	.238	β	.238	.438*	-.185
		Sig.	.108	.002	.183	Sig.	.326	.019	.420

**. significant at the 0.01 level (2-tailed).
*. significant at the 0.05 level (2-tailed).

[9] Please refer to Chapter 3 for detailed results of the exploratory study.

Table 4.3 Spearman correlation coefficients for individual grades and widget indicator scores based on the individual months from the online study in 2015-2016, n=33

			W1 initiative					W2 responsiveness					W3 presence				
			m1	m2	m3	m4	m5	m1	m2	m3	m4	m5	m1	m2	m3	m4	m5
T1	planning & progress	Corr.	.336	.154	.188	.200	.297	.421*	.130	.116	.571**	.580**	.221	.024	.119	.453**	.407*
		Sig.	.056	.391	.295	.265	.094	.015	.470	.522	.001	.000	.217	.895	.511	.008	.019
T2	contribution to team	Corr.	.354*	.118	.231	.299	.290	.374*	.101	.274	.599**	.641**	.103	-.120	.018	.393*	.365*
		Sig.	.043	.512	.195	.091	.102	.032	.577	.123	.000	.000	.569	.507	.921	.024	.037
T3	support	Corr.	.305	.036	.124	.371*	.362*	.331	.039	.149	.641**	.656**	.045	-.142	-.013	.481**	.443**
		Sig.	.084	.844	.491	.034	.039	.060	.830	.407	.000	.000	.805	.431	.942	.005	.010
T4	individual overall	Corr.	.372*	.098	.174	.306	.342	.378*	.064	.212	.609**	.669**	.146	-.082	.072	.458**	.424*
		Sig.	.033	.586	.333	.083	.051	.030	.723	.237	.000	.000	.416	.650	.689	.007	.014

**. significant at the 0.01 level (2-tailed). *. significant at the 0.05 level (2-tailed).

Table 4.4 Standardised path coefficients (β) for the individual grades and the widget indicator scores based on the individual months from the online study in 2015-2016, n=33

			month 1			month 2			month 3			month 4			month 5		
			W1	W2	W3	W1	W2	W3	W1	W2	W3	W1	W2	W3	W1	W2	W3
T1	planning & progress	β	.228	.499**	-.232	.313	.217	-.335	-.036	-.029	.231	-.314	.388*	.418	-.225	.411*	.284
		Sig.	.208	.004	.208	.128	.253	.110	.909	.879	.441	.177	.028	.074	.389	.046	.346
T2	contribution to team	β	.257	.480**	-.272	.352	.290	-.524**	.110	.150	-.016	-.055	.513**	.125	-.219	.622**	.129
		Sig.	.157	.007	.140	.072	.105	.007	.726	.421	.957	.811	.002	.591	.367	.001	.645
T3	support	β	.213	.476**	-.267	.160	.206	-.332	-.113	.045	.208	.037	.436**	.185	-.117	.498**	.213
		Sig.	.253	.009	.157	.456	.290	.122	.724	.814	.491	.870	.008	.417	.634	.008	.451
T4	individual overall	β	.237	.504**	-.261	.288	.246	-.407*	-.017	.054	.154	-.123	.462**	.261	-.194	.526**	.220
		Sig.	.189	.004	.155	.161	.191	.048	.957	.777	.611	.589	.005	.257	.432	.005	.439

**. significant at the 0.01 level (2-tailed). *. significant at the 0.05 level (2-tailed).

When calculating the correlations for the online study per month instead of the whole run, the results are again quite different from those in the exploratory study. In the exploratory study the scores of the indicators ‘W1 initiative’ and ‘W2 responsiveness’ correlated significantly with all four grades in months 1, 2, 3 and 4 with W2 also significantly correlating with the grades T2, T3 and T4 in month5. The indicator ‘W3 presence’ had the smallest number of significant correlations with the different grades that were rather low. The strongest correlations were obtained between W2 and all grades in month2. Looking at the individual months, the correlation results from the online study with the live activity widget here also look quite different (see Table 4.3). Overall there are now less significant correlations and hardly any in month1 or month2. The strongest correlation coefficients (ranging from .571 to .669) are received between the ‘W2 responsiveness’ indicator and the four different grades in month4 and month5. All of them are significant at the 0.01 level. Additionally, the previously low scoring ‘W3 presence’ indicator now obtains high and significant correlations with all four grades in month4 and month5.

Conducting the structural equation modelling over the entire length of the run in the exploratory study showed that all three widget indicator scores could be used as predictors for all four grades except the ‘T1 planning & progress’ / ‘W3 presence’ combination. The ‘W2 responsiveness’ was the strongest and most significant predictor. In our current online study, there are only three predictive relations (see Table 4.2), i.e. the ‘W2 responsiveness’ indicator is a predictor for the grades ‘T2 contribution to the team’, ‘T3 support’ and ‘T4 individual-overall’. None of the other indicators can be used as predictors.

Comparing the regression analysis results for the individual months from the exploratory study with the online study again reveals a number of differences. Previously the ‘W2 responsiveness’ indicator was a predictor for all grades in all months with month1 and especially month2 providing the strongest predictive relations. The ‘W1 initiative’ indicator received a predictive relation with all four grades in month1 and month3 while the ‘W3 presence’ indicator was negatively predictive for the ‘T1 planning & progress’ grade only. In the online study, however, the ‘W2 responsiveness’ indicator can only be used as a predictor in month1, month4 and month5 with the latter one holding the strongest predictive relations (see Table 4.4). The ‘W1 initiative’ indicator is in no predictive relation with any of the grades in any of the months. The widget indicator scores of ‘W3 presence’, though, are in a significant negative predictive relation with the grades ‘T2 contribution to the team’ and ‘T4 individual-overall’ in month2.

4.3.2 Widget Evaluation

In order to gauge how the learners and tutors of EVS evaluate their experience with the activity widget, we asked them to fill out the Evaluation Framework for Learning Analytics (EFLA) questionnaire. As we distributed the questionnaire twice during the course, we are able to compare not only the two user types with one another but also any changes in the users’ perception of the activity widget over time. Figure 4.2

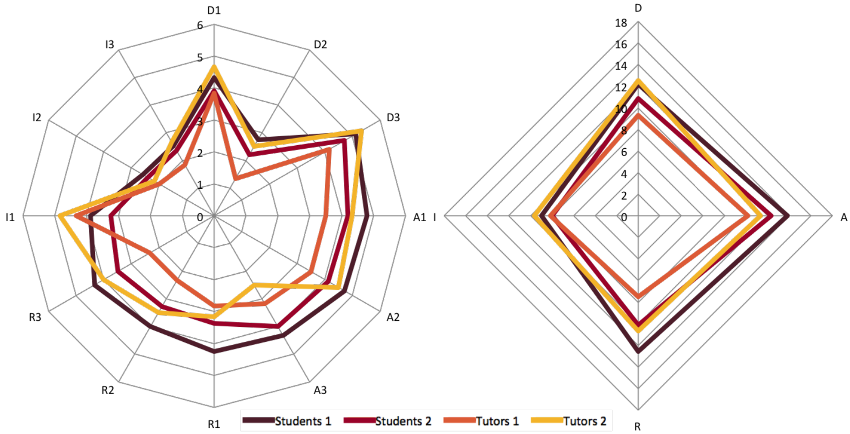


Figure 4.2 Average scores of the twelve EFLA-2 items on the left and the four dimensions on the right for students and tutors for both rounds

shows the average scores of the twelve questionnaire items as well as the combined criteria for both user types and both rounds.

On average students as well as tutors rated awareness and reflection items higher than the items of the data and impact criteria. Also, while the students on average rated the activity widget more positively in the middle of the course, tutors gave more positive ratings at the end of the course.

Conducting a t-test for the four criteria allowed us to see whether the differences between the two user types or between the two rounds were significant or not. Table 4.5 shows the mean, standard deviation and standard error mean for the answers given by students and tutors in rounds 1 and 2. We conducted t-tests for four different settings. First, we compared the answers from the students to those from the tutors in round 1 and round 2. We then compared the answers from round 1 to those from round 2 for each user group. Tables 4.6 and 4.7 show the Levene's test as well as the t-test results for the four different settings.

There are two cases where the differences in ratings are significant. The first one is the rating of the awareness criterion when comparing students and tutors in round 1: $t(28) = 2.158, p = .040$. The second one is the rating of the reflection criterion when comparing round 1 and round 2 of the students: $t(47) = 2.110, p = .040$. None of the other t-tests obtained significant results at the .05 or even the .01 level. In two cases the equality of variance could not be assumed due to the results of the Levene's test. Both of those cases involved the ratings for the reflection criterion from the tutors in round 1, which are rather low, but did not yield significant t-test results. If the equality of variance had been assumed for those cases, however, the difference in ratings between students and tutors for the reflection criterion in phase 1 would have been highly significant (0.006).

Table 4.5 Statistics of the EFLA-2 results for students and tutors for both round

	round	Students n	Mean	Std.Dev.	St.Er.	Tutors n	Mean	St.Dev.	St.Er.
D	1	24	12.21	3.659	.747	6	9.33	5.502	2.246
	2	25	10.84	3.648	.730	6	12.50	4.637	1.893
A	1	24	13.83	3.002	.613	6	10.17	6.014	2.455
	2	25	12.32	4.634	.927	6	11.33	4.633	1.892
R	1	24	12.58	3.296	.673	6	7.50	5.320	2.172
	2	25	10.12	4.720	.944	6	10.67	2.422	.989
I	1	24	9.00	3.901	.796	6	8.17	5.345	2.182
	2	25	8.00	4.223	.845	6	9.67	3.559	1.453

Table 4.6 Results of the Levene's tests and the t-tests for students vs tutors in both rounds

	Round1: students vs tutors					Round2: students vs tutors				
	Levene's test		t-test			Levene's test		t-test		
	F	Sig.	t	df	Sig.	F	Sig.	t	df	Sig.
D	.209	.651	1.555	28	.131	1.006	.324	-.952	29	.349
A	2.903	.099	2.158	28	.040	.129	.722	.468	29	.643
R	4.555	.042	2.236	5.994	.067	2.514	.124	-.273	29	.787
I	1.357	.254	.435	28	.667	1.114	.300	-.891	29	.380

Table 4.7 Results of the Levene's tests and the t-tests for round 1 vs round 2 for both stakeholders

	Students: round1 vs round2					Tutors: Round1 vs Round2				
	Levene's test		t-test			Levene's test		t-test		
	F	Sig.	t	df	Sig.	F	Sig.	t	df	Sig.
D	.132	.718	1.311	47	.196	.024	.879	-1.078	10	.306
A	2.998	.090	1.350	47	.183	.009	.924	-.376	10	.714
R	2.889	.096	2.110	47	.040	7.401	.022	-1.327	6.988	.226
I	1.639	.207	.860	47	.394	1.209	.297	-.572	10	.580

From the open ended questions at the end of each EFLA questionnaire we were able to gather some qualitative feedback about the students' and tutors' impression of the activity widget. Generally most students liked the idea behind the dashboard and appreciated to see their platform activities being set in relation to those of their team members. Many students, however, mentioned several issues they were concerned about: The activity widget was not able to reflect activities outside the platform nor did it take the quality of the posts into account. Some students complained that they noticed people posting irrelevant things in order to achieve higher scores. Some students, though, were made aware that they indeed did less than their team mates and thus participated more in the group.

The tutors also expressed their appreciation of the activity widget in the open ended comments and generally liked having the widget as a reference. For most of them, the activity widget confirmed their own impression about their students throughout the course. With regard to the widget fostering reflection about their own tutoring style it was mentioned that such support would be especially useful in those cases where the student groups do not work together well as the tutors could then use the widget to detect such issues early on. One concern the tutors also had was that the activity widget only reflects actions within the EVS platform and that any work the students do with other online tools is not included.

4.4 Discussion

When student activity is calculated over the whole run of the course, the Spearman correlation results show that in the online study the scores of the 'W2 responsiveness' indicator correlate significantly and positively with the four grades. Research question A1 (With the activity widget in use, do widget indicator scores again correlate significantly and positively with the tutors' gradings of individual students?) can thus be answered with a 'yes'. However, while in the exploratory study the scores of all three widget indicators correlated significantly and positively with at least three if not all four of the grades, in the online study only 'W2 responsiveness' did. But although there are now less correlations that are significant, those that are significant are very strong. Regarding the increase of strength of the Spearman correlation results in our online study, similar results are achieved when calculating the activity for the individual months instead of over the whole run of the course. In comparison to the exploratory study, there are also less correlations that are significant in the individual months but those that are significant are very strong.

Going into the online study, we had expected something like this to happen. Of the three indicators analysed, 'W2 responsiveness', i.e. commenting on the posts, pages or files of others, is the one that best represents team interaction and collaboration. With the activity widget in use during the course, we expected it to foster the students' awareness and reflection about their position within the team and the team as a whole and to thus facilitate collaboration processes.

The standardised path coefficients from the structural equation modelling (see Tables 4.4 and 4.2) show that there are indeed significant predictive relations between some of the widget indicator scores and the grades. As with the correlation coefficients, only the 'W2 responsiveness' scores receive significant results when looking at the entire length of the run. Slightly more diverse results become apparent when looking at the standardised path coefficients for the different months, e.g. in month2 the score of the 'W3 presence' indicator can also be seen as a predictor for some of the grades. However, the best and by far the most frequent predictor for all four grades are the scores from the 'W2 responsiveness' indicator. Research question A2 (With the activity widget in use, are the scores of the responsiveness indicator again better predictors for the students' individual grades than those of the others?) can thus also be answered with a 'yes'. Since the 'W2 responsiveness' indicator scores, although surprisingly at the time, had been by far the best predictor in the exploratory study and since we had anticipated an increase of team interactions due to the widget triggering awareness and reflection processes, we had expected this indicator to be the best overall predictor in the online study as well.

What surprised us, however, was that in the online study none of the predictive relations involved the 'W1 initiative' indicator. As presented earlier, the 2015-2016 run had a significantly lower number of initiative posts per student. When looking into the log data from this year it became apparent that the number of posted files (which is by far the major contributor to the initiative score) was lower during the year of the online study. This may be explained by an increased use of external tools already from early on in the course which cannot be logged and was thus not included in the calculation of the widget indicator scores.

During the exploratory study this use of external tools, especially during the second half of the course where the different group reports needed to be written, turned out to be the most likely explanation for the widget indicators of the first half of the course to be better predictors than those of the second half. As the students in the online study also made use of such external tools, we expected the same to be true for the 2015-2016 run even though the widget was now in use. However, research question A3 (With the activity widget in use, are the widget indicator scores produced in the first half of the course again better predictors than those produced in the second half?) has to be answered with a 'no' as the strongest correlations and best predictive relations are now more likely to happen towards the end of the course.

This shift to the last few months now being the main source for widget scores with predictive power is already indicated by the correlation results: all grade / W2 as well as all grade / W3 combinations in month4 and month5 are significantly and positively correlated with correlation coefficients ranging from .365 to .669. Month2 and month3 do not show any significant correlations in the online study whereas they did so for many grade / widget indicator combinations in the exploratory study. With regard to predictive relations, while none of the widget indicator scores from month3 can be used as a predictor for any of the grades, there are two predictors in month2: The scores of the 'W3 presence' indicator are in a significant negative predictive relation with the grades 'T2 contribution to team' and 'T4 individual-overall'. With regard to the best predictor, the results of the regression analysis in the online study confirm the afore mentioned shift and show that for three grades the best predictors are the scores of the 'W2 responsiveness' indicator in month5, except the grade 'T1 planning & progress' that is best predicted by the 'W2 responsiveness' indicator score in month1.

In the previous years, the widget indicator scores in the last months were poor predictors of the grades, which we attributed to the students mostly using non-logged external tools in this period. In the online study, the widget indicator scores in the last months were the best predictors of the grades. We can think of two probable causes of this shift. First, students that were initially less active may have been stimulated by widget feedback to become more active, resulting in better grades. Second, students, aware that their activities with the external tools were not captured by the activity widget, posted more frequently on the EVS platform as they wanted the widget to reflect their being active in the course.

Another surprising observation for us was the students' neglect of the privacy protection option through the reciprocal privacy model. None of the students disabled this functionality to mask their data from their team. This could be due to the nature of the collaborative learning process that requires to be aware of the status of other students. In fact, we received generally positive responses from the students about the activity widget and that it indeed supported their team awareness processes as well as added a 'fun factor' to the online learning environment.

The results of the formal evaluation of the activity widget using the EFLA questionnaire show that the answer to research question B1 (Is there a difference in widget evaluation results between the mid-course questionnaire and the end-course questionnaire?) is 'yes, but for the students' reflection criterion only' as their reflection ratings in round 2 were significantly lower than those in round 1. In all other cases, neither for the students nor the tutors was there a significant difference in evaluation results between the two rounds. From what we were able to gather from the open ended questions as well as the discussion thread, this difference was most likely due to the students feeling less accurately represented the more the course progressed as the activities in the external tools were not reflected in the widget scores. When comparing the evaluation results from the two user groups with one another, the only significant difference is that of the awareness dimension in round 1. Here, students have rated the awareness items significantly higher than the tutors did. In all other cases, neither in round 1 nor in round 2 was there a significant difference in evaluation results between the two user groups. Research question B2 (Is there a difference in widget evaluation results between students and tutors?) can therefore be answered with 'yes, but for the awareness dimension of phase 1 only'. This is most likely due to the generally positive reception of the activity widget by students already at the beginning of the course while tutors used and thus appreciated the widget more towards the end of the course when they saw their personal impressions about the students confirmed. Except for those two cases, students and tutors thus evaluated the activity widget in a very similar way.

Combining the EFLA results with the comments gathered via the open ended questions allows us to conclude that both students and teachers generally liked and appreciated the activity widget and felt supported in their awareness and reflection processes. Both user groups, however, had issues with the widget's data access (D2) as well as its support of more efficient (I2) and more effective (I3) learning. Additionally, both user groups found it problematic that the activity from external tools could not be included in the widget. Students would also like to see not only the quantity but also the quality of their discussion posts to be taken into account as they otherwise fear that too many irrelevant message are posted to increase the widget indicator score.

We had already identified the risk of students ‘playing the system’ during our exploratory study and had thus provided a detailed user manual at the beginning of the 2015-2016 course explaining the activity widget’s aim and functionalities. This, although being an important step, however, does not seem to have been enough. As emphasised in other studies (Lonn et al., 2015; Beheshitha et al., 2016; Khan and Pardo, 2016) learning analytics visualisations need to be tightly embedded into a course’s instructional design, especially if they are to be used by the students themselves. For the next run of EVS we will therefore carefully take the gathered results into account in order to improve the activity widget as well as the instructional design and to enhance the user experience.

There are several limitations of our study. Due to the change in student population, the students’ behaviour in the five different runs cannot be set into a one-to-one relation. Their previous experience with and usage of online learning platforms as well as external communication and collaboration tools influences the cohort’s actions. The same applies to the tutors. Although many of them have been tutors for EVS for a number of years, their experience and interactions with their student groups also changes from year to year. Related to this aspect of change in student population, student and tutor behaviour as well as external tools is another aspect that has to be kept in mind when looking at the results of our online study: although a number of our observations can be explained as effects of the activity widget being in use, there is no proof that this is the case. Only after observing and analysing further years of the EVS will we be able to attribute differences between the years that did not have the widget and those that did clearly to the use of the widget.

4.5 Conclusions

This chapter presented an empirical study conducted with data collected during the five months of a live Master course where students work collaboratively in virtual teams. We implemented a learning analytics-based activity widget to foster awareness and reflection among the team members into the course’s online learning platform and examined the predictive power of the widget indicators towards the students’ grades of this course in comparison to the data from previous years where the widget had not been in use. Our results indicate that the widget indicator ‘responsiveness’, i.e. the number of response posts made on the course’s platform, is a significant positive predictor towards the grades. In the years without the widget, the students’ behaviour of the first few months of the course held more predictive power, whereas in the year where the widget was implemented into the platform, the last few months of the course had a higher predictive potential. This, in combination with the results from a quantitative as well as qualitative evaluation of the activity widget during the course, suggests that the differences between the years could be explained by the use of the widget and its effective fostering of awareness and reflection. More investigations are needed in order to provide further evidence that can substantiate this hypothesis and confirm the effectiveness of the widget. We will therefore continue to deploy the activity widget in future editions of the course.

Chapter 5

All Stripped Down – Or: Eeny meeny miny moe, catch all items by the toe

This chapter presents the next iteration in the evaluation and improvement process of the evaluation framework for learning analytics. After using the second version of the framework (EFLA-2) in the last chapter to evaluate a learning analytics widget, we now evaluate the framework itself by using the students' and tutors' EFLA-2 questionnaire answers for some quantitative analyses as well as for qualitative analyses. That is, feedback gathered during an experts focus group where all items were systematically and individually discussed is used to identify those aspects of the framework that need improvement. Based on all these results, the third version of the evaluation framework for learning analytics (EFLA-3) is constructed.

This chapter is based on:

Scheffel, M., Drachsler, H., and Specht, M. (in preparation for submission to LAK18). *Eeny meeny miny moe, catch all items by the toe*.

5.1 Introduction

One of the main goals of learning analytics is to understand and optimise learning (Siemens, 2011). During the last few years, more and more educational institutions have thus implemented some form of learning analytics into their learning environments. While some institutions developed elaborate systems to provide their management with information about their data, others aim their analyses at teaching staff or the students. But although learning analytics has been happening for quite a while now, evidence on what works, and what does not, especially with regards to the impact on students and teachers, is sparse. There is currently no standardised way of comparing different learning analytics approaches with one another.

Several indices, instruments, frameworks or models have been created to measure or gauge certain aspects of learning analytics. The Learning Analytics Readiness Instrument (Arnold et al., 2014a), for instance, can be used to identify how ready an institution is to implement learning analytics. Once the analytics are in place, the analytics maturity index (Bichsel, 2012) can be used to measure the level of maturity an institution's currently implemented learning analytics has. Both of those tools, however, are geared towards an institution's administrative or organisational level. Neither of them provide information on how those stakeholders that are actually involved in the learning processes, i.e. the learners and teachers, experience the implemented analytic.

In order to close this gap, we have thus been developing the evaluation framework for learning analytics (EFLA) to help standardise the evaluation of learning analytics tools. Constructing an evaluation framework is not a one-step process. Continuous evaluations are needed to create improved versions in order for such a framework to be widely used and accepted. Starting with a group concept mapping study to identify quality indicators for learning analytics for the first version of the evaluation framework for learning analytics (EFLA-1) (Scheffel et al., 2014)[1], we are continuously using, evaluating and improving the evaluation framework for learning analytics (Scheffel et al., 2015)[2]. Having used the EFLA-2 in a collaborative online learning course (Scheffel et al., 2017b)[3], we now use the EFLA-2 data collected in that study to explore for this current study whether the EFLA-2's structure (see Figure 5.1) needs further improvement, i.e. whether any of the framework's dimensions or items are problematic and should either be adapted or removed. The following research question guided us through this evaluation study:

(RQ) Are there problematic issues with any of the EFLA's items and if so, how can they be addressed to improve the framework?

The next section explains our methodology for this evaluation study, followed by the analysis results. Before concluding the chapter, the results are discussed and set into relation to the research question.

[1] This publication is included as **Chapter 1** in this thesis.
[2] This publication is included as **Chapter 2** in this thesis.
[3] This publication is included as **Chapter 4** in this thesis.

Table 5.1 Dimensions and items of the learner and the teacher section of the second version of the evaluation framework for learning analytics (EFLA-2)

		EFLA-2 items for learners/teachers
Data	**D1:**	I know what data is being collected.
	D2:	I have access to my/my students' data.
	D3:	I understand the presented results.
Awareness	**A1:**	I am aware of my/my students' current learning status.
	A2:	I comprehend my/my students' current learning status.
	A3:	I can project my/my students' future learning status.
Reflection	**R1:**	I reflect on my learning/teaching activities.
	R2:	I reflect on alternative learning/teaching activities.
	R3:	I know when to change my behaviour.
Impact	**I1:**	I can detect whether I am/my students are falling behind.
	I2:	I study/My students learn more efficiently.
	I3:	I study/My students learn more effectively.

5.2 Method

5.2.1 Participants

The European Virtual Seminar on Sustainable Development (EVS)[4] is a Master-level online course coordinated by the Open University of the Netherlands. Approximately ten different European universities (regular ones as well as distance ones) take part in the course each year. De Kraker and Cörvers (2014) offer a detailed description of the course and its aims.

The course takes place every year from November 1 until April 1 of the next year. During this five-months period students from different countries and disciplines are grouped together. In teams of four to seven students, they work together on issues related to sustainability, e.g. climate adaption, nature conservation and waste management. While students from the distance universities are usually between 30 and 50 years of age, those from the regular universities are between 20 and 25. Depending on the number of students in a year, there are usually about nine teams working on different topics. Throughout the course each team is guided by a tutor and an expert on the given topic. In some cases, these two roles are covered by one person.

In the run of 2015-2016, i.e. the year in which the data for this study was collected, there were six teams with a total of 33 students. Every team was guided by a different tutor, i.e. there were six tutors.

The collected EFLA-2 data stems from the evaluation of an awareness widget we

[4] `http://www.ou.nl/evs`

built for the course that displays data based on the automatically collected and generated data of the students' activities on the course platform. The widget is built as an Elgg[5] environment plug-in and can be downloaded under the GNU GPL version 2 (Slootmaker et al., 2015). The widget allows both students and tutors to become aware of the students' activities within their group. A more detailed description of the widget as well as screenshots are given in Chapter 3 (see Figures 3.1 and 3.2 on page 67).

5.2.2 Procedure

The EFLA-2 was turned into two online questionnaires (one for students and one for tutors) using Google Forms[6]. For both stakeholder groups the questionnaire contained a section for each dimension and its items. All items were to be rated on a scale from 0 for no agreement to 6 for high agreement. Every section also contained an open comment box asking users for any possible comments about this section. The EFLA was distributed among all EVS users twice during the 2015/2016 edition of the course. The first invitation to participate was sent on January 12, 2016, i.e. about halfway through the course. The users had ten days to answer. The second invitation to participate was sent on March 18, 2016, i.e. almost at the end of the course. The users had one week to answer.

Using IBM's SPSS Statistics several statistical analyses were conducted: descriptive statistics for the EFLA items, a principal component analysis and a reliability analysis. In addition, graphs showing the average evaluation of each EFLA item for the widget were created for both measurement points. All analyses were done separately for the student data and the tutor data.

On April 13, 2016, a face-to-face full-day experts focus group was organised with the tutors, topic experts and the EVS course coordinator. All participants gave their informed consent to participate in this focus group and for their contributions to be used for the evaluation of the framework. First, the results of an exploratory offline study (Scheffel et al., 2017a)[7] were presented to the focus group followed by a presentation about what the EFLA is and how it had been constructed, evaluated and improved so far. After providing details about how the students and tutors had evaluated the widget at the two points during the course (see Figure 4.2 on page 93), an open discussion took place about all four EFLA-2 dimensions and their items. For each dimension and their items the following questions were used to guide the discussion: (1) What else do we need to measure this better/more precisely?, (2) What would you want to know about a tool?, and (3) How do you explain the differences between student and tutor results? The overarching question for the whole discussion was given as: What do we need to do to improve the EFLA?

[5] `https://elgg.org/`
[6] `http://forms.google.com`
[7] This publication is included as **Chapter 3** in this thesis.

Table 5.2 Descriptive statistics of all EFLA-2 items from both rounds combined for students (left) and tutors (right).

	students				tutors			
	N	Mean	St.D.	Var.	N	Mean	St.D.	Var.
D1	49	4.12	1.509	2.276	12	4.25	1.658	2.750
D2	49	2.47	2.063	4.254	12	1.92	2.466	6.083
D3	49	4.92	1.152	1.327	12	4.75	1.712	2.932
A1	49	4.49	1.570	2.463	12	3.92	1.782	3.174
A2	49	4.41	1.553	2.413	12	4.00	1.809	3.273
A3	49	4.16	1.344	1.806	12	2.83	1.992	3.970
R1	49	3.80	1.514	2.291	12	3.00	1.651	2.727
R2	49	3.63	1.629	2.654	12	2.92	1.379	1.902
R3	49	3.90	1.584	2.510	12	3.17	1.528	2.333
I1	49	3.55	1.768	3.128	12	4.58	1.676	2.811
I2	49	2.49	1.609	2.588	12	2.08	1.564	2.447
I3	49	2.45	1.542	2.378	12	2.25	1.545	2.386

5.3 Results

5.3.1 Statistical Analyses

In the first round 24 students and all six tutors completed the EFLA-2 questionnaire. In the second round 25 students and all six tutors answered. This gives us a total student N of 49 and a tutor N of 12. As the items for students and tutors are formulated slightly differently, all analyses were always done separately for the two stakeholder groups. Table 5.2 shows several descriptive statistics, i.e. mean, standard deviation and variance, for all twelve EFLA-2 items for the students (left) and the tutors (right). For both user groups item D2 seems to be different from all other items as the variance in both cases is much higher than that of the other items (4.254 for the students and 6.083 for the tutors). For the students, item I1 also has a rather high variance (3.128) as does item A3 for the tutors (3.970).

In order to get an idea of the internal coherences of the EFLA-2, i.e. the relationships between the items, we conducted a fixed-four-factor principal component analysis with the data from both stakeholder groups. We did not do this to test the validity of the questionnaire but to get a general overview of which items might be problematic before conducting the experts focus group. Validity testing of the EFLA will be done during the next evaluation iteration.

Table 5.3 shows the rotated matrix results of the four-component analysis for student and tutor data. For the student data all items have a primary loading of .6 or higher. Many of them, however, have quite high cross loads which indicates that these items could possibly also fall into another component. These items are D3, A3, R1, I1 and I2. For the tutor data all items except I1 (.557) have a primary loading of .6 or higher. Item I1 is also prominent in that it has rather high cross loads on more than

Table 5.3 Principal component analysis using Varimax rotation for four components for students' EFLA-2 data (primary loads in yellow, high cross loads in light yellow) and tutors' EFLA-2 data (primary loads in red, high cross loads in light red).

	students				tutors			
	1	2	3	4	1	2	3	4
D1	.685	.350	.043	.411	.708	.177	.496	.366
D2	-.046	.113	.369	.839	.388	-.031	.169	.892
D3	.415	.221	-.235	.695	.772	.230	.548	.088
A1	.893	.266	.125	.060	.426	.320	.709	.369
A2	.902	.216	.150	.169	.545	.475	.625	.155
A3	.400	.674	-.065	.238	.059	.676	.648	.123
R1	.479	.614	.347	.075	.856	.158	.129	.186
R2	.086	.834	.314	.195	.884	.283	.122	.276
R3	.414	.794	.167	.023	.910	.086	.220	.106
I1	.690	.242	.493	-.143	.496	.557	.546	-.312
I2	.129	.512	.602	.154	.159	.940	.209	.126
I3	.195	.141	.900	.087	.221	.922	.156	-.106

one component. The tutor data items of A2 and A3 also emerged with high cross loads. As the four components are not clearly structured along the lines of the four dimensions of the EFLA-2, a need for structural as well as substantial change to the framework's items is indicated.

Although the principal component analysis did not show four clearly structured components in line with the four EFLA-2 dimensions, we conducted a reliability analysis, i.e. we calculated Cronbach's Alpha, for the four EFLA-2 dimensions in order for us to look for further indications of problematic issues with the individual items (see Table 5.4). The lowest reliability score (.630) is received by the students' *Data* scale. All others on the students' as well as on the tutors' side receive a reliability score of at .763 or even above .8 or .9. There are, however, five cases where an increase in Cronbach's Alpha can be achieved if one item is eliminated from the scale. For the students data deleting D2 would result in a *Data* reliability score of .709 and deleting A3 would result in an *Awareness* reliability score of .947. The deletion of both of those items from the tutors' data would also increase Cronbach's Alpha of the respective scales (.941 for *Data* and .933 for *Awareness*). Additionally, for the tutors' data, deleting item I1 from the *Impact* scale would lead to an increased reliability score of .944.

While filling in the EFLA-2 questionnaire, all user had the opportunity to comment on any of the sections. Most of those comments were related to the widget that was being evaluated and only very few comments referred to the questionnaire itself. From those few comments we gathered that both students and tutors of EVS generally did not have much difficulty to answer the questions. However, item D2 about having access to the collected data was problematic for both stakeholder

Table 5.4 Reliability statistics and scale statistics of different item groups for students' EFLA-2 (left) and tutors' EFLA-2 (right)

	students				tutors			
items	Cron.α	Mean	Var.	St.D.	Cron.α	Mean	Var.	St.D.
D1+D2+D3	[1].630	11.51	13.547	3.681	[3].828	10.92	26.265	5.125
A1+A2+A3	[2].859	13.06	15.642	3.955	[4].912	10.75	26.568	5.154
R1+R2+R3	.875	11.33	17.891	4.230	.928	9.08	18.265	4.274
I1+I2+I3	.763	8.49	16.463	4.058	[5].908	8.92	19.356	4.400

[1] increases to .709 if D2 is eliminated
[2] increases to .947 if A3 is eliminated
[3] increases to .941 if D2 is eliminated
[4] increases to .933 if A3 is eliminated
[5] increases to .944 if I1 is eliminated

groups as they often did not understand what was meant with "having access".

5.3.2 Focus Group

During the experts focus group with the tutors, we systematically addressed each dimension of the EFLA individually. Throughout the discussion the focus was set on the overarching question on how to improve the framework. Results of the group concept mapping study (which was used to construct the EFLA-1 and the EFLA-2) were provided to support the discussion.

The *Data* Dimension

Several issues were mentioned by the focus group participants in relation to the *Data* dimension and its items. With regards to the dimension in general it was said that due to the formulation of the statements in the first person singular, i.e. "I know", "I have" and "I understand" the participants felt that it was not the learning analytics tool that was being evaluated but rather that they were being tested. The participants therefore suggested to formulate the statements from a more neutral point of view, e.g. "this learning analytics tool does ..." or "this learning analytics tool provides ..." etc. Another issue that referred to the choice of answers provided for the *Data* dimension was the use of a Likert scale for these items. All participants agreed that the current items were all formulated in a way that asked for a two-dimensional answer of yes or no instead of an agreement with a scale. It was therefore suggested to either change the answer choice or to reformulate the statements in such a way that a Likert scale answer type is supported. Additionally, they suggested to add an "I don't know"-option to the answer options.

With regard to the individual items, none of the participants had problems with answering D1. They did say, however, that answers to this question would most likely not indicate whether the users of a learning analytics tool had been informed about the data collection process, e.g. by a user manual or some form of introduction, but rather whether the users had read or heard, i.e. consumed, these instructions. Item

D2 about data access caused a very intense discussion. Many participants said that it might be unclear what "having access" meant in this context, e.g. seeing one's data in the widget or seeing the actual data logs, and imagined that learners of a course would face the same issues. For some participants, for example, it was not clear that users have the right to see the collected data about them and they doubted that students knew this. Another aspect mentioned about item D2 was that while the other two *Data* items were more about the structural or methodological nature of a learning analytics tool, D2 was more about ethical or transparency issues. The outcome of this discussion resulted in the suggestion to either specifically mention the topic of transparency if this is what the EFLA should be evaluating or to formulate the *Data* items in such a way that it becomes clear that transparency on different levels is the goal. It was then also suggested to make all *Data* items deal with the ethical and transparent handling of the collected data instead of evaluating structural or methodological issues. Item D3 did not seem problematic to the focus group participants in general. They did, however, mention that its transparency aspect about how and why the collected data is being used should be stressed more in order to align it better with the other two *Data* items.

The *Awareness* Dimension

Although the participants did not feel as strong about this as they did for the *Data* section, they did mentioned that formulating the *Awareness* dimension items from a neutral or third person point of view instead of the currently used first person, i.e. "the tool does ..." instead of "I am aware" or "I comprehend", would make the items easier to rate. One aspect the participants had some issues with was the differentiation between items A1 and A2. After having been introduced to the development process of the EFLA at the beginning of the focus group they did understand why the *Awareness* dimension consisted of the three items related to becoming aware, comprehending and projecting a situation, but they agreed that not knowing the theoretical background, any users, whether they are teachers or students, would very likely have difficulties grasping the exact difference between the two items. They argued that for most people becoming aware of something and comprehending it is one intermingled process and not two steps that can clearly be separated from one another. It was therefore suggested to only keep one of these two items. With regards to item A3, participants suggested to replace the verb "project" to simplify the item and make it easier for users of the EFLA to understand it. Suggested alternatives were "foresee", "forecast", "picture", "predict" or "imagine". Another formulation that the focus group participants saw likely to cause problems in understanding the items properly was that of "learning status". They suggested to use "development", "process", "progress", "performance" or "situation".

The *Reflection* Dimension

As with the other dimensions, the focus group participants suggested to use a formulation such as "the learning analytics tool stimulates me to reflect" instead of the currently used "I reflect" as this would make it even easier to understand the

Reflection dimension's items. Additionally, these types of formulations would make it clearer to the EFLA users that a learning analytics tool is being evaluated and not the actions of the users themselves. Similarly to the items A1 and A2 of the *Awareness* dimensions, participants pointed out that it might be difficult for some EFLA users to properly distinguish between the items R1 and R2, i.e. to clearly separate the process of reflecting on one's behaviour from that of reflecting on one's alternative behaviour. It was again suggested to only keep one of the two items. Another issue mentioned by the participants was that the verb "change" used in item R3 might be too strong, i.e. the EFLA users might associate it with "complete reversal" rather than "step-by-step adjustment". Therefore, it was suggested to use the verb "adapt" or "adjust".

The *Impact* Dimension

For the *Impact* dimension, the focus group participants also suggested to not use the first person point of view for the items. Doing this would on the one hand, again, make it clearer to the EFLA users that the goal is to evaluate a learning analytics tool and not their behaviour, and would on the other hand make the whole framework more consistent if it were to be done for all items in all dimensions. Generally, the participants found it difficult to pin down the impact of a learning analytics tool to a few questions. They did understand why the current items had been chosen for the framework but suggested to possibly include an item about motivation as they deemed this highly important for learners as well as teachers of a course. The participants also remarked that the three *Impact* items relate to different things: While I1 evaluates a specific feature of a learning analytics tool, i.e. whether the tool identifies students at risk of falling behind, the item I2 and I3 refer to a more general characteristic of a tool, i.e. making the students' learning processes more efficient and more effective. With regards to the items I2 and I3, the focus group participants stated very clearly that for a teacher or tutor of a course there would be no real way of knowing whether a learning analytics tool helps their students to learn more efficiently or more effectively. They could, however, gauge whether a learning analytics tool helps themselves to teach or tutor more efficiently and more effectively. It was therefore suggested to change the focus of I2 and I3 to the teachers'/tutors' impression of their own situation. Finally, according to the participants, the EFLA would be the easier to answer the more consistent (in terms of item addressee, item goal, etc.) a dimension is. They thus suggested to keep this in mind when improving the framework.

5.4 Discussion

From the results of the quantitative as well as the qualitative analysis it was very clear that several items of the framework had to be adapted and that an improvement of the framework was needed. All changes, however, only involved the items, i.e. the dimensions stayed the same as did the framework's split in a learner and a teacher section. One change that was applied to all items was the shifting of the items' syntactical subject. Instead of all statements reading “I do ...”, they now all refer to the learning analytics tool as the source of action to be evaluated and read “This learning analytics tool does ...”.

For the *Data* dimension, item D1 did not seem to be problematic as no issues with it came up in either the statistical analyses or during the focus group. Item D2 about having access to the collected data was very problematic. Not only did it already stand out in the descriptive statistics for both user groups, it also did so during the principal component analysis where it formed its own component for the tutors and almost formed its own component for the students given the high cross load of D3. D2 equally stuck out in the reliability analysis as it was one of the items that, if deleted, would lead to a better Cronbach’s Alpha value of its dimension’s scale for both students and tutors. The participants of the focus group confirmed these issues and called for a clearer focus for the topic of transparency for this item and the *Data* dimension in general. Although not being a problematic issue in itself, participants also suggested a clearer connection of item D3 to the transparency topic.

Based on these inputs, we decided to focus the *Data* dimension on the three aspects of “what”, “why” and “who” in relation to data handling and transparency and thus make the whole dimension easier to understand. Item D1 about “what data is being collected” is thus kept for the new framework. Item D3 about the aspect of understanding the data is reformulated and shifted to the aspect of “why data is being collected”. The problematic item D2 about having access to one’s own data is reformulated as well and shifted to the aspect of “who has access to the data” in hopes that the item is now less difficult to understand. Due to the aforementioned change of the used point of view, all *Data* items now begin with “For this learning analytics tool it is clear ...". We decided to use the wording of “it is clear” to allow for an easy use of Likert scale ratings as answer options. Adding in an additional “I don’t know”-answer was not an option as this had already been ruled out during the creation of the EFLA-2.

With regards to the *Awareness* dimension, we reformulated all items to read “This learning analytics tool makes me ...”. Additionally, the term “learning status” was changed to “learning situation” due to the focus group’s suggestion. Using the term “situation” also better reflects the wording used in the model of “situation awareness” by Endsley (2000).

Neither item A1 nor item A2 seemed to be very problematic looking at the statistical analyses. Only item A2 about comprehending the students’ learning status stood out slightly in the principal component analysis as it had a very high cross load for the tutors’ data. The focus group participants attributed this to a possible difficulty to properly distinguish the actions of becoming aware of a situation and comprehending a situation. Therefore, in order to make answering the EFLA straight forward and as quick and easy to answer as possibly, we followed the focus group’s suggestion to only keep one of the two items. Item A2 was thus removed and will not be part of the new EFLA version.

In contrast to A1 and A2, item A3 already showed signs of being a problematic item during the statistical analyses. In the descriptive statistics its variance for the tutor data was slightly higher than that of most other items and in the principal component analysis it had high cross loads for both the students' and the tutors' data. For the reliability analysis of the *Awareness* scale, item A3 was flagged for both students and tutors, i.e. its deletion would increase the scale's Cronbach's Alpha value. We assume these problems to be caused by the verb "project" as was also mentioned by the focus group. We therefore decided to replace it with "forecast". Additionally, in order to make the meaning of forecasting even clearer, the wording of "future learning situation" was changed to "possible future learning situation" and the phrase "given my / my students' (un)changed behaviour" was added to the statement. We anticipate that in doing so we have substituted the verb "project" in a semantically adequate way.

The only item from the *Reflection* dimension that had a slight indication of being problematic in the statistical analyses was R1 as it received a high cross load in the principal component analysis for the students' data. Why this was the case was not obvious to us. The focus group had suggested for a possible lack of clear distinction between items R1 and R2 and to thus keep only one of them. In order to make the use of the EFLA quick and easy for future users, we therefore decided to do several things. First, we did remove item R2 but kept item R1. Second, we reworded R1 to clearly refer to reflection "on my past learning / teaching behaviour" so as not to allow for any vagueness. Taking up the pattern of changing the point of view from the EFLA user to the learning analytics tool, the item now begins with "This learning analytics tool stimulates me to reflect ...".

Following the suggestion from the focus group, the verb "change" in item R3 was replaced by "adapt" to better reflect the notion that a learning analytics tool can foster the reflection about changing something in one's behaviour but that this is not necessary in all situations. To support this, we also used the wording of "This learning analytics tool stimulates me" at the beginning of the item and additionally added the phrase "if necessary" at the end.

In the *Impact* dimension, item I1 turned out to show problematic tendencies in all statistical analyses. It had a higher variance than most other items for the student data and high cross loads for students and tutors in the principal component analysis. It was also one of the items for the tutor data that, if deleted, would increase its scale's Cronbach's Alpha value. The reason for us to remove this item from the EFLA, however, was mainly due to the fact that it did indeed deal with a specific feature of a learning analytics tool, i.e. detecting students at risk of falling behind. Although the detection of students at risk is a very common reason for learning analytics tools to be used, it is a feature of a tool rather than an intended general conceptual goal.

One aspect that the focus group picked up from the presentation of the previous EFLA construction and evaluation iterations and declared missing from the current EFLA was that of motivation. As the aspect of motivation had been one of the original quality indicators of the EFLA-1 and was also one of the themes identified for the

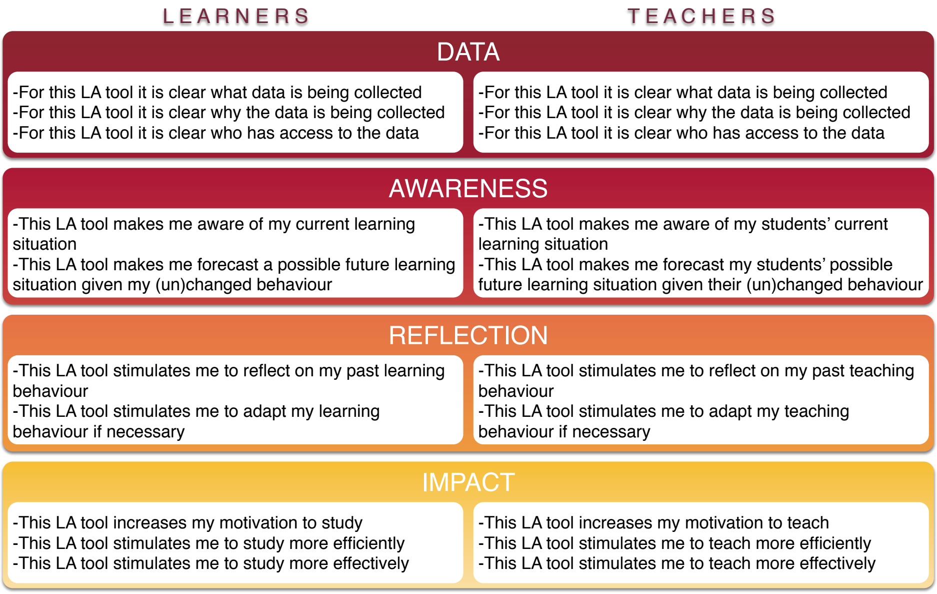


Figure 5.1 Third version of the evaluation framework for learning analytics (EFLA-3)

Impact cluster when creating the EFLA-2, we decided to replace the removed item I1 with one referring to a learning analytics tool's impact on the users' motivation. The new I1 now reads "This learning analytics tool increases my motivation to study / teach".

For items I2 and I3, only I2 was identified as possibly problematic in the statistical analyses as it had a high cross load for the students in the principal component analysis. We assume this to be due to the students not properly understanding the meaning of efficiency although both terms, efficiency and effectiveness had been defined in the questionnaire. A more important aspect, however, was mentioned by the focus group. They clearly said that it is difficult for teachers or tutors to judge whether a learning analytics tool has an impact on the students' learning efficiency or effectiveness. We therefore accepted the focus group's suggesting of shifting the focus of these items to the teaching process. We had already done so with all *Reflection* items as well as with the new item I1. The new version of the teachers' *Impact* dimension thus now only focuses on the teaching process. For both students and teachers, items I2 and I3 now begin with "This learning analytics tool stimulates me to ...".

With these changes we expect all future users of the EFLA being able to quickly and easily evaluate learning analytics tools. Figure 5.1 shows all ten items of the third version of the evaluation framework for learning analytics.

5.5 Conclusion

This chapter presented the evaluation of the second and the creation of the third version of the evaluation framework for learning analytics. We first used several quantitative analyses, i.e. descriptive statistics as well as principal component and reliability analysis of data collected during the evaluation of a widget in an online course, to get an idea of which EFLA-2 items might be problematic and need further improvement. The second step of our study was based on the detailed input from a focus group that discussed every item of the framework individually. On the basis of the quantitative as well as the qualitative analysis results we were able to create the EFLA-3.

The next step will now consist of turning the third evaluation framework for learning analytics into a concrete evaluation instrument, i.e. an online questionnaire, that will be distributed among the students and teachers of a MOOC platform. We will implement widgets for the platform's learning analytics dashboard and use the EFLA-3 to evaluate the widgets. Once the widgets have been evaluated we will again use the collected data to evaluate the framework. This time we will perform validity and reliability analysis to determine whether the dimensions of the EFLA-3 validly represent the underlying components and whether the items reliably measure their underlying dimension.

Part III

Closing Time – Input from the data

Chapter 6

How's It Gonna End – Or: The proof of the pudding

After showing in Chapter 4 that the evaluation framework for learning analytics can be used to evaluate a learning analytics application at several points in time and to reflect differences between the two stakeholder groups, this chapter now explores whether the EFLA can also be used to measure changes in perception of the users between different versions of widgets as well as differences between the two stakeholder groups. In order to do so we have picked two widgets from a MOOC platform's learning analytics dashboard and developed and implemented new versions of those widgets. Additionally, to statistically validate the structure of the EFLA, principal component analysis is used to determine the framework's underlying factors, followed by a reliability analysis of the components' scales. After the elimination of two items, the second round of analysis confirms the assumptions gathered after the first round and the EFLA-4 is created.

This chapter is based on:

Scheffel, M., Drachsler, H., Toisoul, C., Ternier S., and Specht, M. (2017). The Proof of the Pudding: Examining Validity and Reliability of the Evaluation Framework for Learning Analytics. In *Proceedings of the 12th European Conference on Technology Enhanced Learning* (EC-TEL 2017), Berlin, Heidelberg. Springer.

6.1 Introduction

By using learning analytics (LA), i.e. by measuring, collecting, analysing and reporting the learners' data from a course in a useful and meaningful way, awareness and reflection about the learning and teaching processes can be stimulated and students at risk of dropping out can be identified (Long and Siemens, 2011; Reyes, 2015). During the last few years the amount of learning analytics-related research, publications and events has increased steadily (Gašević et al., 2015a). Learning analytics, however, is not to be seen as pure 'number-crunching' on a strictly institutional level or as only being used to improve retention. Instead, it is about creating a holistic view on all learning and teaching processes involved (Gašević et al., 2015b). Therefore, as learning analytics should stimulate the self-regulating skills of the learners (Persico and Pozzi, 2014) and foster awareness and reflection processes for learners and teachers, it is recognised that a good way to present learning analytics to users is through a visual representation (Few, 2006; Verbert et al., 2014; Tempelaar et al., 2015; Bull et al., 2016). Kim et al. (2016) indicate that learners' achievement could be increased by allowing them access to a learning analytics dashboard, i.e. a collection of visualisations. They also point out that learning analytics visualisations should be carefully designed if interest in and usage of the dashboard and analytics is to be maintained by the main stakeholders, i.e. learners and teachers.

With the need for empirical studies growing and more and more discussions about the effect of learning analytics coming up (Siemens et al., 2013; Ferguson and Clow, 2016), a number of studies investigating the impact of learning analytics dashboards have been published in the last few years. Lonn et al. (2015) for example have shown that seeing their academic performance in a learning analytics applications could affect students' interpretation of their data and their success. They stress that learning analytics interventions need to be designed carefully with student goal perception in mind. Beheshitha et al. (2016) randomly assigned learning analytics visualisations to students of a blended learning course and showed that it depended on the students' achievement goal orientation whether the effect of the visualisations on the learning progress was positive or negative. They stress that students' achievement goal orientation and other individual differences need to be taken into account during the learning analytics design process. Finally, Khan and Pardo (2016) showed that depending on the students' information needs and the types of learning activities different kinds of learning analytics dashboards and visualisations are needed for them to be effective. From all three studies it is thus clear that learning analytics visualisations need to be embedded into the instructional design to have a positive effect.

An important aspect that thus needs to be kept in mind when using learning analytics to address issues such as the ones mentioned above is the following: How can we make sure that the learning analytics are valid, reliable, understandable and supportive for the involved stakeholders? We have thus developed an evaluation instrument that allows a standardised approach to the evaluation of learning analytics tools: the evaluation framework for learning analytics (EFLA). The framework consists of four dimensions (*Data*, *Awareness*, *Reflection*, *Impact*) for learners and teachers.

Taking all of this into account, we designed and developed new versions for two widgets from the learning analytics dashboard of the ECO MOOC platform and investigated in a lab experiment whether the current structure of the EFLA appropriately reflects the questionnaire's underlying components and whether the evaluation instrument can be used to measure changes between different versions of widgets. The lab setting was chosen as low numbers of teachers in the ECO environment would not give us sufficient input from that stakeholder group and because it allowed for a controlled experimental setting. We conducted our study with the following research questions in mind:

(RQ-A) Are the changes in the widgets reflected in the EFLA ratings of those widgets, i.e. can the EFLA measure differences between iterations of a widget?

(RQ-B1) Do the four current EFLA dimensions validly represent the underlying structure?

(RQ-B2) Do the items within the dimensions reliably measure the underlying component?

The next section describes the ECO platform's widgets and elaborates on the method of analysis. After the presentation of the results, the discussion section sets the results in relation to the research questions while the final section concludes the chapter.

6.2 Method

6.2.1 Participants

Fifteen PhD candidates (eight women and seven men) and fifteen assistant, associate or full professors (seven women and eight men) from the Faculty of Psychology and Education of the Open University of the Netherlands voluntarily participated in the experiment. The PhD candidates were assigned the role of students while the post-docs were assigned the role of teachers during the experiment. All participants had at least basic knowledge about what learning analytics is. Informed consent was obtained from all participants.

6.2.2 Materials

The ECO Platform and the Learning Analytics Widgets

The European project ECO (Elearning Communication Open-Data)[1] has created a platform that gives free access to MOOCs based on open educational resources and caters to different MOOC providers (Brouns et al., 2014). A learning analytics

[1] `https://ecolearning.eu`

dashboard containing several visualisations is part of the ECO platform to support the ECO users. All users of the portal, i.e. the students as well as the teachers of the MOOCs, see the same visualisations. The visualisations are based on interaction data of the users with the platform and with the MOOCs, e.g. launching a course, accessing pages, watching videos, posting in a forum, uploading homework, etc. While the MOOC providers have their own way of logging the users' activities, each of them additionally provides the logged data as xAPI[2] statements which are then stored in a learning record store (Berg et al., 2016a). All statements used within the ECO project are publicly available via the Dutch xAPI Specification for Learning Activities (DSLA) Registry[3] (Berg et al., 2016b).

One advantage of massive open online courses (MOOCs) is that a diverse public can be provided with education about a vast amount of topics at a low cost (Drachsler and Kalz, 2016). There are, however, still a number of differences in comparison to other, more traditional online courses, e.g. that MOOCs are often not as well-structured and well-prepared as traditional courses. Although many MOOCs are accessible for free, some are commercialised and licensed under different conditions (Miyazoe and Anderson, 2013). Many of the issues and challenges that still need to be solved for MOOCs are related to instructional design, effective implementation, maintenance and enhancement, quality assurance and avoidance of high dropout rates (Brahimi and Sarirete, 2015; Saadatdoost et al., 2015).

It has been shown that learning as well as teaching processes in a MOOC can be supported by providing a visualisation that categorises the learners according to their engagement in the course (Coffrin et al., 2014; MacNeill et al., 2014; Morgado et al., 2014). Cobo et al. (2014) list four categories that can be used to cluster learners based on their activities, e.g. logins, page clicks, forum posts, downloads of and access to learning material etc., within the MOOC: (1) very active/collaborative, (2) active, (3) passive and (4) inactive learners. Such a categorisation could for example be used to identify students that are at risk of dropping out or to provide different types of feedback to different groups of learners.

Two of the existing ECO learning analytics visualisations were chosen for the experiment: the Activity Widget and the Resources Widget. The Activity Widget (see Figure 6.1) shows how active the learners are in a MOOC according to the number of actions done in that MOOC (position of the users on the x-axis vs. number of actions on the y-axis). By hovering over the graph, the exact position and number of activities is shown for that specific user. The Resources Widget (see Figure 6.2) shows what types of resources are present in this course and how often the various resources in the MOOC were accessed by all the users. The length of the bar indicates the frequency of accesses which is also given as a number at the end of the bar. Again, students as well as teachers of a MOOC see the same visualisation.

The second version of the Activity Widget (see Figure 6.3) again shows the total

[2] `https://experienceapi.com`
[3] `http://bit.ly/DutchXAPIreg`

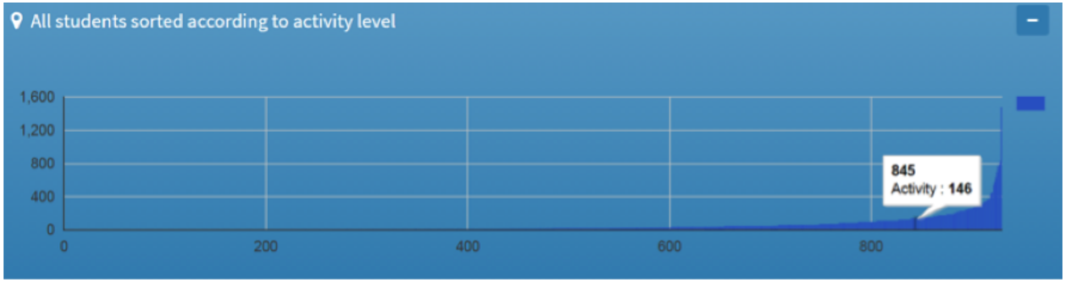


Figure 6.1 Original versions of the Activity Widget

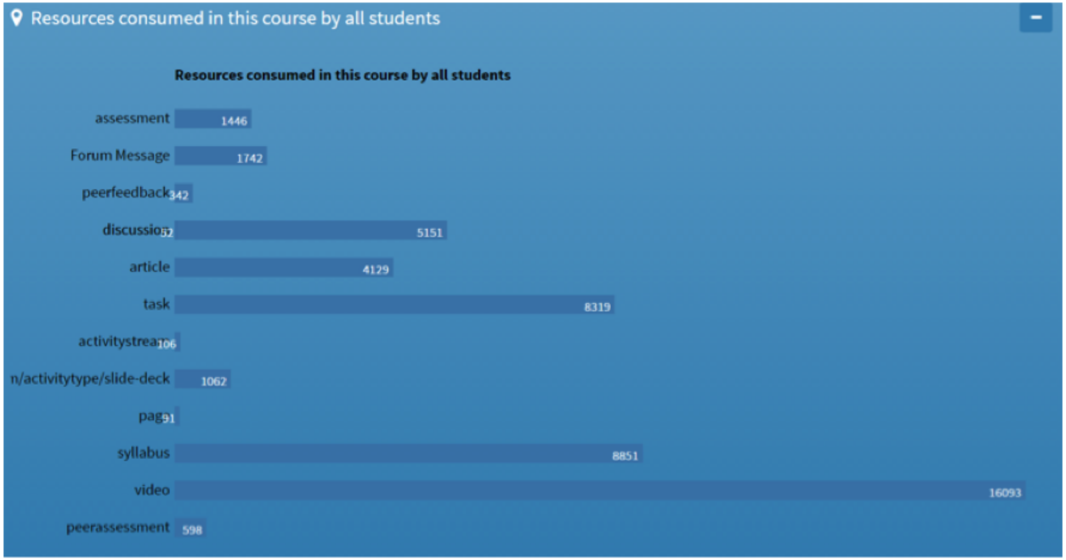


Figure 6.2 Original version of the Resources Widget

Table 6.1 Dimensions and items of the learner and the teacher section of the third version of the evaluation framework for learning analytics (EFLA-3)

		EFLA-3 items for learners/teachers
Data:	**D1**	For this LA tool it is clear what data is being collected.
	D2	For this LA tool it is clear why the data is being collected.
	D3	For this LA tool it is clear who has access to the data.
Awareness:	**A1**	This LA tool makes me aware of my/my students' current learning situation.
	A2	This LA tool makes me forecast my/my students' possible future learning situation given my/their (un)changed behaviour.
Reflection:	**R1**	This LA tool stimulates me to reflect on my past learning/teaching behaviour.
	R2	This LA tool stimulates me to adapt my learning/teaching behaviour if necessary.
Impact:	**I1**	This LA tool increases my motivation to study/teach.
	I2	This LA tool stimulates me to study/teach more efficiently.
	I3	This LA tool stimulates me to study/teach more effectively.

activity per user (position of the users on the x-axis vs. number of actions on the y-axis). Additionally, a user's own position is highlighted in red. With the radio buttons users can choose the type of clustering used in the visualisation. They can choose between the Median with quartiles and an artificial intelligence algorithm that both create four clusters in reference to the four activity types by Cobo et al. (2014). Users can use the information buttons to get more information. In order to protect the users' privacy, none of the users are able to identify who the other users in the visualisation are as the ECO learning analytics dashboard does not distinguish between students and teachers of the course.

The updated version of the Resources Widget (see Figure 6.4) compares a user's sequence of activites, i.e. their MOOC path (white line) with the ideal path through the course (black line) and the paths of other participants. On the x-axis a user can see which activities have been accessed (green) and which have not (red). The icons are indicating the type of activity. The y-axis refers to the weeks of the course. To load the paths of the other participants users can click the "Load all the other student paths" button. With the slider they can filter more or less students, where the more active students are positioned right on the slider. They can use the zoom buttons to zoom in or out on the graph. Teachers could use this tool to identify if learners are using the MOOC as planned by discovering if activities are accessed too early, too late, or not at all. Students could compare themselves to other students and to the model line. All users can use the information buttons to get more information. Students as well as teachers of a MOOC see the same visualisation. Again, in order to protect the users' privacy, none of the other users are identifiable.

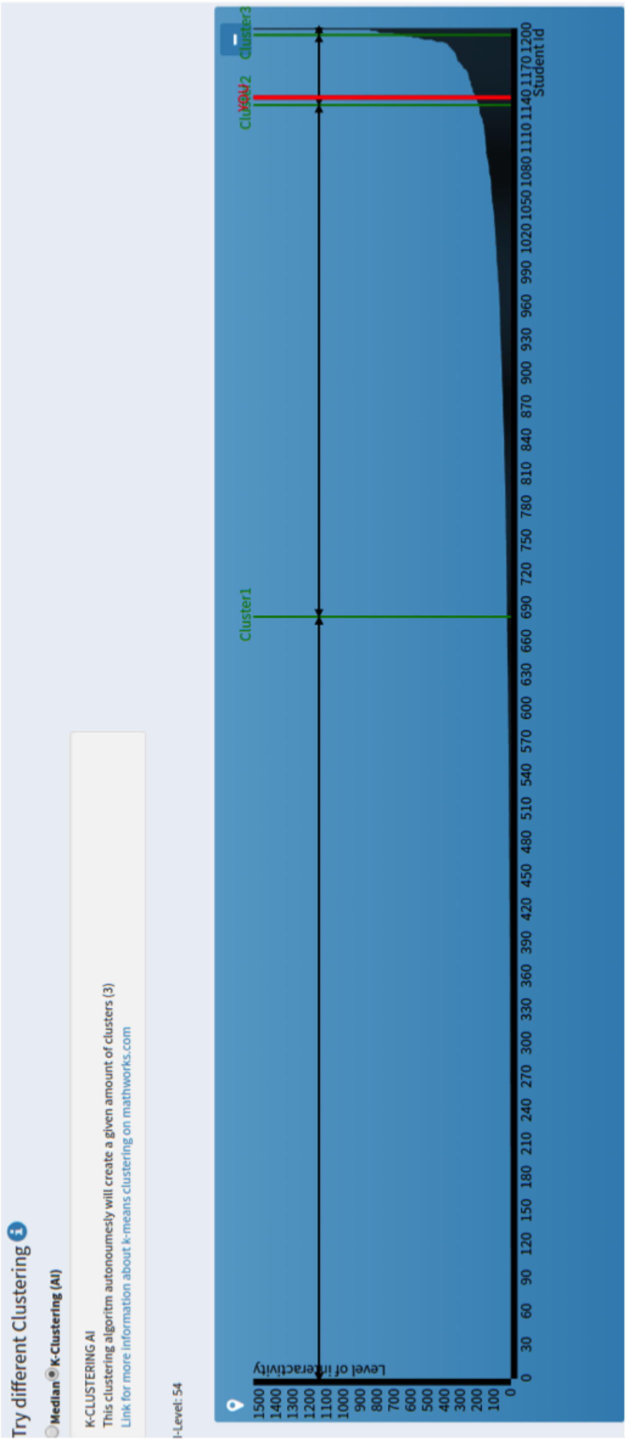


Figure 6.3 Updated versions of the Activity Widget

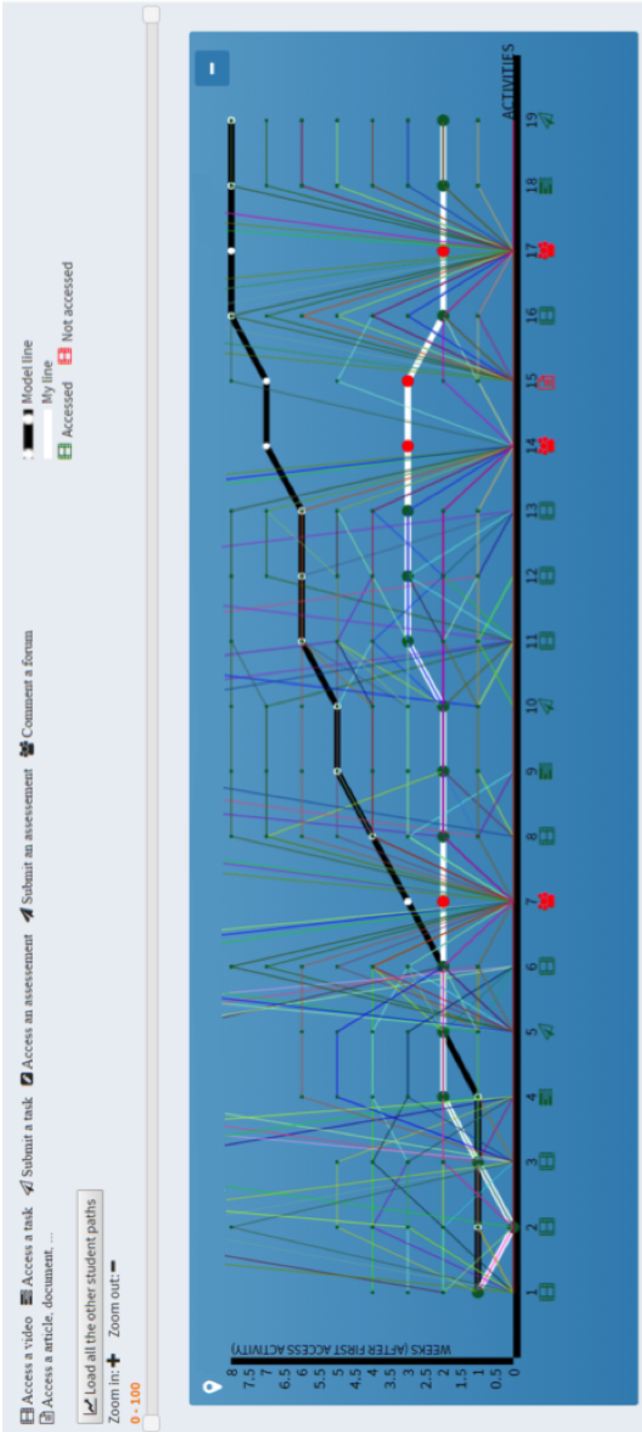


Figure 6.4 Updated version of the Resources Widget

The Evaluation Framework

An institution's need for reflection on how ready they are to implement learning analytics solutions is addressed by the Learning Analytics Readiness Instrument (LARI) (Arnold et al., 2014a). While the LARI has been proven to be an effective instrument to evaluate institutional readiness, there is no standardised instrument so far to evaluate the learning analytics tools once implemented. However, more and more learning analytics tools are being designed, developed and implemented. In order to close this gap, we have therefore developed the evaluation framework for learning analytics (EFLA).

The first version of the evaluation framework for learning analytics (EFLA-1) was constructed through a group concept mapping study with experts from the learning analytics community and consisted of five dimensions (*Objectives*, *Learning Support*, *Learning Measures and Output*, *Data Aspects* and *Organisational Aspects*) with four items each (Scheffel et al., 2014)[4]. After a small evaluation study with learning analytics experts (Scheffel et al., 2015)[5] as well as a revisit of the GCM data and a thorough look at related literature, the second EFLA version was developed. Split into two parts, one for learners and one for teachers, the framework now consisted of four dimensions (*Data*, *Awareness*, *Reflection* and *Impact*) with three items each. This version was turned into an applicable tool, i.e. a questionnaire for students and teachers, and then used in an online course (Scheffel et al., 2017b)[6]. Based on a subsequent evaluation of the EFLA-2, the third version was created. While the dimensions stayed the same, the items were slightly reduced and further refined. Table 6.1 shows version 3 of the EFLA. All items are rated on a scale from 1 for no agreement to 10 for high agreement.

6.2.3 Procedure

All participants were invited to an individual face-to-face session for the experiment. At the beginning of each session, every participant received an introduction to the experiment and was asked to give their informed consent to take part in the study. Following an experimental script (the full script is given in Appendix F), each participant first received some introductory information about the ECO platform and its learning analytics dashboard before getting detailed explanations about the four learning analytics widgets while being shown a screenshot of the corresponding widget. For the two updated widget versions a live demo was also provided.

After each widget explanation, participants were asked to evaluate the widget using the EFLA-3 while assuming either the role of a student (all PhD candidates) or a teacher (all post docs). At the end of each EFLA survey participants had the option to add comments to the survey. When all four widgets had been evaluated, participants were asked to supply some demographic information (gender and age range) and

[4] This publication is included as **Chapter 1** in this thesis.
[5] This publication is included as **Chapter 2** in this thesis.
[6] This publication is included as **Chapter 4** in this thesis.

were given a final opportunity to enter comments about the experiment.

Once all data was collected from the participants, several statistical analyses were calculated using IBM's SPSS Statistics and graphs showing the average evaluation of each EFLA item for the different widgets from both stakeholder groups were created. The statistical analyses included t-tests for the widget evaluation and principal component analysis as well as reliability analysis for the EFLA evaluation.

6.3 Results

6.3.1 Widget Evaluation

Figure 6.5 shows the average scores of the ten EFLA-3 items from students and teachers for both versions of the widgets. On average students and teachers gave better ratings to the second versions of both widgets. The only item students rated lower in an updated widget version is D1 for the Resources Widget. The items that teachers rated lower in an updated widget version are D3 and R2 for the Activity Widget and also D1 for the Resources Widget. While the original versions of the widgets received higher ratings from the teachers, the updated widget versions received higher ratings from the students.

Conducting paired sample t-tests for the ten EFLA-3 items allowed us to see whether the differences in ratings between the two versions of the widgets were significant or not. For the student participants (see Table 6.2) there are several EFLA items where the difference between the ratings of the widgets' two versions is significant. The second version of the Activity Widget received significantly higher ratings for the items A1 ($p = .019$), R1 ($p = .044$), R2 ($p = .008$) and I2 ($p = .022$) while the Resources Widget received significantly higher ratings for all items (p ranges between .000 and .048) except D1. In case of the teachers (see Table 6.3), each widget only has one item where the difference between the two versions is significant: for item I2 of the Activity Widget $t(14) = -2.942, p = .011$ and for item A2 of the Resources Widget $t(14) = -2.839, p = .013$.

6.3.2 EFLA Evaluation

Every participant completed the EFLA survey for both versions of the two learning analytics widgets which gives us a total N of 120 for each EFLA item (60 per stakeholder group, 30 per widget, 15 per widget version). All statistical analyses were conducted separately for the students' and teachers' data due to the different semantics, i.e. different wording leading to different meaning, of the ten EFLA items. The highest N within one analysis is thus 60. Table 6.4 shows the descriptive statistics, i.e. N, minimum value, maximum value, mean, standard deviation and variance, for all ten EFLA items for the students (left) and the teachers (right). Two values seem to be slightly different from the rest: the variance of EFLA item D3 for students as well as for teachers is noticeably higher than all other variance values.

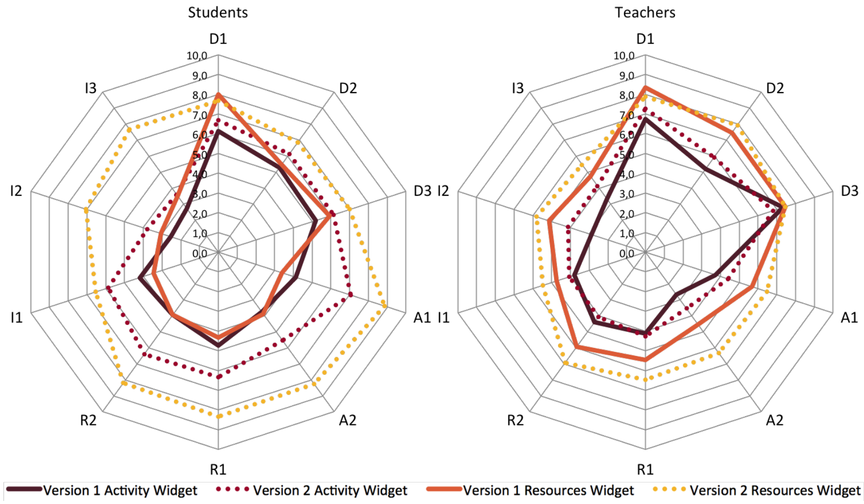


Figure 6.5 Average scores of the EFLA-3 items for students (left) and teachers (right) for both versions of both widgets

First Analysis

Before conducting the principal component analysis (PCA) we first looked at the factorability of the ten EFLA items for students and teachers. For the students' EFLA only few correlations were below .3 and all ten items correlated at least .6 with at least two other items. Additionally, the Kaiser-Meyer-Olkin measure of sampling adequacy was .836, i.e above the recommended value of .6, and Bartlett's test of sphericity was $\chi^2(45) = 462.515, p < .000$. All diagonals of the anti-image correlation matrix were above .7.

For the teachers' EFLA there were also few correlations below .3 and nine items correlated at least .4 with at least two other items (only D3 did not). Additionally, the Kaiser-Meyer-Olkin measure of sampling adequacy was .848, i.e above the recommended value of .6, and Bartlett's test of sphericity was $\chi^2(45) = 405.841, p < .000$. Nine diagonals of the anti-image correlation matrix were above .7 (except D3 where it was .486).

Due to these results, none of the items were discarded at this point and we continued with the PCA using Varimax rotation in order to identify the factors underlying the EFLA. As we had structured the EFLA with four dimensions in mind (data, awareness, reflection, impact), the solution with four components was examined first, followed by those with three and with two components. Table 6.5 shows the results of the PCA for these different settings.

First Principal Component Analysis – Students. For the students' four-components solution all communalities were above .8 except I1 which was .749. Together the four components explained 85.824% of the variance (80.805 for the three components

Table 6.2 Paired sample t-test for the students' EFLA-3 item scores of the two versions for the Activity Widget (left) and the Resources Widget (right), significances are coloured

	Activity Widget					Resources Widget				
	Mean	St.D.	t	df	Sig.	Mean	St.D.	t	df	Sig.
D1	-.533	1.506	-1.372	14	.192	.333	1.447	.892	14	.388
D2	-.867	1.598	-2.101	14	.054	-1.400	2.324	-2.333	14	.035
D3	-.933	2.017	-1.793	14	.095	-1.067	1.907	-2.166	14	.048
A1	-2.933	4.267	-2.662	14	.019	-5.467	2.167	-9.771	14	.000
A2	-1.800	3.489	-1.998	14	.065	-4.400	2.293	-7.432	14	.000
R1	-1.600	2.798	-2.215	14	.044	-4.000	2.420	-6.401	14	.000
R2	-2.467	3.067	-3.114	14	.008	-4.267	2.374	-6.959	14	.000
I1	-1.667	3.374	-1.913	14	.076	-3.067	3.515	-3.379	14	.004
I2	-1.267	1.907	-2.572	14	.022	-4.000	2.535	-6.110	14	.000
I3	-.867	1.922	-1.746	14	.103	-4.133	2.475	-6.469	14	.000

Table 6.3 Paired sample t-test for the teachers' EFLA-3 item scores of the two versions for the Activity Widget (left) and the Resources Widget (right), significances are coloured

	Activity Widget					Resources Widget				
	Mean	St.D.	t	df	Sig.	Mean	St.D.	t	df	Sig.
D1	-.533	2.356	-.877	14	.395	.467	1.552	1.164	14	.264
D2	-.733	2.604	-1.091	14	.294	-.467	1.685	-1.073	14	.301
D3	.533	2.446	.845	14	.413	.000	1.964	.000	14	1.000
A1	-.667	2.350	-1.099	14	.290	-.800	2.111	-1.468	14	.164
A2	-.867	2.134	-1.573	14	.138	-1.800	2.455	-2.839	14	.013
R1	-.133	2.264	-.228	14	.823	-1.000	2.171	-1.784	14	.096
R2	.333	1.543	.837	14	.417	-1.000	2.420	-1.600	14	.132
I1	-.267	1.944	-.531	14	.604	-.733	1.710	-1.661	14	.119
I2	-1.267	1.668	-2.942	14	.011	-.667	1.839	-1.404	14	.182
I3	-.800	1.656	-1.871	14	.082	-.733	1.831	-1.551	14	.143

Table 6.4 Descriptive statistics of all EFLA-3 items from all widgets combined for students (left) and teachers (right)

	students						teachers					
	N	Min.	Max.	Mean	St.D.	Var.	N	Min.	Max.	Mean	St.D.	Var.
D1	60	1	10	7.12	2.450	6.003	60	3	10	7.55	1.908	3.642
D2	60	1	10	5.93	2.968	8.809	60	2	10	6.63	2.091	4.372
D3	60	1	10	6.07	3.194	10.199	60	2	10	7.27	3.162	9.995
A1	60	1	10	5.87	3.105	9.643	60	1	10	5.07	2.642	6.979
A2	60	1	10	5.35	2.839	8.062	60	1	9	4.27	2.421	5.860
R1	60	1	10	5.93	2.711	7.351	60	1	10	5.08	2.438	5.942
R2	60	1	10	5.62	2.853	8.139	60	1	9	5.33	2.319	5.379
I1	60	1	10	5.02	2.902	8.423	60	1	8	4.52	2.259	5.101
I2	60	1	10	4.12	2.811	7.901	60	1	10	4.48	2.411	5.813
I3	60	1	10	4.38	2.946	8.681	60	1	9	4.42	2.309	5.332

Table 6.5 First principal component analysis using Varimax rotation for four, three and two components for students' EFLA-3 data (primary loadings are yellow) and teachers' EFLA-3 data (primary loadings are red)

	four components								three components						two components			
	students				teachers				students			teachers			students		teachers	
	1	2	3	4	1	2	3	4	1	2	3	1	2	3	1	2	1	2
D1	.004	.029	.788	.553	.247	.155	.144	.919	-.012	.063	.864	.200	.799	-.150	.026	.866	.486	.503
D2	.126	.161	.907	-.206	.354	.737	.355	.124	.145	.136	.863	.428	.636	-.038	.191	.863	.636	.348
D3	.135	.162	.875	.103	-.069	-.057	.914	.099	.138	.164	.879	-.043	.235	-.863	.204	.881	-.187	.871
A1	.380	.807	.223	-.026	.226	.721	-.357	.400	.358	.812	.214	.249	.677	.580	.814	.246	.689	-.201
A2	.329	.839	.139	-.067	.503	.673	-.345	-.035	.306	.840	.125	.551	.324	.586	.796	.161	.784	-.338
R1	.505	.703	.066	.347	.878	.277	-.084	.076	.466	.746	.116	.888	.200	.176	.848	.140	.889	.001
R2	.636	.634	.146	.062	.855	.349	-.007	.147	.613	.658	.150	.871	.316	.130	.896	.162	.913	.092
I1	.790	.337	-.038	.101	.870	.076	-.032	.124	.772	.373	-.028	.861	.117	.044	.819	-.038	.789	.081
I2	.879	.306	.174	-.034	.860	.300	.048	.157	.872	.332	.160	.873	.304	.058	.863	.143	.888	.151
I3	.856	.349	.190	-.047	.737	.268	-.166	.274	.848	.373	.174	.730	.324	.231	.873	.160	.829	-.005

Table 6.6 First reliability statistics and scale statistics of different item groups for students' EFLA-3 (left) and teachers' EFLA-3 (right)

	students					teachers				
items	N	Cron.α	Mean	Var.	St.D.	N	Cron.α	Mean	Var.	St.D.
D	3	.855	19.12	58.173	7.627	3	[2].397	21.45	24.489	4.949
A	2	.852	11.22	30.851	5.554	2	.814	9.33	21.650	4.653
R	2	.890	11.55	27.913	5.283	2	.945	10.42	21.468	4.633
I	3	[1].902	13.52	62.762	7.922	3	.878	13.42	39.196	6.261
A+R	4	.916	22.77	105.945	10.293	4	.870	19.75	69.513	8.337
A+R+I	7	.939	36.28	297.698	17.254	7	.924	33.17	194.277	13.938
R+I	5	.928	25.07	157.385	12.545	5	.938	23.83	110.277	10.501

[1] could be increased to .954 if eliminating I1
[2] could be increased to .574 if eliminating D3

with primary loadings). All items in the four-components solution (rotated matrix) had a primary loading of .6 or above. However, only three of the four components contained primary loads. Component 1 was clearly formed by items I1, I2 and I3, component 2 consisted of items A1, A2 and R1 and component 3 was clearly formed by items D1, D2 and D3. Item R2 had two possible primary loads (.636 and .634) and could be part of either component 1 or component 2. Looking at the three-components solution for the students' data, the communalities were all above .736. The three components cumulatively explained 81.427% of the variance. Also, the distinction between the components was clearer than in the four-components solution: component 1 contained items I1, I2 and I3, component 2 contained items A1, A2, R1 and R3 and component 3 contained items D1, D2 and D3. Again all items had a primary loading of .6 or above. The two-components solution for the students' data had communality values above .7 except for A2 (.660) and I1 (.672). Cumulatively the two components explained 75.238% of the variance. This solution had primary loadings for nine items above .8 and one item at .796 with component 1 containing A1, A2, R1, R2, I1, I2 and I3 and component 2 containing the items D1, D2 and D3.

To sum up, the three-components solution seems to be the best result as all components contain primary loads (the four component solution does not) and as it explains more variance than the two-components solution.

First Principal Component Analysis – Teachers. The PCA of the teachers' data provided somewhat less clearly structured solutions. In the four-components solution all communalities were above .7. Together the four components explained 83.866% of the variance. All items had a primary loading of at least .6. Component 1 contained items R1, R2, I1, I2 and I3, while component 2 contained items D2, A1 and A2. Items D1 and D3 each formed their own component. The data items thus did not form one component but are spread over three different ones. The three-components solution for the teachers' data had communality values of at least .7 for all values except for D2 (.589) and I3 (691). Cumulatively 77.409% of variance were explained by the three components. This solution had one clear component containing items R1, R2, I1, I2 and I3 with all primary loadings above .7. D1, D2 and A1 formed one component, as did D3 and A2, all with primary loadings above .5. Both A1 and A2, however, had rather high cross-loads: while A1 had a primary load of .677 in component 2 (together with D1 and D2) it had a cross-load of .580 for component 3 (where it would join A2 and D3). A2 (primary load of .586) on the other hand also had a high cross-load of .551 in factor 1 (where it would join R1, R2, I1, I2 and I3). Finally, in the two-components solution for the teachers' data, the communalities were above .6 except for D1 (.489), D2 (.526) and A1 (515). The two components explained 68.146% of the variance. Component 1 contained D2, A1, A2, R1, R2, I1, I2 and I3 (all with primary loads above .6), while the second component was comprised of items D1 and D3. Again, the data items did not form one clear component. Item D1 (primary load of .503 in factor 2), however, had a rather high cross-load of .486 in component 1 and could thus possibly be positioned there leaving D3 to form its own component.

To sum up, the three-components solution seems to be the best result as all components have at least two primary loads (the four-components solution does not) and as it explains more variance than the two-components solution.

First Reliability Analysis. In order to see how reliable the scales are and to check whether any of the items should be excluded, we calculated the reliability values, i.e. Cronbach's Alpha, for several item combinations based on the principal component analysis results (see Table 6.6): the four EFLA dimensions data, awareness, reflection and impact individually (D,A,R,I), the combination of the awareness and reflection items (A+R), the combination of the awareness, reflection and impact items (A+R+I), and the combination of the reflection and impact items (R+I). Only one scale, i.e. the teachers' three data items on their own, received a low reliability score (.397). All other scales had a reliability score of .8 or higher. For two scales a substantial increase ($> .05$) in Cronbach's Alpha could be achieved by eliminating an item. For the students' EFLA eliminating item I1 in the impact items only scale would result in an alpha of .954 while an elimination of item D3 in the data items only scale of the teachers' EFLA would result in a Cronbach's Alpha of .574.

As the items D3 and I1 seemed to cause problems and hindered a clear component solution, we decided to remove them from the framework and to re-do the analysis with the remaining eight items D1, D2, A1, A2, R1, R2, I2 and I3.

Second Analysis

Before doing the principal component analysis, we again looked at the factorability of the EFLA items. For the students' data there were again few correlations between the items that were below .3 and all items correlated at least .6 with at least one other item. The Kaiser-Meyer-Olkin measure of sampling adequacy was .799 (which is above the recommended value of .6) and Bartlett's test of sphericity was $\chi^2(28) = 359.650, p < .000$. All diagonals of the anti-image correlation matrix were above .7 (except for D1 which was .526).

The teachers' data also showed few correlations below .3 and, except for D1 and D2 which correlated at .4 with three other items, all other items correlated at .6 with at least one other item. Additionally, the Kaiser-Meyer-Olkin measure of sampling adequacy was .826 and Bartlett's test of sphericity was $\chi^2(28) = 338.879, p < .000$. All diagonals of the anti-image correlation matrix were above .7.

Second Principal Component Analysis – Students. Table 6.7 shows the results of the PCA using Varimax rotation for these different settings. For the students' four-component solution all communalities were above .8. Together the four components explained 89.975% of the variance. All items in the four-components solution had a primary loading of .7 or above. Component 1 was clearly formed by items A1, A2, R1 and R2, component 2 consisted of items I2 and I3, component 3 only contained D1 and component 4 only contained D2. Looking at the three-components solution for the students' data, the communalities were all above .793. The three components cumulatively explained 84.559% of the variance. Again, component 1 was clearly formed by items A1, A2, R1 and R2 and component 2 consisted of items I2 and I3. Component 3 was made up of D1 and D2. All primary loadings were above .7. The two-components solution for the students' data had communality values above .7 except for A2 (.691). Cumulatively the two components explained 77.195% of the variance. This solution had primary loadings for all items above .8 with component 1 containing A1, A2, R1, R2, I2 and I3 and component 2 containing the items D1 and D2.

To sum up, the three-components solution seems to be the best result as all components have at least two primary loads (the four-components solution does not) and as it explains more variance than the two-components solution.

Second Principal Component Analysis – Teachers. The principal component analysis of the teachers' data provided the following results. In the four-components solution all communalities were above .792. Together the four components explained 89.644% of the variance. All items had a primary loading of at least .7. Component 1 contained items R1, R2, I2 and I3, while component 2 contained items A1 and A2. Items D1 and D2 each formed their own component. The three-components solution for the teachers' data had communality values of at least .7 for all items except for D2 (.547). Cumulatively 82.201% of variance were explained by the three components. This solution had one clear component containing items R1, R2, I2 and I3 with all primary loadings above .7. A1 and A2 formed component 2, and D1 and D2 formed component 3, all with primary loadings above .7 except for D2 (.524). Finally, in the two-components solution for the teachers' data, the communalities were either just below or well above .7 except for D2 (.545) and A1 (.517). The two components explained 72.445% of the variance. Component 1 contained items A1, A2, R1, R2, I2 and I3, all with primary loads above .7 except for A1 (.566), while the second component was comprised of items D1 (.940) and D2 (.572).

To sum up, the three-components solution seems to be the best result as all components have at least two primary loads (the four-components solution does not) and as it explains more variance than the two-components solution.

Second Reliability Analysis. Again, we calculated reliability values, i.e. Cronbach's Alpha, for several item combinations: the four EFLA dimensions data, awareness, reflection and impact individually (D,A,R,I), the combination of the awareness and reflection items (A+R), the combination of the awareness, reflection and impact items (A+R+I), and the combination of the reflection and impact items (R+I). Table 6.8 gives an overview of these analyses for the students' as well as the teachers' EFLA. Only one scale, i.e. the teachers' data items on their own, receives a noticeably lower reliability score (.574). All other scales have a reliability score of .7 or higher. For none of the scales a substantial increase ($> .05$) in Cronbach's Alpha could be achieved by eliminating an item.

Table 6.7 Second principal component analysis using Varimax rotation for four, three and two components for students' EFLA-3 data (primary loads are yellow) and teachers' EFLA-3 data (primary loads are red)

	four components								three components						two components			
	students				teachers				students			teachers			students		teachers	
	1	2	3	4	1	2	3	4	1	2	3	1	2	3	1	2	1	2
D1	.070	.048	.936	.301	.210	.120	.934	.176	.077	.008	.903	.171	.100	.928	.054	.904	.103	.940
D2	.162	.151	.377	.878	.292	.205	.199	.895	.156	.184	.862	.451	.262	.524	.226	.864	.466	.572
A1	.840	.300	.049	.220	.220	.881	.297	.135	.845	.289	.184	.202	.878	.329	.839	.197	.566	.443
A2	.849	.266	-.013	.157	.481	.767	-.120	.216	.853	.254	.096	.506	.786	-.029	.824	.109	.823	.091
R1	.780	.436	.246	-.153	.874	.218	.067	.240	.798	.386	.087	.893	.236	.144	.863	.100	.876	.213
R2	.717	.540	.063	.123	.834	.226	.152	.348	.731	.523	.131	.869	.247	.265	.896	.142	.849	.333
I2	.380	.891	.052	.112	.869	.230	.167	.199	.405	.881	.115	.872	.241	.221	.864	.121	.852	.289
I3	.419	.863	.049	.126	.785	.339	.246	-.001	.443	.853	.122	.741	.332	.220	.876	.129	.783	.293

Table 6.8 Second reliability statistics and scale statistics of different item groups for students' EFLA-3 (left) and teachers' EFLA-3 (right)

	students					teachers				
items	N	Cron.α	Mean	Var.	St.D.	N	Cron.α	Mean	Var.	St.D.
D	2	.745	13.05	23.608	4.859	2	.574	14.18	11.237	3.352
A	2	.852	11.22	30.851	5.554	2	.814	9.33	21.650	4.653
R	2	.890	11.55	27.913	5.283	2	.945	10.42	21.468	4.633
I	2	.954	8.50	31.712	5.631	2	.881	8.90	19.922	4.463
A+R	4	.916	22.77	105.945	10.293	4	.870	19.75	69.513	8.337
A+R+I	6	.936	31.27	226.029	15.034	6	.916	28.65	149.214	12.215
R+I	4	.925	20.05	104.794	10.237	4	.935	19.32	75.135	8.668

6.4 Discussion

6.4.1 Widget Evaluation

The evaluation of the widgets using the EFLA-3 questionnaire shows that there are indeed significant differences in evaluation results between the different widget versions. Our research question RQ-A can thus be answered with “yes". However, the differences are not significant for all items of all widgets from both stakeholders. Students really seemed to appreciate the second versions of the widgets much more than the first versions. Especially the Resources Widget received significantly higher evaluation results for its second version.

Taking into account the open comments from the questionnaire as well as the questions and comments uttered during the experiment by both stakeholder groups, these results are not really surprising. The teacher participants were much more hesitant and held back by the lab setting of the experiment while the student participants could easily put themselves in the mindset of an online course participant. Another factor that is likely to play a role in influencing the teachers' widget evaluations is that due to the ECO platform's not distinguishing between the user types of learners and teachers when displaying the visualisations, the personalisation aspect of the widgets' second versions was rather pointless for the teachers. That is, they might have felt disregarded.

There are several limitations to this part of our study. First, the experiment did not take place in a live environment. The participants were thus not able to actually use the different widgets themselves within a course environment for an extended period of time. Instead, they were only shown images of the widgets and shown a live demo during the experimental session. While this is an important limitation, however, it is at the same time also a beneficial aspect of our study. Due to the laboratory setting, we were able to control that all participants got the same information and answered all questions. Another limitation of our study is that the roles of the participants were not assigned randomly but according to their current level of employment at the university. While a random role assignment might have provided us with more impartial results, we decided to assign the roles as close as possible to a real-life student-teacher setting. A third limitation for our study is that the widgets were shown in the same order to all participants the evaluation results might thus possibly be biased.

6.4.2 EFLA Evaluation

Although none of the items were discarded before conducting the first principal component analysis, the descriptive statistics (variance) as well as the factorability check (correlations and anti-image correlations for the teachers' data) hinted at possible issues with item D3. We began the first PCA assuming that the EFLA consisted of four distinct dimensions. For the students' data, however, only three components had primary loadings in the four-components solution thus indicating that there are only three underlying components to the EFLA. This was also supported by the other two solutions (the variance explained was higher for the three-components solution compared to the two-components solution).

The first analysis of the teachers' data also showed that a four-components solution did not best represent the data. It also became apparent that D1, D2 and D3 and to some extent A1 and A2 seemed to be problematic for the teachers. Their principal component analysis results for those items were much less clear than those of the students. This had already been foreshadowed during the experiment. The teacher participants asked considerably more questions than the student participants and voiced uncertainty about how to answer some of the questions. This insecurity about the items is likely to be reflected in their answers resulting in partially inconclusive principal component analysis results. The students did not seem to have such issues with the items and their results are thus more confident and possibly more credible.

The reliability analysis confirmed that several items might hinder a clear component solution. Two items, D3 and I1, had to be discarded. The fact that it was precisely those two items that were problematic is reasonable if we look at the actual questions behind those items. D3 says "For this LA tool it is clear who has access to the data". In comparison to this item, D1 and D2 much more clearly address the micro level of the immediately involved learners and teachers themselves (Greller and Drachsler, 2012) which is what the EFLA is about. Both of those items are much more connected to the user's personal point of view whereas D3 could be (mis)interpreted so as to cover the whole learning environment instead of an individual LA tool despite the statement saying "For this LA tool...". Additionally, in order to interpret a visualisation it is important to know what data it is based on and why (i.e. what the purpose is) but to know who else has access to the data does not affect the interpretation. Instead, it is more an issue of an institution's LA policy than an individual visualisation to make sure that privacy and transparency regulations are in place and transparently communicated.

Already during the experiment, student as well as teacher participants mentioned that they had difficulties answering item I1 due to its generality. The item says "This LA tool increases my motivation to study/teach". Whereas I2 and I3 cover the specific aspects of efficiency and effectiveness, item I1 covers motivation in general. Many participants said that their being motivated by a visualisation very much depended on the contents of the widget. For example, if a student sees that he is the lowest performing student, he might not be motivated to study by such a visualisation, while the opposite might be true if he sees himself in the top-performing group. On other days, the same student might feel very motivated to study when seeing that he is lagging behind. General motivation is thus too context-dependent to receive a reliable rating for one visualisation.

The second principal component analysis without the two discarded items confirmed the previous indication that there are three underlying components for the EFLA items. In this solution each component was loaded by at least two items and explained more of the variance than the two-components solution. There is, however,

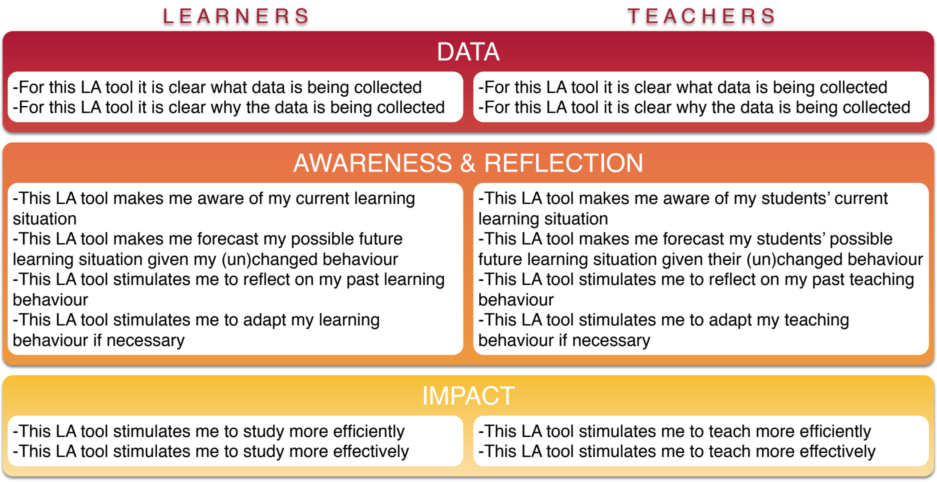


Figure 6.6 Fourth version of the evaluation framework for learning analytics (EFLA-4)

a difference in how the items are spread across the components. For the students' data, D1 and D2 form one component, A1, A2, R1 and R2 form a second one and I2 and I3 form a third. The teachers' data resulted in one component containing D1 and D2, a second one containing A1 and A2 and another one containing R1, R2, I2 and I3. Even though some of the items of the student and teacher EFLA are semantically different, the two EFLA sections are still to be seen as two sides of the same coin.

Thus, in order to decide which of the three-components solutions to use for the next version of the EFLA, we took several aspects into account. First, the teacher participants of our study voiced more insecurities than the student participants did which leads us to put more confidence in the students' results. Second, the reliability results for the students' data showed higher Cronbach's Alpha values than those of the teachers and the explained variance was higher for the students' three component solution. And third, supporting awareness and reflection processes in users in order to impact the learning or teaching processes is an important aim of learning analytics. Awareness and reflection go hand in hand, with the former being a prerequisite of the latter (Butler and Winne, 1995; Endsley, 1995; Schön, 1983).

Based on this, the new version of the EFLA now consists of three dimensions: *Data*, *Awareness & Reflection*, *Impact*. The *Data* dimension contains items D1 and D2 and the *Impact* dimension contains items I2 and I3. Finally, the *Awareness & Reflection* dimension contains the four items A1, A2, R1 and R2. Figure 6.6 shows the dimensions and items for both stakeholders in details.

RQ-B1 thus has to be answered with "no" as the assumed four-components structure did not turn out to be the best solution. However, the three-components solution we settled on does provide a fairly similar EFLA structuring to the one we envisioned as the items were not completely re-arranged within new clusters but two of the original dimensions were combined into one. RQ-B2 also has to be answered with “no” as not all ten EFLA items turned out to reliably measure their component. However, eight of the items did and will thus constitute the new EFLA.

6.5 Conclusion

This chapter presented the results of an empirical lab study where we specifically developed and implemented several widgets for a MOOC platform’s learning analytics dashboard and evaluated them using the Evaluation Framework for Learning Analytics (EFLA). We also evaluated said framework using principal component analysis and reliability analysis. The results of the widget analysis showed that the EFLA can indeed be used to measure differences between different widget iterations. The results of the EFLA analysis show that there are three underlying dimensions in the EFLA instead of four and that not all items in version 3 of the EFLA reliably measured these dimensions.

A new and improved fourth version of the EFLA has thus been created – consisting of the three dimensions *Data*, *Awareness & Reflection*, and *Impact* – that can be used to validly and reliably evaluate learning analytics tools. All items are to be rated on a scale from 1 for ‘strongly disagree’ to 10 for ‘strongly agree’. Figure 6.7 shows the questionnaire templates for the learner and the teacher section of the EFLA together with the scoring instructions. In order to calculate a learning analytics tool’s EFLA score, i.e. a number between 0 and 100, the following steps are needed per stakeholder group: (1) calculate the average value for each item based on the answers given for that item, (2) calculate the average value for each dimension based on the average of its items, (3) calculate the dimensional scores by rounding the result of $((x-1)/9) * 100$ where x is the average value of a dimension, and (4) calculate the overall EFLA score by taking the average of the three dimensional scores.

The framework has been published as open access and is available for everyone interested in evaluating their learning analytics tools. The framework’s template flyer as well as an interactive spreadsheet to automatically calculate the EFLA scores and create visualisations of the scores are available for download via the LACE (Learning Analytics Community Europe) website[7].

[7] `http://www.laceproject.eu/evaluation-framework-for-la/`

The Evaluation Framework for Learning Analytics
EFLA

for LEARNERS

DATA

For this LA tool it is clear what data is being collected

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

For this LA tool it is clear why the data is being collected

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

AWARENESS & REFLECTION

This LA tool makes me aware of my current learning situation

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

This LA tool makes me forecast my possible future learning situation given my (un)changed behaviour

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

This LA tool stimulates me to reflect on my past learning behaviour

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

This LA tool stimulates me to adapt my learning behaviour if necessary

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

IMPACT

This LA tool stimulates me to study more efficiently

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

This LA tool stimulates me to study more effectively

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

for TEACHERS

DATA

For this LA tool it is clear what data is being collected

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

For this LA tool it is clear why the data is being collected

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

AWARENESS & REFLECTION

This LA tool makes me aware of my students' current learning situation

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

This LA tool makes me forecast my students' possible future learning situation given their (un)changed behaviour

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

This LA tool stimulates me to reflect on my past teaching behaviour

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

This LA tool stimulates me to adapt my teaching behaviour if necessary

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

IMPACT

This LA tool stimulates me to teach more efficiently

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

This LA tool stimulates me to teach more effectively

	1	2	3	4	5	6	7	8	9	10	
strongly disagree	O	O	O	O	O	O	O	O	O	O	strongly agree

EFLA scoring per stakeholder group

Step 1: Calculate the average value for each item based on the answers given for that item.

Step 2: Calculate the average value for each dimension based on the average of its items.

Step 3: Calculate the dimensional scores by rounding the results of ((x-1)/9)*100 where x is the average value of a dimension.

Step 4: Calculate the overall EFLA score by taking the average of the three dimensional scores.

Figure 6.7 EFLA questionnaire template with scoring instructions

General Discussion

The research reported in this thesis described the continuous process of iteratively creating, using, evaluating and improving the evaluation framework for learning analytics (EFLA). The framework addresses the current lack of evaluation instruments by offering a standardised way to evaluate learning analytics tools and to measure and compare the impact of learning analytics on educational practices. The main objectives of the research presented in this thesis therefore were threefold: (1) identify quality indicators for learning analytics, (2) create an applicable evaluation instrument based on these indicators, and (3) validate the evaluation instrument.

The thesis approached these objectives in three distinct parts. In Part I, based on the input from the learning analytics community, the results of a group concept mapping study were used to devise the first version of the framework (EFLA-1). Further input from learning analytics experts as well as a literature review was then used to evaluate and revise the framework to create the second version of the framework (EFLA-2). Part II then dealt with the questions whether a learning analytics widget can influence the students' grades in a course and whether the evaluation framework can be used to evaluate such a widget at several points in time and to reflect differences in perception between the two stakeholder groups. The widget evaluation as well as the results from an experts focus group were then used to evaluate the framework and to create the third version of the framework (EFLA-3). Finally, Part III of the thesis showed that the evaluation framework can also be used to measure changes between different versions of widgets as well as differences between the two stakeholder groups and presented the validity and reliability analyses that led to the creation of the fourth and final version of the evaluation framework (EFLA-4).

This concluding chapter first summarises and discusses the main outcomes from these three parts to then address the limitations of the presented research as well as implications for the field and future research.

Main findings

Motivated by the lack of evaluation standards that define quality indicators for and thus enable the evaluation of learning analytics tools, **Chapter 1** presented the first version of the evaluation framework for learning analytics (EFLA-1). The framework was based on the results of a group concept mapping study with experts from the learning analytics field. Based on the prompt *'One specific quality indicator to evaluate the effects of learning analytics is ...'* and the study participants' subsequent ratings of the collected statements according to their importance and feasibility, several topic areas and themes were identified for the construction of the evaluation framework. Ultimately, five dimensions with four items each were used to build EFLA-1 (see Figure 1.9 on page 25):

- ***Objectives***: *Awareness*, *Reflection*, *Motivation*, *Behavioural Change*
- ***Learning Support***: *Perceived Usefulness*, *Recommendation*, *Activity Classification*, *Detection of Students at Risk*
- ***Learning Measures and Output***: *Comparability*, *Effectiveness*, *Efficiency*, *Helpfulness*
- ***Data Aspects***: *Transparency*, *Data Standards*, *Data Ownership*, *Privacy*
- ***Organisational Aspects***: *Availability*, *Implementation*, *Training of Educational Stakeholders*, *Organisational Change*

A focused literature review based on the framework's dimensions was then added to further extend the group concept mapping study with the latest insights from the learning analytics community. This contextualisation was used to better position the framework's dimensions and their relevance for the evaluation of learning analytics tools in the field. The review, for example, covered works about the theoretical background of concepts such as awareness and reflection, presented studies about the design, implementation and usage of learning analytics tools, mentioned possible risks with regards to ethical and privacy issues and took into account strategies and experiences needed to institutionalise learning analytics.

As the first step to turning the evaluation framework for learning analytics into an applicable evaluation instrument, in **Chapter 2** the EFLA-1 was turned into a questionnaire and applied to a number of learning analytics tools. In doing so the learning analytics experts participating in this study were able to identify problematic issues with the framework that needed to be addressed to improve the framework (see Figure 2.1 on page 41). The most critical requirements identified by the quantitative and qualitative evaluation results were:

- the framework and its questionnaire need to be more condense;
- dimension titles and item names need to be clear and easy to understand;
- there need to be different questionnaires for the different user types;
- the questionnaire needs to be answered by those that actually use the LA tools;
- users have to be able to relate to the items and to provide information about them;
- the items must be motivated, i.e. concept-driven and not feature driven.

To condense the framework and reduce the number of dimensions and items, the group concept mapping data was reconsidered as this ensured the continued input from the learning analytics community. Additionally, a review of the literature on other evaluation instruments, frameworks and categorisations in the field of learning analytics and related domains was conducted in order to gain further input from previously gathered experiences. These two steps revealed that the evaluation framework should consist of the four dimensions *Data*, *Awareness*, *Reflection* and *Impact* and have one section to be answered by learners and one by teachers. This implied that any reference to organisational, institutional, administrative or management-related aspects had to be removed from the framework and that the focus was instead placed on those two stakeholder groups that are directly involved in the learning process and directly impacted by the use of learning analytics tools (i.e. learners and teachers).

In order to then theoretically ground and motivate the items within the dimensions, a further review of related literature was conducted. This determined that the *Data* dimension was to consist of three items dealing with different aspects of transparency with regards to the collection of data, access to the data and presentation of data. The three selected *Awareness* items were mainly based on the situation awareness model by Endsley (2000), while the three chosen *Reflection* items were mainly inspired by the questionnaire about reflective thinking by Kember et al. (2000). The *Impact* dimension was determined to cover the three aspects of detecting students at risk as well as making learning more efficient and more effective. Thus, the second version of the evaluation framework (EFLA-2) was presented (see Figure 2.5 on page 54):

- **Data**: (1) I know what data is being collected, (2) I have access to my/my students' data, and (3) I understand the presented results
- **Awareness**: (1) I am aware of my/my students' current learning status, (2) I comprehend my/my students' current learning status, and (3) I am able to project my/my students' future learning status
- **Reflection**: (1) I reflect on my learning/teaching activities, (2) I reflect on alternative learning/teaching activities, and (3) I know when to change my learning/teaching behaviour
- **Impact**: (1) I can detect whether I am/my students are falling behind, (2) I study/My students learn more efficiently, and (3) I study/My students learn more effectively

This structure of EFLA-2 was able to reduce the framework from five dimensions with four items each to four dimensions with three items each. It also met all other requirements set by the previous evaluation study: The dimension and item names were clear and easy to understand; there were two sections of the framework, one for students and one for teachers; the framework questionnaire was meant to be answered by those two stakeholder groups that are actively using the tools; and finally, all items were motivated, i.e. concept-driven, and theoretically-grounded. Also, instead of the (mostly) single-word items in EFLA-1, the items of EFLA-2 were now formulated as statements to ease the framework's use as a questionnaire, e.g. 'I know what data is being collected' or 'I reflect on my learning activities', as this made it easier for EFLA users to relate to the items.

The next chapter formed the introduction to Part II of the thesis that dealt with the application of the EFLA-2 to a collaborative learning support widget and the framework's subsequent evaluation and improvement. **Chapter 3** presented an exploratory study about the reflective and predictive power of widget indicators of a learning analytics-based awareness widget towards students' grades by instantiating these indicators with data from four previous runs of the European Virtual Seminar on Sustainable Development (EVS). That is, although the activity widget had not been implemented into the course platform in those years, the log data from these years was explored to see what the widget indicator scores would have been if the widget had been used in those years.

The study investigated (1) whether the widget indicator scores correlated with the tutor gradings of individual students, (2) whether the scores of some widget indicators were better predictors for the students' individual grades and (3) whether certain points in time produced indicator scores that were better grade predictors than others. It was hypothesised that significant positive correlations existed between the widget indicators and the grades, that the widget indicator 'presence' was a better predictor than the other ones and that the widget indicator scores produced in the second half of the course were better predictors towards the grades than those in the first half of the course.

The results of the correlation analysis and the structural equation modelling of the exploratory study showed that most of the indicators indeed significantly and positively correlated with the grades and that they could be used as predictors. The scores of the 'presence' indicator, however, did not turn out to be better predictors for the grades, neither for the whole run nor for the individual months. Instead, the 'responsiveness' indicator achieved the best results. Looking at the individual months, the analysis showed that the months in the first half of the course yielded better correlation and structural equation modelling results than those in the second half. This unexpected outcome was attributed to an unforeseen large usage of communication tools outside of the course's learning environment.

Pertaining to the discussion about the effectiveness of learning analytics visualisations that was referred to in the related work section of Chapter 3, this study brought valuable contributions to it as evidences for the effectiveness of dashboards for reflection and awareness of pure online collaborative learning processes are clearly indicated. With the predictive power of the indicators from the widget being investigated, the study was able to show that the final grades and widget indicator scores would have been significantly and positively correlated had the widget been implemented in these previous years of the course. This overall positive result provided a useful empirical basis for the development of instructional designs and activities within the EVS online course. The confidence was high that the widget, once it was implemented in the course's environment, would foster reflection and awareness of the collaborative learning processes, would provide valuable feedback to the learners on different activities of collaborative learning, and would contribute to an adjustment of the learning design of the course.

The follow-up study was presented in **Chapter 4**. The learning analytics-based activity widget to foster awareness and reflection among the team members was implemented into the course's online learning platform and the predictive power of the widget indicators towards the students' grades of this course were examined and compared to the data from previous years where the widget had not been in use. This empirical study therefore explored (1) whether the widget indicator scores again correlated significantly and positively with the tutors' gradings of individual students, (2) whether the scores of the 'responsiveness' indicator were again better predictors for the students' individual grades than those of the others, and (3) whether the widget indicator scores produced in the first half of the course were again better predictors than those produced in the second half.

With regards to the first research question the results of the correlation and structural equation modelling analyses showed that in the year the widget was used the widget indicator scores also correlated significantly and positively with the students' grades. In comparison to the results from the exploratory study where there had been positive and significant correlations with some of the grades for all indicators, this time only the 'responsiveness' indicator showed significant correlations. The second research question, i.e. that 'responsiveness' was to be seen as the best predictor, could also be answered positively. Although 'presence' also received some significant predictor results, 'responsiveness' remained the best predictor for the whole year as well as for individual months.

For the third research question the results showed that with the widget in use the widget indicator scores from the first half of the course were not the better predictors. Instead, with the widget in use, the original hypothesis from the exploratory study that the later months produce better prediction indicators was confirmed. The results suggested that this shift might have been caused by the availability of the widget as less active students at the beginning of the course might have been stimulated to become more active or because students – knowing their activities on external tools was not reflected in the widget – were more active on the platform to better reflect their overall activity level in the course. Overall, the study results suggested that the differences between the years could be explained by the use of the widget and its effective fostering of awareness and reflection.

The implementation of the widget in a live environment was also used to have the course's students and tutors evaluate the widget by using the EFLA-2. In this second study in Chapter 4 it was tested whether the framework could be used to evaluate a learning analytics application at several points in time and to reveal differences in perception between the two stakeholder groups. To capture the students' and tutors' experiences with the widget over time, the EFLA-2 questionnaire was sent out twice: the first evaluation was conducted in the middle of the course and the second one at the end. That is, the study investigated (1) whether there was a difference in widget evaluation results between the mid-course questionnaire and the end-course questionnaire and (2) whether there was a difference in widget evaluation results between students and tutors.

The results of the two widget evaluation rounds showed that the evaluation framework for learning analytics can be used to evaluate a learning analytics application at several points in time and to reflect differences between the two stakeholder groups. Comparing the ratings from the two rounds with one another revealed that there was a difference in widget evaluation results between the mid-course questionnaire and the end-course questionnaire but for the students' *Reflection* dimension only as their *Reflection* ratings in round 2 were significantly lower than those in round 1. This difference was most likely due to the students feeling less accurately represented the more the course progressed as the activities in the external tools were not reflected in the widget scores. When comparing the evaluation results from the two user groups with one another, the only significant difference was that of the *Awareness* dimension in round 1. Here, students have rated the *Awareness* items significantly higher than the tutors did. This was most likely due to the generally positive reception of the activity widget by students already at the beginning of the course while tutors used and thus appreciated the widget more towards the end of the course when they saw their personal impressions about the students confirmed. Overall, students and tutors thus evaluated the activity widget in a very similar way.

After using EFLA-2 to evaluate the learning analytics widget developed and implemented for the collaborative online learning course, the evaluation of the framework itself was reported in **Chapter 5**. The evaluation was split into two parts. First, the results from several quantitative analyses, i.e. descriptive statistics as well as principal component and reliability analysis of the data collected during the previous widget evaluation, were used to identify which EFLA-2 items were problematic and needed further improvement. The second step of the framework evaluation study was based on the detailed input from a focus group that discussed all dimensions and their items in detail. The overarching question that guided the discussion was: *What needs to be done to improve the EFLA?*

The reported results of the quantitative as well as the qualitative analysis revealed that several items needed to be adapted to improve the framework. All changes, however, involved only the items, i.e. the dimensions stayed the same as did the framework's split in a learner and a teacher section. One change that was applied to all items was the shifting of the items' syntactical subject as this was perceived as making it clearer to EFLA users that a learning analytics tool is being evaluated and not the actions of the users themselves. Instead of all statements reading "I do ...", they were changed to refer to the learning analytics tool as the source of action to be evaluated and read "This learning analytics tool does ...".

In the *Data* dimension the two items about data access and the presentation of data were identified as problematic. Using the input from the focus group the dimension was adapted to focus on the three aspects of 'what', 'why' and 'who' in relation to data handling and transparency. For the *Awareness* dimension the third item about projecting a future learning status was identified as slightly problematic while the item about comprehending a learning status was seen as too similar to the one about becoming aware of a learning status. This resulted in the removal of one item and the rewording of another so that the dimension focused on becoming aware and forecasting a possible future situation only. Similar issues were identified in the *Reflection* dimension. The item of reflecting about alternative behaviour was too similar to the one about reflecting about behaviour in general and thus removed. The adapted dimension's focus was set on reflecting about past behaviour and being stimulated to change behaviour if needed. As the aspect of motivation was identified to be missing from the framework, an item that gathered many issues during the quantitative and qualitative results in the *Impact* dimension was removed and one about motivation was added instead. Additionally, for the teachers' section of the framework, the focus was changed to their teaching activities instead of the students' learning activities in terms of efficiency and effectiveness. Chapter 5 concluded with the presentation of the EFLA-3, the third version of the evaluation framework for learning analytics (see Figure 5.1 on page 112):

- **Data**: For this LA tool it is clear (1) what data is being collected, (2) why the data is being collected, (3) who has access to the data
- **Awareness**: This LA tool makes me (1) aware of my/my students' current learning situation, (2) forecast my/my students' possible future learning situation given my/their (un)changed behaviour
- **Reflection**: This LA tool stimulates me (1) to reflect on my past learning/teaching behaviour, (2) to adapt my learning/teaching behaviour if necessary
- **Impact**: This LA tool (1) increases my motivation to study/teach, (2) stimulates me to study/teach more efficiently, (3) stimulates me to study/teach more effectively

Overall, the detailed discussion of all items facilitated the identification of several issues hampering the applicability of the evaluation framework. Continuing the efforts of reducing the framework to a core set of meaningful and easily applicable items, the framework was further condensed from four dimensions and a total of twelve items to four dimensions and a total of ten items.

Chapter 6 constituted Part III of this thesis. It contained the last iteration of the usage, evaluation and improvement of the evaluation framework for learning analytics. The chapter was divided into two studies: one covered the application of the EFLA-3 to several learning analytics widgets while the other dealt with the validity and reliability analysis of the framework itself.

While Chapter 4 investigated whether the evaluation framework for learning analytics could be used to measure differences between different points in time for the same learning analytics widget for the two stakeholders, this chapter's first study focused on the comparability of widget versions, i.e. it investigated whether the EFLA could be used to measure changes between different versions of widgets and the two stakeholder groups. The idea behind this was that changes in a widget would be reflected in the EFLA ratings of the new widget version. The reported study took place in a controlled experimental lab setting using two existing widgets from the learning analytics dashboard of the ECO MOOC platform and their two specifically developed and implemented updated versions. The new versions had been designed based on input from related literature about dashboards in massive open online courses. All study participants filled in the EFLA-3 questionnaire for all four widgets.

The evaluation of the widgets using the EFLA-3 questionnaire showed that there were indeed significant differences in evaluation results between the different widget versions for both stakeholder groups. On average students and teachers gave better ratings to the second versions of both widgets. While the original versions of the widgets received higher ratings from the teachers, the updated widget versions received higher ratings from the students. The number of significant differences was much larger for the students than for the teachers which pointed to the students appreciating the second versions of the widgets much more than the first versions.

Apart from the quantitative results, additional qualitative data was collected as well. The qualitative feedback collected from the study participants mirrored the quantitative results as they revealed that teacher participants were much more hesitant and held back by the lab setting of the experiment while the student participants could easily put themselves in the mindset of an online course participant. Another factor that was likely to play a role in influencing the teachers' widget evaluations was that due to the ECO platform's not distinguishing between the user types of learners and teachers when displaying the visualisations, the personalisation aspect of the widgets' second versions was rather pointless for the teachers so that they might have felt disregarded. Nevertheless, the study was able to show that the evaluation framework for learning analytics can be used to measure changes in perception between different versions of widgets as well as differences between the two stakeholder groups.

Finally, the EFLA-3 questionnaire data collected during the empirical widget evaluation study was used to establish the validity and reliability of the evaluation framework by means of principal component analysis and reliability analysis. The two research questions guiding this study were (1) whether the four EFLA-3 dimensions validly represented the underlying structure and (2) whether the items within the dimensions reliably measured the underlying component.

The results of a first round of principal component analysis hinted at the fact that the EFLA-3's four dimensional structure might not be valid. Additionally, the reliability analysis for the different scales then revealed that there were items that needed to be removed in order to increase the scales' reliability scores. The main reason why one of the *Data* items did not fit the framework was seen to be due to the fact that in order to interpret a visualisation it was important to know what data it was based on and why (i.e. its purpose), but to know who else had access to the data would not affect the interpretation. The item eliminated from the *Impact* dimension dealt with user motivation and was deemed too general and too content- and context-dependent, i.e. a learning analytics tool might motivate a user on one day but not on the next due to its content or external factors.

The second round of principal component analysis confirmed that the evaluation framework for learning analytics should consist of three dimensions to validly reflect the underlying components. The two dimensions of *Awareness* and *Reflection* needed to be combined into one dimension. The reliability analysis did not reveal any further need for item removal. Thus, the fourth and final version of the evaluation framework (EFLA-4) was presented (see Figure 6.6 on page 136):

- **Data**: For this LA tool it is clear (1) what data is being collected, (2) why the data is being collected
- **Awareness & Reflection**: This LA tool (1) makes me aware of my/my students' current learning situation, (2) makes me forecast my/my students' possible future learning situation given my/their (un)changed behaviour, (3) stimulates me to reflect on my past learning/teaching behaviour, (4) stimulates me to adapt my learning/teaching behaviour if necessary
- **Impact**: This LA tool (1) stimulates me to study/teach more efficiently, (2) stimulates me to study/teach more effectively

With this last evaluation and improvement iteration, a further condensation and increased applicability of the evaluation framework for learning analytics was thus achieved. EFLA-4 now consisted of three dimensions with a total of eight items. With this structure the validity and reliability of the framework was confirmed. The framework was made openly and publicly available, i.e the framework's template flyer (see Figure 6.7 on page 138) as well as an interactive spreadsheet to automatically calculate the EFLA scores and create visualisations of the scores have been made accessible via the LACE (Learning Analytics Community Europe) website[8] for all those who are interested to use the framework in order to evaluate their learning analytics tools.

Limitations of this research

There were several aspects that had to be kept in mind while the presented research was conducted. Some of these limitations were more related to the development, implementation and evaluation of the involved learning analytics widgets while others pertained more to the construction and evaluation of the evaluation framework for learning analytics itself. However, none of the limitations were so strong that they could dispute the presented conclusions.

The limiting issue identified in Part I was related to the participants of the group concept mapping study. Most of the participants, especially those of phase 2 where the demographics were available, were working at a university and were more research- than practice-oriented. As most participants were thus involved in the field of higher education, input from workplace learning as well as from schools was low. This, however, was not seen as problematic in the context of this research as the

[8] `http://www.laceproject.eu/evaluation-framework-for-la/`

identification of quality indicators for learning analytics can be assumed to be rather similar across the different sectors, i.e. very similar overall results would have been expected had the participants mainly represented one of the other sectors. The focus here was set on getting input from experts, i.e. those people actively involved in the field. As the application of learning analytics and research in this field was and still is much more prominent in the higher education sector than in the others (Ferguson et al., 2016a), using input from this sector was the most feasible way to go with regards to this research.

In Part II several limitations needed to be considered in relation to the presented awareness widget. One aspect of using an activity-based widget to support group awareness in a collaborative learning environment was to avoid the time consuming, tedious and disruptive aspects of lengthy questionnaires. It was understood that analysing distal data such as activity logs from a learning environment could of course never be used as a one-to-one replacement for proximal data such as questionnaires or interviews, i.e. proximal variables have indeed more predictive power than distal variables (Fishbein and Ajzen, 2010). However, the use of learning analytics can contribute to and enrich reflection and awareness processes for learners as well as teachers especially and can be used as an additional indication towards group activities due to its non-disruptiveness and its taking into account of the full student cohort at the same time. Another limitation related to the widget presented in Part II was that although the research looked at behavioural data, it did not examine learning as a process itself. Neither did it explore whether any learning actually took place (for the purpose of the studies it was assumed that a student's grade is an indicator of knowledge level) nor did it actually observe learning where and how it took place, e.g. in the form of brain activity and modifications. Bio-psychological and educational neuroscience research is of huge importance for discovering the phenomenon of learning. On many levels, however, the brain and its ways of working are still a mystery. And although the recent year saw learning analytics researchers contributing to this field by combining log data with data from biophysical sensors, addressing and taking into account these issues was out of the scope of this thesis.

Apart from the limitations that were encountered using a widget with such a type of visualised information, there were also some risks associated with it. One of them was that students could use it 'strategically', e.g. by posting many short, largely irrelevant messages to improve their scores. In the exploratory study this risk did not play a role. In the later study, however, it did. From the results gathered it did not look like the students in this research 'played the system' but there was evidence that some students did post more in order to better reflect their platform-external activities. Although this limitation did not play a role in the reported research, it is an issue that needs to be taken into account whenever such a type of information visualisation is used. Similarly, if learning analytics tools meant to foster collaborative processes have privacy options implemented, e.g. such as the reciprocal privacy model described in this research, certain risks would have to be taken into account. Theoretically, if many or even all students within a group chose not to share their data, the widget's intention to support awareness and reflection of collaborative learning processes would be seriously interfered with or even prevented. A further risk in the presented research was thus that by providing the students with privacy mechanisms, the likelihood of the widget being able to be the supportive tool it was meant to be decreased. As none of the students changed the privacy setting, the presented research was not influenced. It is, however, a very important aspect to keep in mind when learning analytics tools with such privacy settings are being explored and analyses.

Further limitations that were revealed in Part II were related to the comparison of the results from the different runs of the EVS course. Due to the change in student population, the students' behaviour in the five different runs could not be set into a one-to-one relation. Their previous experience with and usage of online learning platforms as well as external communication and collaboration tools influenced the cohort's actions. The same applied to the tutors. Although many of them had been tutors for EVS for a number of years, their experience and interactions with their student groups also changed from year to year. Related to this aspect of change in student population, student and tutor behaviour as well as external tools was another aspect that had to be kept in mind when looking at the results of the online study: although a number of the observations could be explained as effects of the activity widget being in use, there was no proof that this was the case. Only after observing and analysing further years of the EVS will it be possible to clearly attribute differences between the years that did not have the widget and those that did to the use of the widget.

There are several limitation to the research reported in Part III of this thesis. First, the data collection for the reported study did not take place in a live environment but in a lab setting. This meant that the participants were not able to actually use the different widgets themselves within a course environment for an extended period of time. Instead, they were only shown images of the widgets and some live demos while listening to descriptions of the widget. The original idea of evaluating the ECO widgets with ECO users had to be discarded due to the limited amount of teacher participants available. While not collecting data in a live environment was an important limitation, it was at the same time, however, also a beneficial aspect of the presented research. Due to the laboratory setting, it could be controlled that there was an equal number of student and teacher participants, that all participants were given the same information and that they all answered all questions. The second limitation with regards to the study set up in Part III of this research was that the roles of the participants were not assigned randomly but according to their current level of employment at the university. While a random assignment might have provided more impartial results, it was instead decided to assign the roles as close as possible to a real-life student-teacher setting, i.e. PhD candidates took over the roles of students while all assistant, associate and full professors took over the roles of teachers. The third limitation with regards to the lab setting of the presented study was that the widgets were shown in the same order to all participants. The widget evaluation results might therefore possibly be biased.

Implications and Future Research

Becoming its own distinct research field in 2011, the expectations for learning analytics to solve the retention problem, to increase student success and to support learning and teaching processes were very high. However, a number of years later now, empirical evidence as to whether learning analytics tools have the desired effect or not is still very sparse and there is still a gap between the potential of learning analytics identified by research and how much of this potential has been achieved so far (Ferguson et al., 2016a). Some even say that after a phase of hype, learning analytics "has entered a trough of disillusionment" (Sharkey and Harfield, 2016, para. 1) but at the same time argue that "it is a good and necessary stage in analytics' maturity. It is a crucible out of which we expect to see more realistic expectations, increased accountability, and true innovation in the service of institutional performance and student success" (Sharkey and Harfield, 2016, para. 11). In early 2016, at the peak of the learning analytics hype, the NMC Horizon report saw learning analytics to be on the 'one year or less time-to-adoption horizon' (Johnson et al., 2016). One year later, this position is taken over by 'adaptive learning technologies', i.e. the focus is now more set on an holistic view of the students and the enhancement and personalisation of their learning processes. With regards to tracking and evaluating evidence, Adams Becker et al. (2017) therefore ask: "What good is a new approach or technology deployment if the results are not carefully measured and analyzed, with the program adjusted based on the results?" (p.6). Using the System Usability Scale (SUS) (Brooke, 1996) as an inspiration, the research presented in this thesis addressed these issues by creating and validating an evaluation framework for learning analytics that helps standardise the evaluation of learning analytics tools and allows for measuring and comparing the impact of learning analytics on educational practices. Similar to the SUS, the EFLA makes use of the subjective assessments about learning analytics tools by their users in order to obtain a general indication of the overall quality of a tool in a quick and simple, yet thoroughly developed, validated and reliable way.

The main objectives of the presented research were to identify quality indicators for learning analytics, to create an applicable evaluation instrument based on these indicators and to validate the evaluation instrument. The process of creating, iteratively applying, evaluating and improving the evaluation framework for learning analytics (EFLA) resulted in a valid and reliable version of the framework that offers a standardised way to evaluate learning analytics tools and to measure and compare the impact of learning analytics on educational practices of learners and teachers. The methodologies covered in the different studies in order to reach these objectives ranged from the involvement of the learning analytics community and the consultation of related research to the employment of learning analytics tools in different courses and the involvement of learners and teachers in the evaluation processes up to the statistical validation of the framework. Taking the current state of the research field into account, the results of the conducted research entail several practical implications in relation to the usage of the evaluation framework.

During the EFLA's evaluation and improvement cycles, different course contexts (collaborative formal course vs. informal MOOC) as well as different widgets (activity widget for collaborative learning settings vs. activity and resources widget for an individual learning setting) have been used in addition to different study procedures (using the EFLA to evaluate the same widget at different points in time vs. using the EFLA to evaluate different widgets at the same time) as well as different study settings (live study vs. lab study). These different studies showed that the evaluation framework for learning analytics can be used successfully in a broad range of circumstances. In addition, with its validity and reliability established, the framework is now ready to close the gap between an institution's measurement of learning analytics readiness and its learning analytics maturity. It can complement the evaluation performed on an institutional level by providing input directly from the learners and teachers, thus allowing to create a holistic view of the learning and teaching processes involved. Furthermore, it also addresses the need for evidence of impact that the field has been calling for.

Therefore, the learning analytics community now has the opportunity to verify the EFLA's applicability and benefit. Just as the System Usability Scale (SUS) was able to fully claim its success after being picked up and used by the research community (Brooke, 2013), usage of the EFLA in the field and on a larger scale will be needed to fully exploit its potential. The ready-for-download questionnaire templates[9] as well as the scoring spreadsheet[10] ease the adoption of the evaluation framework and allow for a similarly easy and 'quick and dirty' evaluation as the SUS. Once the EFLA has been used in a large number of cases and the scores of 50, 100 or even more learning analytics tools are available, follow-up studies evaluating the usage and the results of the framework have to be conducted to discuss and induce possible adaptations. Additionally, with a large pool of tool evaluations using the EFLA available, an average EFLA score across all evaluations as well as the distribution of all scores can be determined which would then allow for the EFLA scores to be turned into percentile rankings or grades.

But in order to 'go big' two aspects will have to be addressed. On the one hand, the 'market', i.e. the educational institutions, companies, teachers, students or researchers who design, develop, implement, set in place, research, provide, buy, etc. the learning analytics tools, will need to be convinced that using the EFLA to evaluate their tools is beneficial. On the other hand, the users, i.e. the learners and teachers who are being presented with the learning analytics tools and are using them, will need to be convinced to answer the EFLA questionnaire as without participation from the users, no evaluation will be possible. There are several related incentives that can be offered to both of these groups.

First, the EFLA provides insights. Using the framework to evaluate learning analytics tools can provide insight to the learners' and teachers' perception of and experience with the tool. It can reveal problematic aspects and identify ways to provide students

[9] `https://rebrand.ly/EFLAtemplate`
[10] `https://rebrand.ly/EFLAscoring`

with a more adaptive and less one-size-fits-all learning experience, as suggested by the field (Teasley, 2017). Once such issues are identified, they can be addressed in updated and improved versions of the learning analytics tool. The evaluated tool could be a whole dashboard as well as a single visualisation. The level of detail chosen is left to those who conduct the evaluation.

Second, the EFLA facilitates comparability. The framework can be used to compare learning analytics tools within one setting, e.g. two widgets for one course, or between different settings, e.g. widgets and dashboards from different courses or even from different educational institutions. Knowing how a tool performs according to the different EFLA dimensions can help to position it in the growing collection of tools available and can stimulate further development. If the results of EFLA scores were to be made publicly available, e.g. by commercial learning analytics providers or by those who publish their tools as open source, the comparability could be taken one step further. A tool's EFLA score could thus be used for advertisement purposes on both a commercial and an academic level.

Third, the EFLA supplies evidence. With the rising urge to ground learning analytics tools more in learning theories (Jivet et al., 2017), the framework can be used to ascertain whether a learning analytics tool fulfilled its intended purpose, i.e. whether it actually had an impact on learning and teaching processes and made them more efficient and more effective. This is what the community itself (e.g. Ferguson and Clow (2017)) and also policy makers have been calling for (e.g. ET2020 Working Group on Digital Skills and Competences (2016)). Additionally, by being able to show, i.e. by providing evidence, that a learning analytics tool really does fulfil its intended purpose, the rationale for collecting user data, which is seen as one of the biggest privacy-related issues, can be accounted for.

Whichever of these aspects poses as the main incentive for the 'market' to apply the EFLA for the evaluation of their learning analytics tools, they will have to be forwarded to the users to assure their participation. Only if the users see and understand the rationale behind having to answer the EFLA questionnaire (possibly again and again) and perceive the added value for themselves, will they be willing to provide their input. In the end, it needs to come down to both sides understanding that, as Baker (2016) describes, the goal of collecting and analysing data in the educational domain is not to create intelligent systems or stupid systems, but to create intelligent and successful students and to promote education. Finally, as was postulated in the 2017 NMC Horizon report by Adams Becker et al. (2017, p. 6): "If education is viewed as a vehicle for advancing the global economy, then it must be the North Star that guides societies to the next big thing, illuminating new ideas that solve pressing challenges and creating opportunities to shape a better future."

References

Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., and Ananthanarayanan, V. (2017). *NMC Horizon Report: 2017 Higher Education Edition*. The New Media Consortium, Austin, Texas.

Aguilar, S., Lonn, S., and Teasley, S. D. (2014). Perceptions and Use of an Early Warning System During a Higher Education Transition Program. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, LAK '14, pages 113–117, New York, NY, USA. ACM. DOI: 10.1145/2567574.2567625.

Arnold, K. E. (2010). Signals: Applying Academic Analytics. *EDUCAUSE Review*. available online at http://er.educause.edu/articles/2010/3/signals-applying-academic-analytics.

Arnold, K. E., Lonn, S., and Pistilli, M. D. (2014a). An Exercise in Institutional Reflection: The Learning Analytics Readiness Instrument (LARI). In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, LAK '14, pages 163–167, New York, NY, USA. ACM. DOI: 10.1145/2567574.2567621.

Arnold, K. E., Lynch, G., Huston, D., Wong, L., Jorn, L., and Olsen, C. W. (2014b). Building Institutional Capacities and Competencies for Systemic Learning Analytics Initiatives. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, LAK '14, pages 257–260, New York, NY, USA. ACM. DOI: 10.1145/2567574.2567593.

Arnold, K. E. and Pistilli, M. D. (2012). Course Signals at Purdue: Using Learning Analytics to Increase Student Success. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge*, LAK '12, pages 267–270, New York, NY, USA. ACM. DOI: 10.1145/2330601.2330666.

Arroway, P., Morgan, G., O'Keefe, M., and Yanosky, R. (2016). *Learning Analytics in Higher Education (Research Report)*. EDUCAUSE Center for Applied Research, Louisville, CO. available online at https://library.educause.edu/~/media/files/library/2016/2/ers1504la.pdf.

Baker, R. S. (2016). Stupid Tutoring Systems, Intelligent Humans. *International Journal of Artificial Intelligence in Education*, 26(2):600–614. DOI: 10.1007/s40593-016-0105-0.

Baker, R. S. and Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In Larusson, J. A. and White, B., editors, *Learning Analytics: From Research* *to Practice*, pages 61–75. Springer New York, New York, NY. DOI: 10.1007/978-1-4614-3305-7_4.

Baron, J. and Jayaprakash, S. (2014). SNAPP (Social Network Analysis & Pedagogical Practices) for Sakai cle v2.8x & v2.9x. website. https://confluence.sakaiproject.org/x/MYEPBQ.

Beheshitha, S., Hatala, M., Gašević, D., and Joksimovic, S. (2016). The Role of Achievement Goal Orientations When Studying Effect of Learning Analytics Visualizations. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, LAK '16, pages 54–63, New York, NY, USA. ACM. DOI: 10.1145/2883851.2883904.

Bennett, T. and Yalams, S. (2013). Correlates of students' attendance to class, participation and performances in engineering modules. In *2013 IEEE Global Engineering Education Conference (EDUCON)*, pages 947–951. DOI: 10.1109/EduCon.2013.6530221.

Berg, A., Scheffel, M., Drachsler, H., Ternier, S., and Specht, M. (2016a). Dutch Cooking with xAPI Recipes: The Good, the Bad and the Consistent. In *Proceedings of the International Conference on Advanced Learning Technologies*, ICALT '16, pages 234–236. DOI: 10.1109/ICALT.2016.48.

Berg, A., Scheffel, M., Drachsler, H., Ternier, S., and Specht, M. (2016b). The Dutch xAPI Experience. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, LAK '16, pages 544–545, New York, NY, USA. ACM. DOI: 10.1145/2883851.2883968.

Bichsel, J. (2012). *Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations (Research Report)*. EDUCAUSE Center for Applied Research, Louisville, CO. available online at http://net.educause.edu/ir/library/pdf/ers1207/ers1207.pdf.

Blikstein, P., Merceron, A., and Siemens, G., editors (2015). *LAK '15: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, New York, NY, USA. ACM.

Boekaerts, M. and Niemivirta, M. (2000). Chapter 13 – Self-Regulated Learning: Finding a Balance between Learning Goals and Ego-Protective Goals. In Boekaerts, M., Pintrich, P. R., and Zeidner, M., editors, *Handbook of Self-Regulation*, pages 417–450. Academic Press, San Diego. DOI: 10.1016/B978-012109890-2/50042-1.

Bolton, G. (2010). *Reflective practice: Writing & professional development*. Sage, London, UK, 3rd edition edition.

Borkowski, J., Chan, L., and Muthukrishna, N. (2000). A Process-oriented Model of Metacognition: Links between Motivation and Executive Functioning. In Schraw, G. and Impara, J., editors, *Issues in the Measurement of Metacognition*. Buros Institute of Mental Measurements, University of Nebraska, Lincoln, NE.

Brahimi, T. and Sarirete, A. (2015). Learning Outside the Classroom Through MOOCs. *Computers in Human Behavior*, 51(PB):604–609. DOI: 10.1016/j.chb.2015.03.013.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In Jordan, P. W., Weerdmeester, B., Thomas, A., and Mclelland, I. L., editors, *Usability evaluation in industry*. Taylor and Francis, London.

Brooke, J. (2013). SUS: A Retrospective. *Journal of Usability Studies*, 8(2):29–40.

Brouns, F., Mota, J., Morgado, L., Jansen, D., Fano, S., Silva, A., and Teixeira, A. (2014). A Networked Learning Framework for Effective MOOC Design: The ECO Project Approach. In *Proceedings of the 8th EDEN Research Workshop*, pages 161–171, Budapest, Hungary. EDEN.

Browne, M. and Cudeck, R. (1989). Single Sample Cross-validation Indices for Covariance Structures. *Multivariate Behavioral Research*, 24:445–455. DOI: 10.1207/s15327906mbr2404_4.

Bruyckere, P. D., Kirschner, P. A., and Hulshof, C. D., editors (2015). *Urban Myths about Learning and Education*. Academic Press, San Diego. DOI: 10.1016/B978-0-12-801537-7.00008-1.

Buckingham Shum, S., Ferguson, R., and Deakin Crick, R. (2012a). Enquiryblogger: Blog-based Learning Analytics for Learning Power & Authentic Enquiry. slides. available online at http://de.slideshare.net/sbs/enquirybloggeranalyticscalrg2012.

Buckingham Shum, S., Gašević, D., and Ferguson, R., editors (2012b). *LAK '12: Proceedings of the Second International Conference on Learning Analytics and Knowledge*, New York, NY, USA. ACM.

Bull, S., Ginon, B., Boscolo, C., and Johnson, M. (2016). Introduction of Learning Visualisations and Metacognitive Support in a Persuadable Open Learner Model. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, LAK '16, pages 30–39, New York, NY, USA. ACM. DOI: 10.1145/2883851.2883853.

Butler, D. and Winne, P. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65(3):245–281. DOI: 10.3102/00346543065003245.

Charlton, B. (2000). *Psychiatry and the Human Condition*. Radcliffe Medical Press, Oxford, UK.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., and Thüs, H. (2012). A Reference Model for Learning Analytics. *International Journal for Technology Enhanced Learning*, 4(5/6):318–331. DOI: 10.1504/IJTEL.2012.051815.

Cisco (2016). Cisco VNI Forecast and Methodology, 2015-2020. White paper, Cisco. available online at http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html.

Clow, D. (2012). The Learning Analytics Cycle: Closing the Loop Effectively. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge*, LAK '12, pages 134–138, New York, NY, USA. ACM. DOI: 10.1145/2330601.2330636.

Clow, D., Cross, S., Ferguson, R., and Rienties, B. (2014). D2.5 Evidence Hub Review. Deliverable, LACE Project. available online at http://www.laceproject.eu/deliverables/d2-5-evidence-hub/.

Cobo, A., Rocha, R., and Rodriguez-Hoyos, C. (2014). Evaluation of the Interactivity of Students in Virtual Learning Environments Using a Multicriteria Approach and Data Mining. *Behaviour & Information Technology*, 33(10):1000–1012. DOI: 10.1080/0144929X.2013.853838.

Coffrin, C., Corrin, L., de Barba, P., and Kennedy, G. (2014). Visualizing Patterns of Student Engagement and Performance in MOOCs. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, LAK '14, pages 83–92, New York, NY, USA. ACM. DOI: 10.1145/2567574.2567586.

Colvin, C., Rogers, T., Alexandra, W., Dawson, S., Gašević, D., Buckingham Shum, S., Nelson, K., Alexander, S., Lockyer, L., Kennedy, G., Corrin, L., and Fisher, J. (2016). *Student Retention and Learning Analytics: a Snapshot of Australian Practices and a Framework for Advancement*. Australian Government Office for Learning and Teaching, Canberra, ACT.

Credé, M., Roch, S. G., and Kieszczynka, U. M. (2010). Class Attendance in College: A Meta-Analytic Review of the Relationship of Class Attendance With Grades and Student Characteristics. *Review of Educational Research*, 80(2):272–295. DOI: 10.3102/0034654310362998.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3):319–340. DOI: 10.2307/249008.

Dawson, S., Drachsler, H., and Penstein Rose, C., editors (2016). *LAK '16: Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, New York, NY, USA. ACM.

de Kraker, J. and Cörvers, R. (2014). European Virtual Seminar on Sustainable Development: international, multi-disciplinary learning in an online social network. *E-learning and Education for Sustainability, Series 'Environmental Education, Communication and Sustainability'*, 35:117–136.

Desouza, K. C. and Smith, K. L. (2016). Predictive Analytics: Nudging Shoving and Smacking Behaviors in Higher Education. *EDUCAUSE Review*, pages 10–20. available online at http://er.educause.edu/∼/media/files/articles/2016/8/erm1651.pdf.

Drachsler, H., Bogers, T., Vuorikari, R., Verbert, K., Duval, E., Manouselis, N., Beham, G., Lindstaedt, S., Stern, H., Friedrich, M., and Wolpers, M. (2010). Issues and Considerations Regarding Sharable Data Sets for Recommender Systems in Technology Enhanced Learning. *Procedia Computer Science*, 1(2):2849–2858. DOI: 10.1016/j.procs.2010.08.010.

Drachsler, H. and Greller, W. (2012). The Pulse of Learning Analytics – Understandings and Expectations from the Stakeholders. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge*, LAK '12, pages 120–129, New York, NY, USA. ACM. DOI: 10.1145/2330601.2330634.

Drachsler, H. and Greller, W. (2016). Privacy and Analytics – it's a DELICATE issue. A Checklist to establish trusted Learning Analytics. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, LAK '16, pages 89–98, New York, NY, USA. ACM. DOI: 10.1145/2883851.2883893.

Drachsler, H., Greller, W., Griffiths, D., Hoel, T., and Kickmeier-Rust, M. (2016). Is Privacy a Show-stopper for Learning Analytics? A Review of Current Issues and their Solutions. *Learning Analytics Review*, (6). available online at http://www.laceproject.eu/learning-analytics-review/files/2016/04/LACE-review-6_privacy-show-stopper.pdf.

Drachsler, H., Greller, W., Stoyanov, S., Fetahu, B., Rex, C., Herder, E., Daga, E., Adamou, A., Parodi, E., and Marco, M. (2013). D2.1 Evaluation Criteria and Methods. Deliverable, LinkedUp Project. available online at http://linkedup-project.eu/files/2013/09/LinkedUp-D2.1-EvaluationCriteria.pdf.

Drachsler, H., Hummel, H., and Koper, R. (2009). Identifying the Goal, User Model and Conditions of Recommender Systems for Formal and Informal Learning. *Journal of Digital Information*, 10(2):4–24. available online at https://journals.tdl.org/jodi/index.php/jodi/article/view/442.

Drachsler, H. and Kalz, M. (2016). The MOOC and Learning Analytics Innovation Cycle (MOLAC): A Reflective Summary of Ongoing Research and its Challenges. *Journal of Computer Assisted Learning*, 32(3):281–290. DOI: 10.1111/jcal.12135.

Drachsler, H., Stoyanov, S., Guy, M., and Scheffel, M. (2014). D2.2.2 Final Version of the LinkedUp Evaluation Framework. Deliverable, LinkedUp Project. available online at http://linkedup.websites.okfn.org/files/2014/11/LinkedUp_D2.2.2.pdf.

Duval, E. (2011). Attention please!: Learning analytics for visualization and recommendation. In *Proceedings of the First International Conference on Learning Analytics and Knowledge*, LAK '11, pages 9–17, New York, NY, USA. ACM. DOI: 10.1145/2090116.2090118.

Dyckhoff, A. L., Lukarov, V., Muslim, A., Chatti, M. A., and Schroeder, U. (2013). Supporting action research with learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 220–229, New York, NY, USA. ACM. DOI: 10.1145/2460296.2460340.

ECAR-ANALYTICS Working Group (2015). The Predictive Learning Analytics Revolution: Leveraging Learning Data for Student Success. Ecar working group paper, Educause Center for Analysis and Research, Louisville, CO. available online at https://net.educause.edu/ir/library/pdf/ewg1510.pdf.

EDUCAUSE (2014). Managing data risk in student success systems. Educause ipas summit report. available online at https://net.educause.edu/ir/library/pdf/PUB9015.pdf.

Elias, T. (2011). Learning Analytics: Definitions, Processes and Potential. Technical report. available online at http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf.

Elkina, M., Fortenbacher, A., and Merceron, A. (2013). The Learning Analytics Application LEMO – Rationals and First Results. *International Journal of Computing*, 12(3):226–234.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64. DOI: 10.1518/001872095779049543.

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: a critical review. In Endsley, M. R. and Garland, D. J., editors, *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates, Mahwah, NJ, USA.

Engelfriet, A., Manderveld, J., and Jeunink, E. (2015). Learning Analytics under the Dutch Data Protection Act. Policy paper, SURFnet. available online at https://www.surf.nl/en/knowledge-base/2015/guide-to-learning-analytics-under-the-personal-data-protection-act.html.

ET2020 Working Group on Digital Skills and Competences (2016). Learning Analytics – Key messages. report, European Commission. available online at https://ec.europa.eu/education/sites/education/files/2016-pla-learning-analytics_en.pdf.

Ferguson, R. (2012a). Learning Analytics: Drivers, Developments and Challenges. *International Journal of Technology Enhanced Learning*, 4(5/6):304–317. DOI: 10.1504/IJTEL.2012.051816.

Ferguson, R. (2012b). The State of Learning Analytics in 2012: A Review and Future Challenges. Technical Report KMI-12-01, The Open University. available online at http://kmi.open.ac.uk/publications/techreport/kmi-12-01.

Ferguson, R. (2014). Learning analytics don't just measure students' progress – They can shape it. *theguardian.com*. available online at https://www.theguardian.com/education/2014/mar/26/learning-analytics-student-progress.

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., and Vuorikari, R. (2016a). Research Evidence on the Use of

Learning Analytics – Implications for Education Policy. Joint research centre science for policy report, European Union. DOI: 10.2791/955210.

Ferguson, R., Buckingham Shum, S., and Deakin Crick, R. (2011). Discussion paper: Enquiryblogger – Using Widgets to Support Awareness and Reflection in a PLE Setting. In *Workshop on Awareness and Reflection in Personal Learning Environments*, PLE conference.

Ferguson, R. and Clow, D. (2016). Learning Analytics Community Exchange: Evidence Hub. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, LAK '16, pages 520–521, New York, NY, USA. ACM. DOI: 10.1145/2883851.2883878.

Ferguson, R. and Clow, D. (2017). Where is the Evidence?: A Call to Action for Learning Analytics. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge*, LAK '17, pages 56–65, New York, NY, USA. ACM. DOI: 10.1145/3027385.3027396.

Ferguson, R., Hoel, T., Scheffel, M., and Drachsler, H. (2016b). Guest Editorial: Ethics and Privacy in Learning Analytics. *Journal of Learning Analytics*, 3(1):5–15.

Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, Inc.

Fishbein, M. and Ajzen, I. (2010). *Predicting and Changing Behavior: The Reasoned Action Approach*. Psychology Press, New York, NY, USA.

Florian, B., Glahn, C., Drachsler, H., Specht, M., and Fabregat Gesa, R. (2011). Activity-Based Learner-Models for Learner Monitoring and Recommendations in Moodle. In Kloos, C. D., Gillet, D., Crespo García, R. M., Wild, F., and Wolpers, M., editors, *Towards Ubiquitous Learning: Proceedings of the 6th European Conference of Technology Enhanced Learning (EC-TEL 2011)*, volume 6964 of *LNCS*, pages 111–124, Berlin, Heidelberg. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-23985-4_10.

Friedman, B. (1997). *Human Values and the Design of Computer Technology*. Cambridge University Press, Cambridge, MA.

Gašević, D., Dawson, S., Mirriahi, N., and Long, P. (2015a). Learning Analytics – A Growing Field and Community Engagement. *Journal of Learning Analytics*, 2(1):1–6. DOI: 10.18608/jla.2015.21.1.

Gašević, D., Dawson, S., and Siemens, G. (2015b). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1):64–71. DOI: 10.1007/s11528-014-0822-x.

Govaerts, S., Verbert, K., and Duval, E. (2011). Evaluating the Student Activity Meter: Two Case Studies. In Leung, H., Popescu, E., Cao, Y., Lau, R., and Nejdl, W., editors, *Advances in Web-Based Learning – Proceedings of the ICWL 2011*, volume 7048 of *LNCS*, pages 188–197, Berlin, Heidelberg. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-25813-8_20.

Govaerts, S., Verbert, K., Duval, E., and Pardo, A. (2012). The Student Activity Meter for Awareness and Self-reflection. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 869–884, New York, NY, USA. ACM. DOI: 10.1145/2212776.2212860.

Greller, W. and Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society*, 15(3):42–57.

Hattie, J. and Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1):81–112. DOI: 10.3102/003465430298487.

Hoel, T., Griffiths, D., and Chen, W. (2017). The Influence of Data Protection and Privacy Frameworks on the Design of Learning Analytics Systems. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge*, LAK '17, pages 243–252, New York, NY, USA. ACM. DOI: 10.1145/3027385.3027414.

Hoyle, R. (1995). *Structural Equation Modelling: Concepts, Issues, and Applications*. Sage, Thousand Oaks, CA.

ISO9241-11:1998(E) (1998). Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. ISO 9241, International Organization for Standardization, Geneva, Switzerland.

Jivet, I., Scheffel, M., Drachsler, H., and Specht, M. (2017). Awareness is not enough. Pitfalls of learning analytics dashboards in the educational practice. In *Proceedings of the 12th European Conference on Technology Enhanced Learning (EC-TEL 2017)*, LNCS, Berlin, Heidelberg. Springer.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., and Hall, C. (2016). *NMC Horizon Report: 2016 Higher Education Edition*. The New Media Consortium, Austin, Texas.

Kane, M. and Trochim, W. (2007). *Concept Mapping for Planning and Evaluation*. Sage Publication, Thousand Oaks, CA.

Kay, D., Korn, N., and Oppenheim, C. (2012). Legal, Risk and Ethical Aspects of Analytics in Higher Education. *CETIS Analytics Series*, 1(6). available online at http://publications.cetis.org.uk/wp-content/uploads/2012/11/Legal-Risk-and-Ethical-Aspects-of-Analytics-in-Higher-Education-Vol1-No6.pdf.

Kember, D., Leung, D. Y., Jones, A., Loke, A. Y., McKay, J., Sinclair, K., Tse, H., Webb, C., Yuet Wong, F. K., Wong, M., and Yeung, E. (2000). Development of a Questionnaire to Measure the Level of Reflective Thinking. *Assessment & Evaluation in Higher Education*, 25(4):381–395. DOI: 10.1080/713611442.

Khan, I. and Pardo, A. (2016). Data2U: Scalable Real Time Student Feedback in Active Learning Environments. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, LAK '16, pages 249–253, New York, NY, USA. ACM. DOI: 10.1145/2883851.2883911.

Kim, J., Jo, I.-H., and Park, Y. (2016). Effects of Learning Analytics Dashboard: Analyzing the Relations Among Dashboard Utilization, Satisfaction, and Learning Achievement. *Asia Pacific Education Review*, 17(1):13–24. DOI: 10.1007/s12564-015-9403-8.

Kirschner, P., Kreijns, K., Phielix, C., and Fransen, J. (2015). Awareness of cognitive and social behaviour in a CSCL environment. *Journal of Computer Assisted Learning*, 31(1):59–77.

Koedinger, K. R., Baker, R., Cunningham, K., Skogsholm, A., Leber, B., and Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. *Handbook of Educational Data Mining*, 43.

Korkofingas, C. and Macri, J. (2013). Does Time Spent Online have an Influence on Student Performance? Evidence for a Large Business Studies Class. *Journal of University Teaching & Learning Practice*, 10(2). available online at http://ro.uow.edu.au/jutlp/vol10/iss2/2.

Kreijns, K., Kirschner, P. A., and Jochems, W. (2003). Identifying the Pitfalls for Social Interaction in Computer-supported Collaborative Learning Environments: A Review of the Research. *Computers in Human Behavior*, 19(3):335–353. DOI: 10.1016/S0747-5632(02)00057-2.

Kreijns, K., Van Acker, F., Vermeulen, M., and van Buuren, H. (2014). Community of Inquiry: Social Presence Revisited. *E-Learning and Digital Media [Special Issue: Inquiry into 'Communities of Inquiry': Knowledge, Communication, Presence, Community*, 11(1):5–19. DOI: 10.2304/elea.2014.11.1.5.

Latif, E. and Miles, S. (2013). Class Attendance and Academic Performance: A Panel Data Analysis. *Economic Papers*, 32(4):470–476. DOI: 10.1111/1759-3441.12054.

Law, E. L.-C. and Wild, F. (2015). A Multidimensional Evaluation Framework for Personal Learning Environments. In Kroop, S., Mikroyannidis, A., and Wolpers, M., editors, *Responsive Open Learning Environments: Outcomes of Research from the ROLE Project*, pages 49–77. Springer International Publishing. DOI: 10.1007/978-3-319-02399-1_3.

Long, P. and Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5):31–40. available online at https://net.educause.edu/ir/library/pdf/ERM1151.pdf.

Lonn, S., Aguilar, S., and Teasley, S. (2015). Investigating Student Motivation in the Context of a Learning Analytics Intervention During a Summer Bridge Program. *Computers in Human Behavior*, 47:90–97. DOI: 10.1016/j.chb.2014.07.013.

Lonn, S. and Teasley, S. D. (2014). Student Explorer: A Tool for Supporting Academic Advising at Scale. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 175–176, New York, NY, USA. ACM. DOI: 10.1145/2556325.2567867.

Louis, W. R., Bastian, B., McKimmie, B., and Lee, A. J. (2016). Teaching psychology in Australia: Does class attendance matter for performance. *Australian Journal of Psychology*, 68:47–51. DOI: 10.1111/ajpy.12088.

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

MacNeill, S., Campbell, L. M., and Hawksey, M. (2014). Analytics for Education. *Journal of Interactive Media in Education*, 1(62014):Art. 7. DOI: 10.5334/2014-07.

Marsh, H., Bella, J., and Hau, K. (1996). An Evaluation of Incremental Fit Indices: A Clarification of Mathematical and Empirical Properties. In Marcoulides, G. and Schumacker, R., editors, *Advanced structural equation modeling: Issues and techniques*, pages 315–353. Erlbaum, Mahwah, NJ, US.

Martynoga, B. (2015). Head of the class – Teaching methods are often based on convention over evidence. Can neuroscience help pupils to learn? *thelongandshort.org*. available online at http://thelongandshort.org/society/head-of-the-class-neuroscience-education.

McAlpine, L. and Weston, C. (2000). Reflection: issues related to improving professors' teaching and students' learning. *Instructional Science*, 28(5):363–385. DOI: 10.1023/A:1026583208230.

Méndez, G., Ochoa, X., and Chiluiza, K. (2014). Techniques for data-driven curriculum analysis. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, LAK '14, pages 148–157, New York, NY, USA. ACM. DOI: 10.1145/2567574.2567591.

Miyazoe, T. and Anderson, T. (2013). Interaction Equivalency in an OER, MOOCS and Informal Learning Era. *Journal of Interactive Media in Education*, 2013(2):Art. 9. DOI: 10.5334/2013-09.

Morgado, L., Mota, J., Quintas-Mendes, A., Fano, S., Fueyo, A., Tomasini, A., Giannatelli, A., Silva, A., Jansen, D., and Brouns, F. (2014). D2.2 instructional design and scenarios for moocs version 1. Deliverable, Elearning, Communication and Open-data (ECO) project.

Mory, E. H. (2004). Feedback Research Revisited. In Jonassen, D. H., editor, *Handbook of Research on Educational Communications and Technology*, pages 745–783. Lawrence Erlbaum Associates, Mahwah, NJ, US.

Norris, D. M. and Baer, L. L. (2013). Building Organizational Capacity for Analytics. *EDUCAUSE*. available online at https://net.educause.edu/ir/library/pdf/PUB9012.pdf.

Oxford Dictionaries (2017). https://www.oxforddictionaries.com. website.

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

Pardo, A. and Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3):438–450. DOI: 10.1111/bjet.12152.

Persico, D. and Pozzi, F. (2014). Informing learning design with learning analytics to improve teacher inquiry. *British Journal of Educational Technology*, 46(2):230–248. DOI: 10.1111/bjet.12207.

Phielix, C., Prins, F. J., Kirschner, P. A., Erkens, G., and Jaspers, J. (2011). Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool. *Computers in Human Behavior*, 27(3):1087–1102. DOI: 10.1016/j.chb.2010.06.024.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., and Kirschner, P. A. (2016). Investigating Collaborative Learning Success with Physiological Coupling Indices Based on Electrodermal Activity. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, LAK ’16, pages 64–73, New York, NY, USA. ACM. DOI: 10.1145/2883851.2883897.

Pintrich, P. R. (2000). Chapter 14 - The Role of Goal Orientation in Self-Regulated Learning. In Boekaerts, M., Pintrich, P. R., and Zeidner, M., editors, *Handbook of Self-Regulation*, pages 451–502. Academic Press, San Diego. DOI: 10.1016/B978-012109890-2/50043-3.

Prinsloo, P. and Slade, S. (2017). An Elephant in the Learning Analytics Room: The Obligation to Act. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge*, LAK ’17, pages 46–55, New York, NY, USA. ACM. DOI: 10.1145/3027385.3027406.

Puustinen, M. and Pulkkinen, L. (2001). Models of Self-regulated Learning: A review. *Scandinavian Journal of Educational Research*, 45(3):269–286. DOI: 10.1080/00313830120074206.

Reyes, J. A. (2015). The skinny on big data in education: Learning analytics simplified. *TechTrends*, 59(2):75–80. DOI: 10.1007/s11528-015-0842-1.

Rivera-Pelayo, V., Rodriguez-Triana, M. J., Petrushyna, Z., Braun, S., Loureiro, A., and Kawase, R. (2013). Special issue on learning analytics. *Journal of Technology Enhanced Learning*, 5(2).

Saadatdoost, R., Sim, A. T. H., Jafarkarimi, H., and Mei Hee, J. (2015). Exploring MOOC from education and Information Systems perspectives: a short literature review. *Educational Review*, 67(4):505–518. DOI: 10.1080/00131911.2015.1058748.

Santos, J., Verbert, K., and Duval, E. (2012). Empowering Students to Reflect on their Activity with StepUp!: Two Case Studies with Engineering Students. In *Proceedings of the 2nd workshop on Awareness and Reflection*, ARTEL'12. CEUR-WS.org.

Santos, J., Verbert, K., Govaerts, S., and Duval, E. (2013). Addressing learner issues with stepup!: An evaluation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 14–22, New York, NY, USA. ACM. DOI: 10.1145/2460296.2460301.

Scheffel, M., Drachsler, H., de Kraker, J., Kreijns, K., Slootmaker, A., and Specht, M. (2017a). Widget, widget on the wall, am I performing well at all? *IEEE Transactions on Learning Technologies*, 10(1):42–52. DOI: 10.1109/TLT.2016.2622268.

Scheffel, M., Drachsler, H., Kreijns, K., de Kraker, J., and Specht, M. (2017b). Widget, Widget As You Lead, I Am Performing Well Indeed!: Using Results from an Exploratory Offline Study to Inform an Empirical Online Study About a Learning Analytics Widget in a Collaborative Learning Environment. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge*, LAK '17, pages 289–298, New York, NY, USA. ACM. DOI: 10.1145/3027385.3027428.

Scheffel, M., Drachsler, H., and Specht, M. (2015). Developing an Evaluation Framework of Quality Indicators for Learning Analytics. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, LAK '15, pages 16–20, New York, NY, USA. ACM. DOI: 10.1145/2723576.2723629.

Scheffel, M., Drachsler, H., Stoyanov, S., and Specht, M. (2014). Quality Indicators for Learning Analytics. *Educational Technology & Society*, 17(4):117–132.

Scheffel, M., Drachsler, H., Toisoul, C., Ternier, S., and Specht, M. (2017c). The Proof of the Pudding: Examining Validity and Reliability of the Evaluation Framework for Learning Analytics. In *Proceedings of the 12th European Conference on Technology Enhanced Learning (EC-TEL 2017)*, LNCS, Berlin, Heidelberg. Springer.

Scheffel, M., Niemann, K., Leony, D., Pardo, A., Schmitz, H.-C., Wolpers, M., and Delgado Kloos, C. (2012). Key action extraction for learning analytics. In Ravenscroft, A., Lindstaedt, S., Delgado Kloos, C., and Hernández-Leo, D., editors, *21st Century Learning for 21st Century Skills: Proceedings of the 7th European Conference on Technology Enhanced Learning (EC-TEL 2012)*, volume 7563 of *LNCS*, pages 320–333, Berlin, Heidelberg. Springer. DOI: 10.1007/978-3-642-33263-0_25.

Scheffel, M., Niemann, K., Pardo, A., Leony, D., Friedrich, M., Schmidt, K., Wolpers, M., and Kloos, C. D. (2011). Usage Pattern Recognition in Student Activities. In Kloos, C. D., Gillet, D., Crespo García, R. M., Wild, F., and Wolpers, M., editors, *Towards Ubiquitous Learning: Proceedings of the 6th European Conference of Technology Enhanced Learning (EC-TEL 2011)*, volume 6964 of *LNCS*, pages 341–355, Berlin, Heidelberg. Springer. DOI: 10.1007/978-3-642-23985-4_27.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. Temple Smith, London, UK.

Sclater, N. (2016). Developing a Code of Practice for Learning Analytics. *Journal of Learning Analytics*, 3(1):16–52. DOI: 10.18608/jla.2016.31.3.

Sclater, N. and Bailey, P. (2015). Code of Practice for Learning Analytics. Guide, Jisc. available online at https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics.

Sharkey, M. and Harfield, T. (2016). Traversing the Trough of Disillusionment: Where do Analytics Go from Here? blog post, Next Generation Learning Challenges. available online at http://nextgenlearning.org/blog/traversing-trough-disillusionment-where-do-analytics-go-here.

Shelton, K. (2010). A Quality Scorecard for the Administration of Online Education Programs: A Delphi Study. *Journal of Asynchronous Learning Networks*, 14(4):36–62.

Shelton, K. and Pedersen, K. (2015). Benchmarking Quality in Online Learning Programs in Higher Education. In *Proceedings of Global Learn 2015*, pages 280–295, Berlin, Germany. Association for the Advancement of Computing in Education (AACE). available online at https://www.learntechlib.org/p/150872.

Siemens, G. (2011). 1st International Conference on Learning Analytics and Knowledge 2011 – About. conference website. available online at https://tekri.athabascau.ca/analytics/about.

Siemens, G. and Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge*, LAK '12, pages 252–254, New York, NY, USA. ACM. DOI: 10.1145/2330601.2330661.

Siemens, G., Dawson, S., and Lynch, G. (2013). Improving the quality and productivity of the higher education sector – policy and strategy for system-level deployment of learning analytics. Discussion paper for the Australian Government, Society for Learning Analytics Research (SoLAR). available online at https://pdfs.semanticscholar.org/d307/34019734c219f6ab83af3dcebfa39c9e8b84.pdf.

Slade, S. and Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioural Scientist*, 57(10):1509–1528. DOI: 10.1177/0002764213479366.

Slootmaker, A., Scheffel, M., Kreijns, K., De Kraker, J., and Drachsler, H. (2015). Performance dashboard to support awareness and reflection in elgg communities (version 1.15) [software]. available online at http://hdl.handle.net/1820/6290.

Stewart, M., Stott, T., and Nuttall, A.-M. (2011). Student Engagement Patterns over the Duration of Level 1 and Level 3 Geography Modules: Influences on Student Attendance, Performance and Use of Online resources. *Journal of Geography in Higher Education*, 35(1):47–65. DOI: 10.1080/03098265.2010.498880.

Strang, K. D. (2016). Beyond engagement analytics: which online mixed-data factors predict student learning outcomes? *Education and Information Technologies*, pages 1–21. DOI: 10.1007/s10639-016-9464-2.

Taibi, D. and Dietze, S. (2013). Fostering analytics on learning analytics research: the lak dataset. In d'Aquin, M., Dietze, S., Drachsler, H., Herder, E., and Taibi, D., editors, *LAK (Data Challenge)*, volume 974 of *CEUR Workshop Proceedings*. CEUR-WS.org. available online at http://ceur-ws.org/Vol-974/lakdatachallenge2013_preface.pdf.

Teasley, S. D. (2017). Student Facing Dashboards: One Size Fits All? *Technology, Knowledge and Learning*, pages 1–8. DOI: 10.1007/s10758-017-9314-3.

Tempelaar, D. T., Rienties, B., and Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47:157–167. DOI: 10.1016/j.chb.2014.05.038.

Touré-Tillery, M. and Fishbach, A. (2014). How to Measure Motivation: A Guide for the Experimental Social Psychologist. *Social and Personality Psychology Compass*, 8(7):328–341. DOI: 10.1111/spc3.12110.

UN Independent Expert Advisory Group (2014). A World That Counts – Mobilising the Data Revolution for Sustainable Development. Report, United Nations. available online at http://www.undatarevolution.org/report/.

van den Hoven, J. (2008). Information Technology, Privacy and the Protection of Personal Data. In van den Hoven, J. and Weckert, J., editors, *Information Technology and Moral Philosophy*, pages 301–321. Cambridge University Press, Cambridge, MA.

van Trigt, M. (2016). How Data Can Improve the Quality of Higher Education. White paper, SURFnet. available online at https://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2016/white paper-learning-analytics_en-def.pdf.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., and Santos, J. (2013). Learning Analytics Dashboard Applications. *American Behavioral Scientist*, 57(10):1500–1509. DOI: 10.1177/0002764213479363.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Assche, F., Parra, G., and Klerkx, J. (2014). Learning Dashboards: An Overview and Future Research Opportunities. *Personal Ubiquitous Computing*, 18(6):1499–1514. DOI: 10.1007/s00779-013-0751-2.

Verbert, K., Manouselis, N., Drachsler, H., and Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society*, 15(3):133–148.

Verpoorten, D. (2012). *Reflection Amplifiers in Self-regulated Learning*. PhD thesis, Open Universiteit, Heerlen, The Netherlands.

Verpoorten, D., Westera, W., and Specht, M. (2011). Infusing Reflective Practice in eLearning Courses – Can Widgets Help? *International Journal of Technology Enhanced Learning*, 3(1):93–109. DOI: 10.1504/IJTEL.2011.039066.

Willis, J., Campbell, J., and Pistilli, M. (2013). Ethics, Big Data, and Analytics: A Model for Application. *EDUCAUSE online*. available online at http://er.educause.edu/articles/2013/5/ethics-big-data-and-analytics-a-model-for-application.

Winne, P. and Hadwin, A. (1998). Studying as Self-regulated Learning. In Hacker, D. and Dunlosky, J., editors, *Metacognition in Educational Theory and Practice*, The Educational Psychology Series, pages 277–304. Erlbaum, Mahwah, NJ.

Winne, P. H. (1995). Inherent Details in Self-Regulated Learning. *Educational Psychologist*, 30(4):173–187. DOI: 10.1207/s15326985ep3004_2.

Winne, P. H. (2006). How Software Technologies Can Improve Research on Learning and Bolster School Reform. *Educational Psychologist*, 41(1):5–17. DOI: 10.1207/s15326985ep4101_3.

Wise, A. F. (2014). Designing Pedagogical Interventions to Support Student Use of Learning Analytics. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, LAK '14, pages 203–211, New York, NY, USA. ACM. DOI: 10.1145/2567574.2567588.

Zimmerman, B. J. (1995). Self-Regulation Involves More Than Metacognition: A Social Cognitive Perspective. *Educational Psychologist*, 30(4):217–221. DOI: 10.1207/s15326985ep3004_8.

Zimmerman, B. J. (2000). Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective. In Boekaerts, M., Pintrich, P. R., and Zeidner, M., editors, *Handbook of Self-Regulation*, pages 13–39. Academic Press, San Diego. DOI: 10.1016/B978-012109890-2/50031-7.

Zinn, C. and Scheuer, O. (2006). Getting to Know Your Student in Distance Learning Contexts. In Nejdl, W. and Tochtermann, K., editors, *Innovative Approaches for Learning and Knowledge Sharing: Proceedings of the first European Conference on Technology Enhanced Learning*, EC-TEL'06, pages 437–451, Berlin, Heidelberg. Springer. DOI: 10.1007/11876663_34.

Appendices

Appendix A

List of 103 statements from the group concept mapping study

1. full access to the used data behind a study
2. the effectiveness for learning versus cognitive overload of the LA tools / dashboard
3. that student modelling into learning styles is possible
4. to allow instructors to support decisions as a data-based system
5. that teachers are active
6. if learners can influence how data are provided
7. to detect that students are using the learning tool in the wrong way
8. the portability of the LA results
9. the portability of the collected data
10. that teachers are sufficiently trained to use LA tools in the right manner
11. that administrators invest in scaling successful tools across their programming
12. the transparency of the used data
13. if the learning analytics tools provide high-level information (e.g. information on cognitive learning activities instead of low-level interaction log data)
14. the robustness against manipulations from attacks, manipulation and fraud
15. if students can get a report on higher levels of academically-purposeful behavior
16. if faculty agree that the provided information is accurate
17. the time spent in the learning experience
18. the comparison of a generic educational quality metric (learning gains, standardized tests, drop-out rates) before and after the inclusion of learning analytics
19. if the learning analytics tools can adapt to the learner's "understanding competence" (they should visualise in a way that the learner can understand the presented information)
20. that student retention is increasing year by year after the introduction of LA systems
21. students' success in terms of aroused interest
22. interventions can be tracked and assessed for accuracy
23. if teachers are able to gain new insights using the given LA methods

24. if students agree that the provided information is accurate
25. that students are sufficiently trained to use LA tools in the right manner
26. to compare learning performance during a course and the academic background of students in order to better identify course prerequisites
27. that students take the learning process more serious when they know their performance is measured
28. that LA users suggest how to improve their learning
29. if students agree that the provided information is useful
30. students' success in terms of grades
31. an opt-out option for data subjects to remove all their data
32. that the LA tool uses a specific data standard
33. if faculty reflect on how their students are learning and alter instruction methods/content delivery to help students attain learning objectives
34. when teachers find the work needed to use the tools worth while
35. the affective quality of the analytics (how able are the analytics to stipulate a positive user experience)
36. the level of control over the collected data
37. that teachers take action in certain situations (if the analysed data shows problems of a learner)
38. to improve the structure of a course based on time-based student performances
39. the ability to explain what could help to further improve
40. the ability to explain what went well
41. tracking the use of resources outside the class that students use to complement the resources provided by the instructor for a class
42. that teachers react in a more personalized way to how their students are dealing with learning material
43. transparency of the used algorithm
44. when no risk issues are left
45. the visualization of personal data to the data subject
46. dependent of the context and objective of implementing learning analytics in that context
47. tracking the efficacy of help resources to ensure that when students are seeking assistance, they are receiving the assistance they sought

48. that change in workplace learning is measurable
49. that methods are open access
50. that teachers are motivated
51. tracking help-seeking behavior of students when they are directed to resources to help them maintain or improve their grades
52. that students are motivated
53. if parents accept the results as useful
54. institutional culture changes such that people readily "buy-in" to using data to forward teaching, learning, and institutional practice
55. to be able to differentiate whether the students' behaviour is good or bad for learning
56. that students regularly utilize the tools provided.
57. that students demonstrate gains on measures of student learning even after controlling for powerful covariates (such as, prior performance and engagement)
58. the ability to explain what went wrong
59. that teachers change their instructional design according to LA results during a course
60. to provide models of how students who successfully complete a course using resources made available to them
61. that teachers are engaged
62. how easily stakeholders understand the affordances of the intervention
63. that privacy is ensured
64. if teachers accept the results as useful
65. to be transparent what the teacher knows about the students and how he uses this information
66. an early detection of students at risk
67. that students are engaged
68. that LA users start to talk about their performance
69. that students change their learning behaviour based on the provided information from learning analytics tools
70. improved student achievement
71. the comparability of different methods

72. the level of adoption of LA processes and techniques in informal educational settings
73. that the user can adapt the LA display to meet his / her needs
74. that change in workplace learning is sustainable
75. that students are more motivated to control their learning process
76. that LA results are compared with other (traditional) measures
77. if students use the learning analytics tools often during their learning process (during knowlege acquisition)
78. students' success in terms of satisfaction
79. that teachers recommend other teachers to use LA tools
80. to compare learning performance during a course and the academic background of students in order to enhance learning resources for a heterogeneous cohort
81. that data are open access
82. that the cost of the courses decrease strongly following optimization
83. that teachers change their behaviour in some aspects
84. that the learning outcome of students increase if they use learning analytics tools
85. a reference data source for the whole LAK community
86. that students recommend other students utilize the LA tools
87. if learners can influence which data are provided
88. tracking the efficacy of help resources to ensure that their problems/issues were addressed by said resource
89. that students do not drop out
90. that students are more aware of their learning progress
91. interventions can be tracked and assessed for usefulness
92. that students change their behaviour in some aspects
93. that users can download their own data
94. if the learning analytics tools can distinguish between cognitive and metacognitive learner actions to be presented
95. giving feedback to teachers about the learning resources they provide their students with
96. the evidence of intended changes in student behaviours

97. that the tool is integrated in the stakeholder community’s work practice
98. that students are active
99. that students compare their learning process with peers
100. that there is uptake of LA in the business world (as new tool or feature in existing product)
101. the level of adoption of LA processes and techniques in formal educational settings
102. the extent to which the achievement of learning objectives can be demonstrated
103. that students become more self-regulated in their learning processes

Appendix B

List of eight clusters and their statements from the group concept mapping study

Cluster 1: Data: open access

	Statement	Bridging
	cluster as a whole	*0.25*
1.	full access to the used data behind a study	0.07
8.	the portability of the LA results	0.32
9.	the portability of the collected data	0.30
12.	the transparency of the used data	0.06
14.	the robustness against manipulations from attacks, manipulation and fraud	0.38
32.	that the LA tool uses a specific data standard	0.19
43.	transparency of the used algorithm	0.22
44.	when no risk issues are left	0.60
49.	that methods are open access	0.23
81.	that data are open access	0.07
85.	a reference data source for the whole LAK community	0.29

Count	Std. Dev.	Variance	Min	Max	Average	Median
11	0.15	0.02	0.06	0.60	0.25	0.19

Cluster 2: Data: privacy

	Statement	Bridging
	cluster as a whole	*0.31*
6.	if learners can influence how data are provided	0.72
31.	an opt-out option for data subjects to remove all their data	0.15
36.	the level of control over the collected data	0.15
45.	the visualization of personal data to the data subject	0.41
63.	that privacy is ensured	0.10
65.	to be transparent what the teacher knows about the students and how he uses this information	0.44
87.	if learners can influence which data are provided	0.35
93.	that users can download their own data	0.14

Count	Std. Dev.	Variance	Min	Max	Average	Median
8	0.20	0.04	0.10	0.72	0.31	0.26

	Cluster 3: Acceptance & uptake	
	Statement	Bridging
	cluster as a whole	*0.86*
2.	the effectiveness for learning versus cognitive overload of the LA tools / dashboard	0.90
3.	that student modelling into learning styles is possible	1.00
10.	that teachers are sufficiently trained to use LA tools in the right manner	0.94
11.	that administrators invest in scaling successful tools across their programming	0.98
46.	dependent of the context and objective of implementing learning analytics in that context	0.93
53.	if parents accept the results as useful	0.66
54.	institutional culture changes such that people readily "buy-in" to using data to forward teaching, learning, and institutional practice	0.85
62.	how easily stakeholders understand the affordances of the intervention	0.69
71.	the comparability of different methods	0.77
72.	the level of adoption of LA processes and techniques in informal educational settings	0.80
97.	that the tool is integrated in the stakeholder community's work practice	0.90
100.	that there is uptake of LA in the business world (as new tool or feature in existing product)	0.99
101.	the level of adoption of LA processes and techniques in formal educational settings	0.82

Count	Std. Dev.	Variance	Min	Max	Average	Median
13	0.11	0.01	0.66	1.00	0.86	0.85

Cluster 4: Learning outcome

	Statement	Bridging
	cluster as a whole	*0.46*
13.	if the learning analytics tools provide high-level information (e.g. information on cognitive learning activities instead of low-level interaction log data)	0.42
16.	if faculty agree that the provided information is accurate	0.56
18.	the comparison of a generic educational quality metric (learning gains, standardized tests, drop-out rates) before and after the inclusion of learning analytics	0.48
19.	if the learning analytics tools can adapt to the learner's "understanding competence" (they should visualise in a way that the learner can understand the presented information)	0.51
23.	if teachers are able to gain new insights using the given LA methods	0.19
25.	that students are sufficiently trained to use LA tools in the right manner	0.87
35.	the affective quality of the analytics (how able are the analytics to stipulate a positive user experience)	0.44
50.	that teachers are motivated	0.19
61.	that teachers are engaged	0.20
73.	that the user can adapt the LA display to meet his / her needs	0.55
76.	that LA results are compared with other (traditional) measures	0.53
94.	if the learning analytics tools can distinguish between cogntive and meta-cognitive learner actions to be presented	0.75
95.	giving feedback to teachers about the learning resources they provide their students with	0.24

Count	Std. Dev.	Variance	Min	Max	Average	Median
13	0.20	0.04	0.19	0.87	0.46	0.44

Cluster 5: Teacher awareness

	Statement	Bridging
	cluster as a whole	*0.41*
4.	to allow instructors to support decisions as a data-based system	0.58
5.	that teachers are active	0.49
33.	if faculty reflect on how their students are learning and alter instruction methods/content delivery to help students attain learning objectives	0.35
34.	when teachers find the work needed to use the tools worth while	0.37
37.	that teachers take action in certain situations (if the analysed data shows problems of a learner)	0.22
42.	that teachers react in a more personalized way to how their students are dealing with learning material	0.25
59.	that teachers change their instructional design according to LA results during a course	0.32
64.	if teachers accept the results as useful	0.30
74.	that change in workplace learning is sustainable	0.73
79.	that teachers recommend other teachers to use LA tools	0.44
82.	that the cost of the courses decrease strongly following optimization	0.70
83.	that teachers change their behaviour in some aspects	0.18

Count	Std. Dev.	Variance	Min	Max	Average	Median
12	0.17	0.03	0.18	0.73	0.41	0.29

Cluster 6: Learning performance

	Statement	Bridging
	cluster as a whole	*0.31*
17.	the time spent in the learning experience	0.52
26.	to compare learning performance during a course and the academic background of students in order to better identify course prerequisites	0.11
28.	that LA users suggest how to improve their learning	0.43
38.	to improve the structure of a course based on time-based student performances	0.16
48.	that change in workplace learning is measurable	0.22
68.	that LA users start to talk about their performance	0.59
80.	to compare learning performance during a course and the academic background of students in order to enhance learning resources for a heterogeneous cohort	0.17
102.	the extent to which the achievement of learning objectives can be demonstrated	0.31

Count	Std. Dev.	Variance	Min	Max	Average	Median
8	0.17	0.03	0.11	0.59	0.31	0.19

Cluster 7: Learning support

	Statement	Bridging
	cluster as a whole	*0.45*
7.	to detect that students are using the learning tool in the wrong way	0.16
15.	if students can get a report on higher levels of academically-purposeful behavior	0.76
22.	interventions can be tracked and assessed for accuracy	0.55
24.	if students agree that the provided information is accurate	0.73
29.	if students agree that the provided information is useful	0.64
39.	the ability to explain what could help to further improve	0.22
40.	the ability to explain what went well	0.27
41.	tracking the use of resources outside the class that students use to complement the resources provided by the instructor for a class	0.59
47.	tracking the efficacy of help resources to ensure that when students are seeking assistance, they are receiving the assistance they sought	0.53
51.	tracking help-seeking behavior of students when they are directed to resources to help them maintain or improve their grades	0.63
55.	to be able to differentiate whether the students' behaviour is good or bad for learning	0.38
56.	that students regularly utilize the tools provided.	0.59
58.	the ability to explain what went wrong	0.31
60.	to provide models of how students who successfully complete a course using resources made available to them	0.36
66.	an early detection of students at risk	0.14
77.	if students use the learning analytics tools often during their learning process (during knowledge acquisition)	0.46
88.	tracking the efficacy of help resources to ensure that their problems/issues were addressed by said resource	0.40
91.	interventions can be tracked and assessed for usefulness	0.44

Count	Std. Dev.	Variance	Min	Max	Average	Median
18	0.18	0.03	0.14	0.76	0.45	0.58

Cluster 8: Student awareness

	Statement	Bridging
	cluster as a whole	*0.11*
20.	that student retention is increasing year by year after the introduction of LA systems	0.10
21.	students' success in terms of aroused interest	0.01
27.	that students take the learning process more serious when they know their performance is measured	0.06
30.	students' success in terms of grades	0.01
52.	that students are motivated	0.13
57.	that students demonstrate gains on measures of student learning even after controlling for powerful covariates (such as, prior performance and engagement)	0.02
67.	that students are engaged	0.35
69.	that students change their learning behaviour based on the provided information from learning analytics tools	0.08
70.	improved student achievement	0.00
75.	that students are more motivated to control their learning process	0.11
78.	students' success in terms of satisfaction	0.03
84.	that the learning outcome of students increase if they use learning analytics tools	0.04
86.	that students recommend other students utilize the LA tools	0.43
89.	that students do not drop out	0.06
90.	that students are more aware of their learning progress	0.12
92.	that students change their behaviour in some aspects	0.08
96.	the evidence of intended changes in student behaviours	0.13
98.	that students are active	0.19
99.	that students compare their learning process with peers	0.18
103.	that students become more self-regulated in their learning processes	0.05

Count	Std. Dev.	Variance	Min	Max	Average	Median
20	0.11	0.01	0.00	0.43	0.11	0.07

Appendix C

Go-zone graphs of all eight clusters of the group concept mapping study

Figure B1: Go-zone graph of cluster 1 *Data: open access*

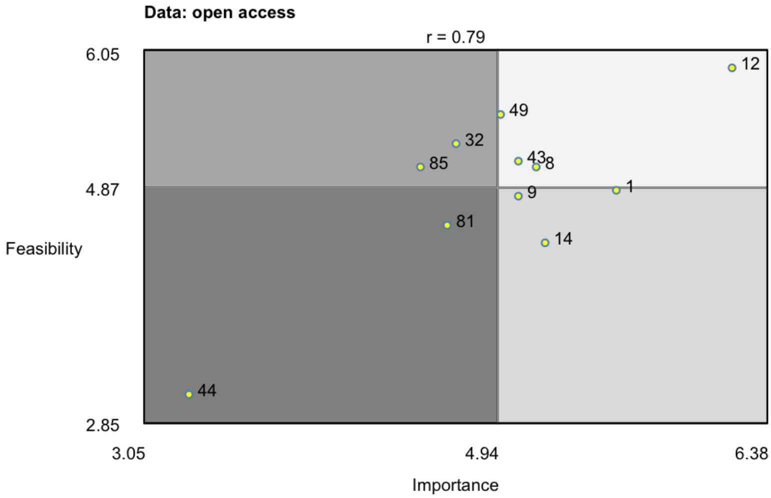


Figure B2: Go-zone graph of cluster 2 *Data: privacy*

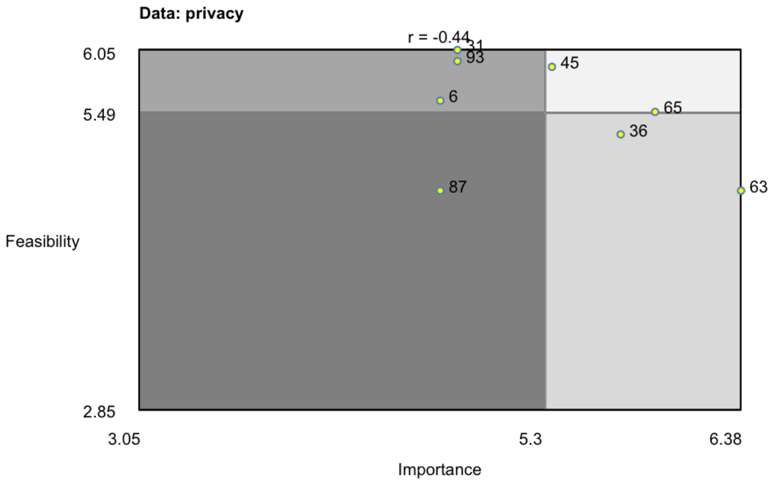


Figure B3: Go-zone graph of cluster 3 *Acceptance & uptake*

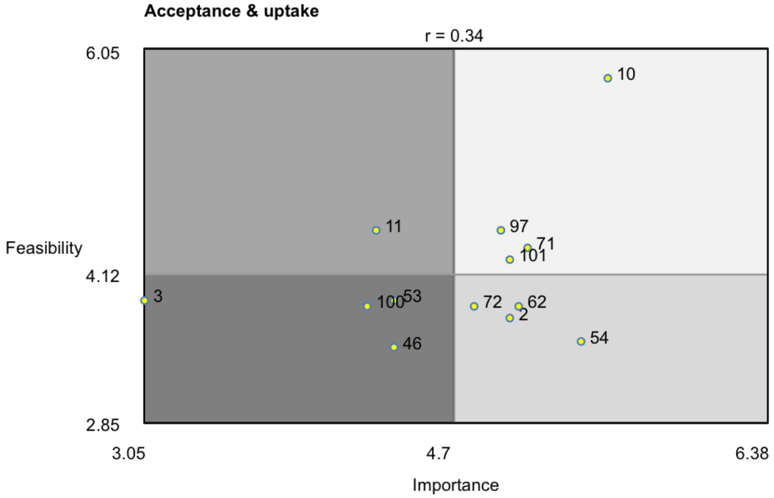


Figure B4: Go-zone graph of cluster 4 *Learning outcome*

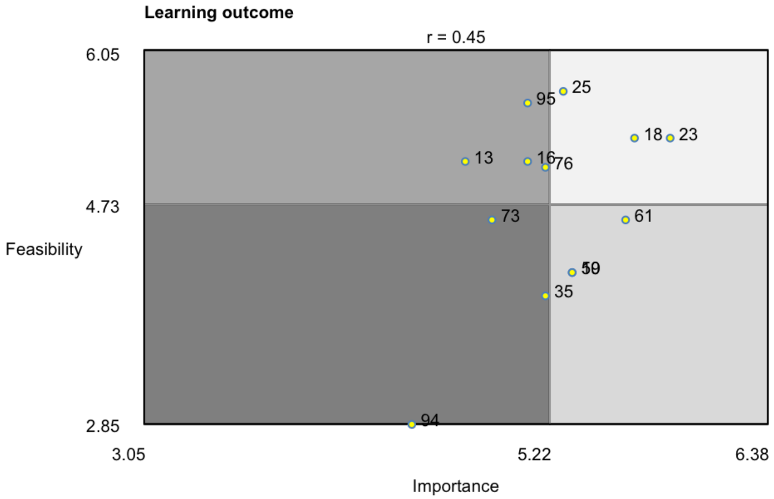


Figure B5: Go-zone graph of cluster 5 *Teacher awareness*

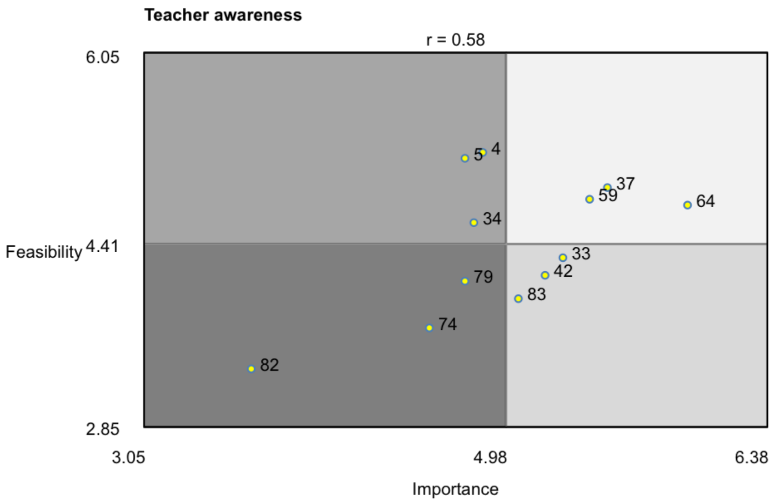


Figure B6: Go-zone graph of cluster 6 *Learning performance*

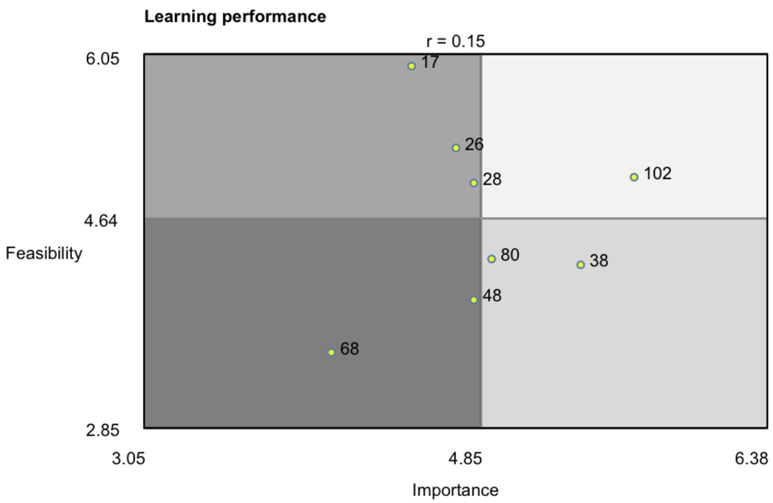


Figure B7: Go-zone graph of cluster 7 *Learning support*

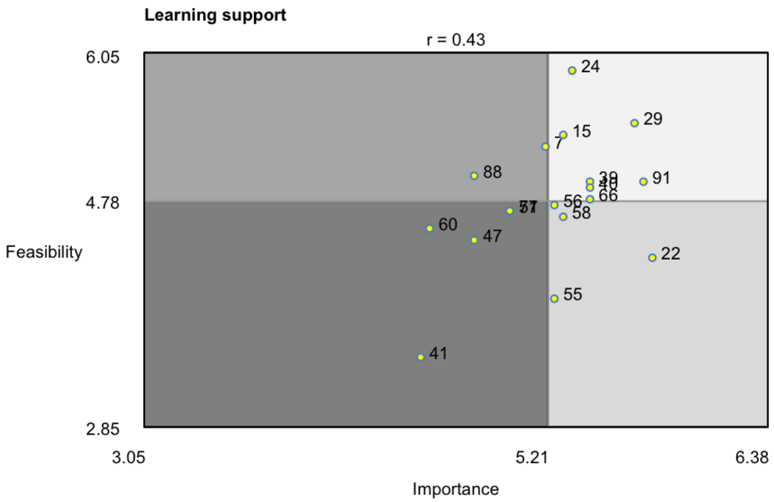
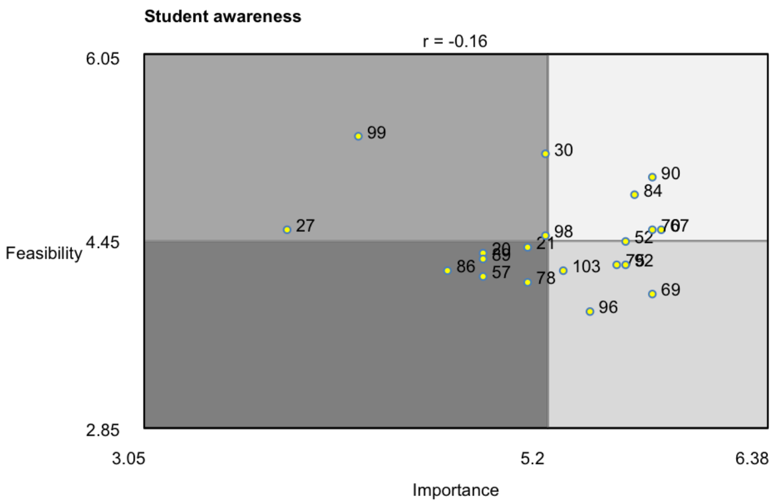


Figure B8: Go-zone graph of cluster 8 *Student awareness*



Appendix D

List of the four clusters and their statements for the framework construction study

Cluster 1: Data Aspects

	Statement	Bridging
	cluster as a whole	*0.27*
1.	full access to the used data behind a study	0.07
6.	if learners can influence how data are provided	0.72
8.	the portability of the LA results	0.32
9.	the portability of the collected data	0.30
12.	the transparency of the used data	0.06
14.	the robustness against manipulations from attacks, manipulation and fraud	0.38
31.	an opt-out option for data subjects to remove all their data	0.15
32.	that the LA tool uses a specific data standard	0.19
36.	the level of control over the collected data	0.15
43.	transparency of the used algorithm	0.22
44.	when no risk issues are left	0.60
45.	the visualization of personal data to the data subject	0.41
49.	that methods are open access	0.23
63.	that privacy is ensured	0.10
65.	to be transparent what the teacher knows about the students and how he uses this information	0.44
81.	that data are open access	0.07
85.	a reference data source for the whole LAK community	0.29
87.	if learners can influence which data are provided	0.35
93.	that users can download their own data	0.14

Count	Std. Dev.	Variance	Min	Max	Average	Median
19	0.18	0.03	0.06	0.72	0.27	0.22

Cluster 2: Impact & Integration

	Statement	Bridging
	cluster as a whole	*0.63*
2.	the effectiveness for learning versus cognitive overload of the LA tools / dashboard	0.90
3.	that student modelling into learning styles is possible	1.00
4.	to allow instructors to support decisions as a data-based system	0.58
10.	that teachers are sufficiently trained to use LA tools in the right manner	0.94
11.	that administrators invest in scaling successful tools across their programming	0.98
13.	if the learning analytics tools provide high-level information (e.g. information on cognitive learning activities instead of low-level interaction log data)	0.42
16.	if faculty agree that the provided information is accurate	0.56
18.	the comparison of a generic educational quality metric (learning gains, standardized tests, drop-out rates) before and after the inclusion of learning analytics	0.48
19.	if the learning analytics tools can adapt to the learner's "understanding competence" (they should visualise in a way that the learner can understand the presented information)	0.51
23.	if teachers are able to gain new insights using the given LA methods	0.19
25.	that students are sufficiently trained to use LA tools in the right manner	0.87
34.	when teachers find the work needed to use the tools worth while	0.37
35.	the affective quality of the analytics (how able are the analytics to stipulate a positive user experience)	0.44
46.	dependent of the context and objective of implementing learning analytics in that context	0.93
50.	that teachers are motivated	0.19
53.	if parents accept the results as useful	0.66
54.	institutional culture changes such that people readily "buy-in" to using data to forward teaching, learning, and institutional practice	0.85
61.	that teachers are engaged	0.20
62.	how easily stakeholders understand the affordances of the intervention	0.69
64.	if teachers accept the results as useful	0.30
71.	the comparability of different methods	0.77
72.	the level of adoption of LA processes and techniques in informal educational settings	0.80
73.	that the user can adapt the LA display to meet his / her needs	0.55
76.	that LA results are compared with other (traditional) measures	0.53
94.	if the learning analytics tools can distinguish between cognitive and meta-cognitive learner actions to be presented	0.75
95.	giving feedback to teachers about the learning resources they provide their students with	0.24

Cluster 2: Impact & Integration (continued)

	Statement	Bridging
97.	that the tool is integrated in the stakeholder community's work practice	0.90
100.	that there is uptake of LA in the business world (as new tool or feature in existing product)	0.99
101.	the level of adoption of LA processes and techniques in formal educational settings	0.82

Count	Std. Dev.	Variance	Min	Max	Average	Median
29	0.26	0.07	0.19	1.00	0.63	0.19

Cluster 3: Teacher Aspects

	Statement	Bridging
	cluster as a whole	*0.36*
5.	that teachers are active	0.49
17.	the time spent in the learning experience	0.52
26.	to compare learning performance during a course and the academic background of students in order to better identify course prerequisites	0.11
28.	that LA users suggest how to improve their learning	0.43
33.	if faculty reflect on how their students are learning and alter instruction methods/content delivery to help students attain learning objectives	0.35
37.	that teachers take action in certain situations (if the analysed data shows problems of a learner)	0.22
38.	to improve the structure of a course based on time-based student performances	0.16
42.	that teachers react in a more personalized way to how their students are dealing with learning material	0.25
48.	that change in workplace learning is measurable	0.22
59.	that teachers change their instructional design according to LA results during a course	0.32
68.	that LA users start to talk about their performance	0.59
74.	that change in workplace learning is sustainable	0.73
79.	that teachers recommend other teachers to use LA tools	0.44
80.	to compare learning performance during a course and the academic background of students in order to enhance learning resources for a heterogeneous cohort	0.17
82.	that the cost of the courses decrease strongly following optimization	0.70
83.	that teachers change their behaviour in some aspects	0.18
102.	the extent to which the achievement of learning objectives can be demonstrated 0.31	

Count	Std. Dev.	Variance	Min	Max	Average	Median
17	0.19	0.03	0.11	0.73	0.36	0.22

Cluster 4: Learner Aspects

	Statement	Bridging
	cluster as a whole	*0.27*
7.	to detect that students are using the learning tool in the wrong way	0.16
15.	if students can get a report on higher levels of academically-purposeful behavior	0.76
20.	that student retention is increasing year by year after the introduction of LA systems	0.10
21.	students' success in terms of aroused interest	0.01
22.	interventions can be tracked and assessed for accuracy	0.55
24.	if students agree that the provided information is accurate	0.73
27.	that students take the learning process more serious when they know their performance is measured	0.06
29.	if students agree that the provided information is useful	0.64
30.	students' success in terms of grades	0.01
39.	the ability to explain what could help to further improve	0.22
40.	the ability to explain what went well	0.27
41.	tracking the use of resources outside the class that students use to complement the resources provided by the instructor for a class	0.59
47.	tracking the efficacy of help resources to ensure that when students are seeking assistance, they are receiving the assistance they sought	0.53
51.	tracking help-seeking behavior of students when they are directed to resources to help them maintain or improve their grades	0.63
52.	that students are motivated	0.13
55.	to be able to differentiate whether the students' behaviour is good or bad for learning	0.38
56.	that students regularly utilize the tools provided.	0.59
57.	that students demonstrate gains on measures of student learning even after controlling for powerful covariates (such as, prior performance and engagement)	0.02
58.	the ability to explain what went wrong	0.31
60.	to provide models of how students who successfully complete a course using resources made available to them	0.36
66.	an early detection of students at risk	0.14
67.	that students are engaged	0.35
69.	that students change their learning behaviour based on the provied information from learning analytics tools	0.08
70.	improved student achievement	0.00
75.	that students are more motivated to control their learning process	0.11
77.	if students use the learning analytics tools often during their learning process (during knowledge acquisition)	0.46
78.	students' success in terms of satisfaction	0.03
84.	that the learning outcome of students increase if they use learning analytics tools	0.04
86.	that students recommend other students utilize the LA tools	0.43
88.	tracking the efficacy of help resources to ensure that their problems/issues were addressed by said resource	0.40

Cluster 4: Learner Aspects (continued)

	Statement	Bridging
89.	that students do not drop out	0.06
90.	that students are more aware of their learning progress	0.12
91.	interventions can be tracked and assessed for usefulness	0.44
92.	that students change their behaviour in some aspects	0.08
96.	the evidence of intended changes in student behaviours	0.13
98.	that students are active	0.19
99.	that students compare their learning process with peers	0.18
103.	that students become more self-regulated in their learning processes	0.05

Count	Std. Dev.	Variance	Min	Max	Average	Median
38	0.23	0.05	0.00	0.76	0.27	0.33

Appendix E

Go-zone graphs of the four clusters of the framework construction study

Figure E1: Go-zone graph of cluster 1 *Data aspects*

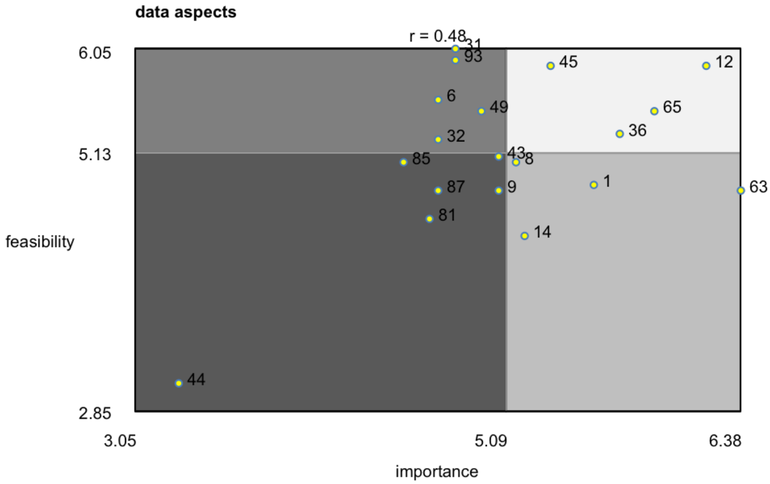


Figure E2: Go-zone graph of cluster 2 *Impact & integration*

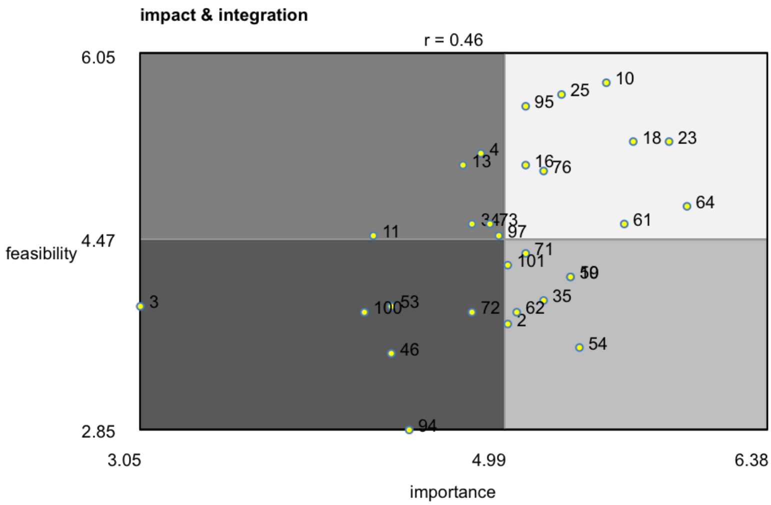


Figure E3: Go-zone graph of cluster 3 *Teacher aspects*

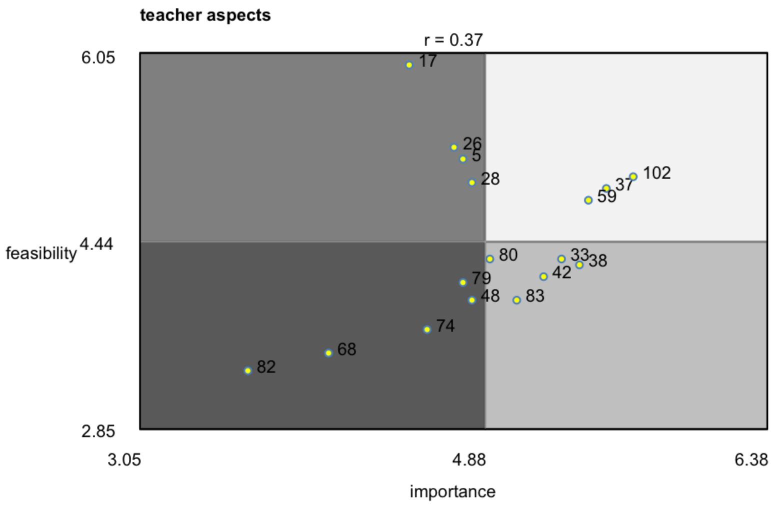
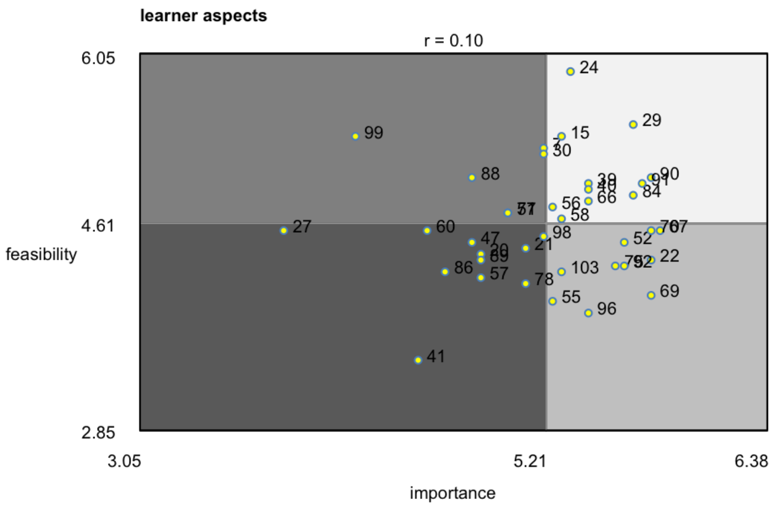


Figure E4: Go-zone graph of cluster 4 *Learner Aspects*



Appendix F

Experimental script used in the ECO lab study

Introduction

Welcome to this experiment! Today I will show you several learning analytics widgets that are being used by the ECO project and will ask you to evaluate them from either a student's or a teacher's point of view.

First, I will shortly tell you a bit about the ECO project and its learning analytics dashboard in general before explaining several widgets to you in more detail. After each detailed widget explanation, I will ask you to fill in the EFLA for that specific widget (either from a student's or from a teacher's point of view).

The EFLA is an evaluation instrument that contains 10 statements that can be rated on a scale from 1 for no agreement to 10 for high agreement. This allows us to then compare the different widgets' roles and their impact on the users with one another.

Before we continue, I ask you to fill in the Informed Consent form. By signing this form you declare that the details about this evaluation have been explained to you and that know that you can ask questions at any time and that you can withdraw at any time.

The ECO Project and its dashboard

ECO is a European project based on Open Educational Resources (OER) that gives free access to a list of MOOCs (Massive Open Online Courses) in 6 languages. The main goal of this project is to broaden access to education and to improve the quality and cost-effectiveness of teaching and learning in Europe. As an ECO user, you can either take part in MOOCs or create your own or both! Project funded from the European Community's CIP (Programme under grant agreement no 621127).

SHOW IMAGE (see Figure F1 below)

As part of the ECO platform, a learning analytics dashboard containing several visualisations has been developed to support the ECO users. The visualisations are based on interaction data of the users with the platform All users of the portal, i.e. the students as well as the teachers of the MOOCs, see the same visualisations.

SHOW IMAGE (see Figure F2 below)

There are several menu points. A user can choose to look at data relating to the ECO platform in general or at data relating to a specific MOOC.

SHOW IMAGE (see Figure F3 below)

Evaluation

We are now going to look at some visualisations in more detail. More specifically, I will show you two widgets and then ask you to evaluate them using the EFLA. I will then show you adapted versions of those widgets and ask you to evaluate them again.

Activity Widget Version 1

This widget shows how active each learner was in the MOOC according to the number of actions done in the MOOC. The x axis shows the position of a user within the MOOC. By hovering over the graph you can see more information about the different positions. In this screenshot for example, the learner in position 845 has done 146 activities. Students as well as teachers of a MOOC see the same visualisation.

SHOW IMAGE (see Figure 6.1)

Please go to the survey page now and fill in the EFLA about the Activity Widget Version 1. While doing so, please assume the role of a student/teacher. If you would like to add any comments, you can do so at the end of the section. Please let me know when you have finished evaluating the Activity Widget Version 1 and I will show you the next widget.

Resources Widget Version 1

This widget shows you what types of resources are present in this course and how often all users together have accessed the various resources in the MOOC. The length of the bar indicates the frequency of accesses which is also given as number at the end of the bar. Students as well as teachers of a MOOC see the same visualisation.

SHOW IMAGE (see Figure 6.2)

Please go to the next section on the survey page now and fill in the EFLA about the Resources Widget Version 1. While doing so, please assume the role of a student/teacher. If you would like to add any comments, you can do so at the end of the section. Please let me know when you have finished evaluating the Resources Widget Version 1 and I will show you the next widget.

Activity Widget Version 2

This widget shows the total activity per user. A user's activity is highlighted in red. With the radio buttons they can choose the type of clustering used in the visualisation. They can choose between the Median with quartiles and an Artificial Intelligence algorithm that creates four clusters. Teachers could use this tool to cluster feedback to groups of users. Learners can use the tool to see themselves in relation to their peers. Users can use the information buttons to get more information. Students as well as teachers of a MOOC see the same visualisation.

SHOW IMAGE and LIVE VIEW (see Figure 6.3)

Please go to the survey page now and fill in the EFLA about the Activity Widget Version 2. While doing so, please assume the role of a student/teacher. If you would like to add any comments, you can do so at the end of the section. Please let me know when you have finished evaluating the Activity Widget Version 2 and I will show you the next widget.

Resources Widget Version 2

This visualisation compares a user's MOOC path with the ideal path of the course and the paths of other participants. To load the paths of the other participants they can use the "Load all the other student paths". With the slider they can filter more or less students, where the more active students are positioned right on the slider. They can use the zoom buttons to zoom in or out on the graph. On the x-axis a user can see which activities have been accessed (green) and which not (red). The icons are indicating the type of activity. Teachers could use this tool to identify if learners are using the MOOC as planned by discovering if activities are accessed too early, too late, or not at all. A student could compare themselves to other students. All users can use the information buttons to get more information. Students as well as teachers of a MOOC see the same visualisation.

SHOW IMAGE and LIVE VIEW (see Figure 6.4)

Please go to the survey page now and fill in the EFLA about the Resources Widget Version 2. While doing so, please assume the role of a student/teacher. If you would like to add any comments, you can do so at the end of the section. After that please continue to the demographics section of the survey.

THANK YOU!

Figures used in the experimental script

Figure F1: Screenshot of the ECO website

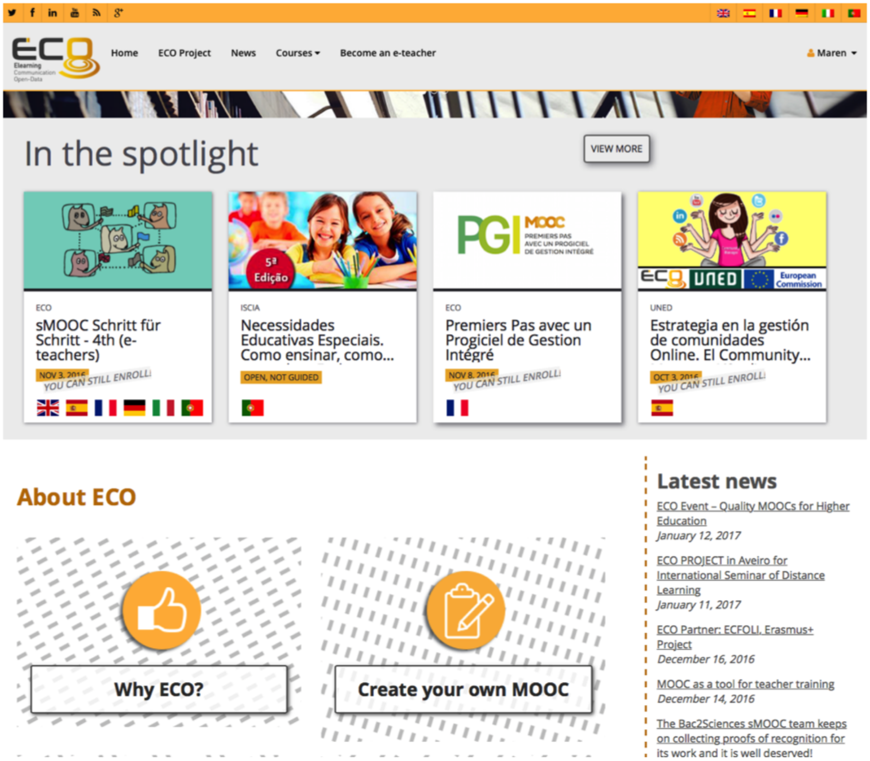


Figure F2: Screenshot of the ECO learning analytics dashboard

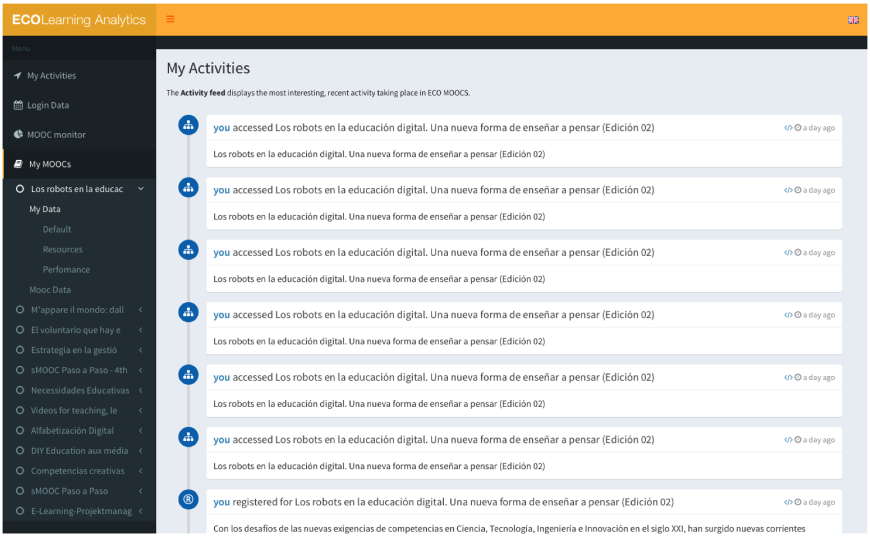
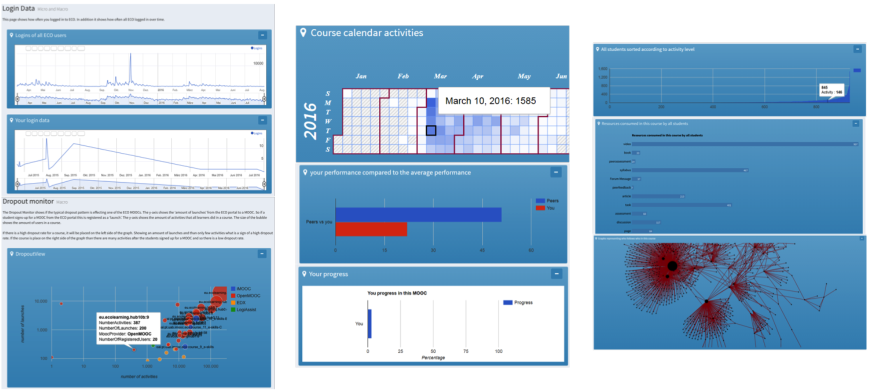


Figure F3: Example visualisations from the ECO learning analytics dashboard



List of Tables

List of Figures

Summary

Becoming its own distinct research field in 2011, the expectations for learning analytics to solve the retention problem, to increase student success and to support learning and teaching processes were very high. However, a number of years later now, empirical evidence as to whether learning analytics tools have the desired effect or not is still very sparse and there is still a gap between the potential of learning analytics identified by research and how much of this potential has been achieved so far (Ferguson et al., 2016a). Using the System Usability Scale (SUS) (Brooke, 1996) as an inspiration, the research reported in this thesis describes the continuous process of iteratively creating, using, evaluating and improving the evaluation framework for learning analytics (EFLA).

The framework addresses the current lack of evaluation instruments by offering a standardised way to evaluate learning analytics tools and to measure and compare the impact of learning analytics on educational practices. Similar to the SUS, the EFLA makes use of the subjective assessments about learning analytics tools by their users in order to obtain a general indication of the overall quality of a tool in a quick and simple, yet thoroughly developed, validated and reliable way. The main objectives of the research presented in this thesis therefore are threefold: (1) identify quality indicators for learning analytics, (2) create an applicable evaluation instrument based on these indicators, and (3) validate the evaluation instrument.

The thesis approaches these objectives in three distinct parts. In Part I, based on the input from the learning analytics community, a group concept mapping study is conducted and its results are used to devise the first version of the framework. Further input from learning analytics experts as well as a literature review is then used to evaluate and revise the EFLA to create the second EFLA version. Part II then deals with the questions whether a learning analytics widget can influence the students' grades in a course and whether the EFLA can be used to evaluate such a widget at several points in time and to reflect differences in perception between the two stakeholder groups. The widget evaluation as well as the results from an experts focus group are then used to evaluate the EFLA and to create the third EFLA version. Finally, Part III of the thesis shows that the EFLA can also be used to measure changes between different versions of widgets as well as differences between the two stakeholder groups and presents the validity and reliability analyses that leads to the creation of the fourth and final version of the EFLA.

During the EFLA's evaluation and improvement cycles, different course contexts (collaborative formal course vs. informal MOOC) as well as different widgets (activity widget for collaborative learning settings vs. activity and resources widget for an

individual learning setting) are used in addition to different study procedures (using the EFLA to evaluate the same widget at different points in time vs. using the EFLA to evaluate different widgets at the same time) as well as different study settings (live study vs. lab study). With its validity and reliability established, the framework is ready to close the gap between an institution's measurement of learning analytics readiness and its learning analytics maturity. It can complement the evaluation performed on an institutional level by providing input directly from the learners and teachers, thus allowing to create a holistic view of the learning and teaching processes involved. Furthermore, it also addresses the need for evidence of impact that the field has been calling for.

The learning analytics community now has the opportunity to verify the EFLA's applicability and benefit. Just as the System Usability Scale (SUS) was able to fully claim its success after being picked up and used by the research community (Brooke, 2013), usage of the EFLA in the field and on a larger scale will be needed to fully exploit its potential. The ready-for-download questionnaire templates[11] as well as the scoring spreadsheet[12] ease the adoption of the evaluation framework and allow for a similarly easy and 'quick and dirty' evaluation as the SUS.

But in order to 'go big' two aspects will have to be addressed. On the one hand, the 'market', i.e. those who design, develop, implement, set in place, research, provide, buy, etc. the learning analytics tools, will need to be convinced that using the EFLA to evaluate their tools is beneficial. On the other hand, the users, i.e. those who are being presented with the learning analytics tools and are using them, will need to be convinced to answer the EFLA questionnaire as without participation from the users, no evaluation will be possible. There are several related incentives that can be offered to both of these groups.

First, the EFLA provides insights. Using the framework to evaluate learning analytics tools can provide insight to the learners' and teachers' perception of and experience with the tool. It can reveal problematic aspects and identify ways to provide students with a more adaptive and less one-size-fits-all learning experience, as suggested by the field (Teasley, 2017). Once such issues are identified, they can be addressed in updated and improved versions of the learning analytics tool. The evaluated tool could be a whole dashboard as well as a single visualisation. The level of detail chosen is left to those who conduct the evaluation.

Second, the EFLA facilitates comparability. The framework can be used to compare learning analytics tools within one setting, e.g. two widgets for one course, or between different settings, e.g. widgets and dashboards from different courses or even from different educational institutions. Knowing how a tool performs according to the different EFLA dimensions can help to position it in the growing collection of tools available and can stimulate further development. If the results of EFLA scores were to be made publicly available, e.g. by commercial learning analytics providers

[11] https://rebrand.ly/EFLAtemplate
[12] https://rebrand.ly/EFLAscoring

or by those who publish their tools as open source, the comparability could be taken one step further. A tool's EFLA score could thus be used for advertisement purposes on both a commercial and an academic level.

Third, the EFLA supplies evidence. With the rising urge to ground learning analytics tools more in learning theories (Jivet et al., 2017), the framework can be used to ascertain whether a learning analytics tool fulfilled its intended purpose, i.e. whether it actually had an impact on learning and teaching processes and made them more efficient and more effective. This is what the community itself (e.g. Ferguson and Clow (2017)) and also policy makers have been calling for (e.g. ET2020 Working Group on Digital Skills and Competences (2016)). Additionally, by being able to show, i.e. by providing evidence, that a learning analytics tool really does fulfil its intended purpose, the rationale for collecting user data, which is seen as one of the biggest privacy-related issues, can be accounted for.

Whichever of these aspects poses as the main incentive for the 'market' to apply the EFLA for the evaluation of their learning analytics tools, they will have to be forwarded to the users to assure their participation. Only if the users see and understand the rationale behind having to answer the EFLA questionnaire (possibly again and again) and perceive the added value for themselves, will they be willing to provide their input. In the end, it needs to come down to both sides understanding that, as Baker (2016) describes, the goal of collecting and analysing data in the educational domain is not to create intelligent systems or stupid systems, but to create intelligent and successful students and to promote education.

Samenvatting

Nadat learning analytics in 2011 een separaat onderzoeksgebied werd, waren de verwachtingen hooggespannen. Verwacht werd dat hiermee het drop-out probleem ondervangen zou worden, dat de prestaties van studenten verhoogd konden worden en dat het de leer- en onderwijsprocessen in het algemeen zou kunnen ondersteunen. Echter, een aantal jaren later is empirisch bewijs over de vraag of learning analytics tools wel of niet het gewenste effect hebben erg gering. Er is nog steeds een kloof tussen het potentieel van learning analytics dat door onderzoek is vastgesteld en de mogelijkheden ervan die tot nu toe benut zijn (Ferguson et al., 2016a). Geïnspireerd door de System Usability Scale (SUS) (Brooke, 1996), beschrijven de onderzoeken in dit proefschrift een iteratief proces van creëren, gebruiken, evalueren en verbeteren van het evaluatieraamwerk voor learning analytics (EFLA).

Het ontwikkelde raamwerk voorziet in een lacune op het gebied van evaluatie-instrumenten en biedt een gestandaardiseerde manier aan om learning analytics tools te evalueren en de impact van learning analytics op de onderwijspraktijk te meten en te vergelijken. Net als de SUS maakt het EFLA gebruik van de subjectieve beoordelingen van gebruikers over learning analytics tools en geeft het een algemene indicatie van de door hen ervaren globale kwaliteit van een tool op een snelle, eenvoudige, doch grondig ontwikkelde, gevalideerde en betrouwbare manier. De belangrijkste doelstellingen van de onderzoeken die in dit proefschrift worden gepresenteerd, zijn drieledig: (1) identificeren van kwaliteitsindicatoren voor learning analytics, (2) ontwikkelen van een toepasbaar evaluatie-instrument op basis van deze indicatoren en (3) het valideren van het evaluatie-instrument.

De doelstellingen van dit proefschrift worden in drie verschillende delen behandeld. In deel I wordt, op basis van de input van de learning analytics gemeenschap, een group concept mapping studie uitgevoerd en worden de resultaten ervan gebruikt om de eerste versie van het raamwerk te ontwikkelen. Verdere input van learning analytics experts en een literatuurstudie worden daaropvolgend gebruikt om het EFLA te evalueren en aan te passen, om zo tot een tweede EFLA versie te komen. Deel II behandelt vervolgens de vragen of een learning analytics widget de resultaten van een leerling kan beïnvloeden, of het EFLA gebruikt kan worden om een widget op verschillende momenten te beoordelen en om verschillen in perceptie tussen de twee belangengroepen weer te geven. Zowel de resultaten van de widget evaluatie alsook de resultaten van een focusgroep bestaande uit deskundigen, worden vervolgens gebruikt om het EFLA te evalueren en de derde EFLA-versie te ontwikkelen. Tot slot wordt in Deel III van het proefschrift aangetoond, dat het EFLA ook kan worden gebruikt om veranderingen tussen verschillende versies van widgets te meten, evenals verschillen tussen twee belangengroepen. Verder worden de validatie- en betrouwbaarheidsanalyses beschreven, die leiden tot het ontwikkelen van de vierde en definitieve versie van het EFLA.

Tijdens de evaluatie- en verbetercycli van het EFLA worden verschillende cursuscontexten (formele cursus versus informele MOOC-cursus) en verschillende widgets (activiteitenwidget voor collaboratief leren versus activiteiten- en resourcewidget voor individueel leren) gebruikt, alsook verschillende studieprocedures (gebruik van het EFLA om dezelfde widget op verschillende tijdstippen te evalueren versus het gebruik van het EFLA om verschillende widgets op hetzelfde tijdstip te evalueren) en verschillende leeromgevingen (live studie versus lab studie). Nu de validiteit en betrouwbaarheid vastgesteld is, kan het raamwerk gebruikt worden om de kloof te dichten tussen de bereidheid binnen een instelling om learning analytics toe te passen en de volwassenheid van de reeds geïmplementeerde learning analytics toepassingen. Het raamwerk kan vervolgens de evaluatie op institutioneel niveau aanvullen door het meenemen van bijdragen van leerlingen en docenten, wat leidt tot een holistisch beeld van de leer- en onderwijsprocessen. Bovendien richt het zich ook op de vraag uit het vakgebied om bewijs te leveren van de impact van learning analytics.

De learning analytics community heeft nu de mogelijkheid om de toepasbaarheid en de voordelen van het EFLA zelf te verifiëren. Verwacht wordt, dat het volledige potentieel van het EFLA pas duidelijk wordt als het binnen het vakgebied op grotere schaal gebruikt gaat worden. Dit was ook het geval na de introductie van de System Usability Scale (SUS), die na gebruik door de onderzoeksgemeenschap als succesvol bestempeld kon worden (Brooke, 2013). Door templates van vragenlijsten[13] en het scoreblad[14] aan te bieden als download, wordt adoptie van het evaluatieraamwerk eenvoudig wat zal leiden tot een vergelijkbare gemakkelijke ‚Äòquick and dirty‚Äô evaluatie zoals bij de SUS.

Echter, om grote successen te boeken moeten er twee aspecten worden aangepakt. Enerzijds moet de 'markt', dat wil zeggen degenen die de learning analytics tools ontwerpen, ontwikkelen, implementeren, onderzoeken, leveren, kopen, etc. ervan overtuigd worden dat ze kunnen profiteren van het gebruik van het EFLA om hun tools te evalueren. Anderzijds zullen de gebruikers, dat wil zeggen degenen aan wie de learning analytics tools worden aangeboden en die ze gebruiken, ervan overtuigd moeten worden om de EFLA-vragenlijst in te vullen, omdat zonder deelname van de gebruikers geen integrale evaluatie mogelijk zal zijn. Er zijn verschillende stimulansen die aan beide groepen kunnen worden aangeboden.

Ten eerste biedt het EFLA inzichten. Als het raamwerk wordt gebruikt om learning analytics tools te evalueren, kan het inzicht verschaffen in de perceptie en ervaringen met de tool van lerenden en docenten. Het kan problematische aspecten onthullen en manieren identificeren om leerlingen een meer gepersonaliseerde en minder one-size-fits-all leerervaring te geven, zoals door het vakgebied voorgesteld wordt (Teasley,

[13] https://rebrand.ly/EFLAtemplate
[14] https://rebrand.ly/EFLAscoring

2017). Zodra dergelijke problematische aspecten zijn geïdentificeerd, kunnen ze worden aangepakt wat leidt tot verbeterde versies van de learning analytics tool. De geëvalueerde tool kan een volledig dashboard zijn, alsook een enkele visualisatie. Dit kan geheel bepaald worden door degenen die de evaluaties uitvoeren.

Ten tweede maakt het EFLA eenvoudig vergelijken mogelijk. Het raamwerk kan worden gebruikt om learning analytics tools te vergelijken binnen een setting, bijvoorbeeld twee widgets voor een cursus, of tussen verschillende settings, bijvoorbeeld widgets en dashboards uit verschillende cursussen of zelfs uit verschillende onderwijsinstellingen. Door kennis op te doen over hoe een tool volgens de verschillende EFLA-dimensies functioneert, kan het optimaal gepositioneerd worden in de groeiende verzameling tools die beschikbaar zijn, wat weer kan leiden tot verdere ontwikkeling van de tool. Als bijvoorbeeld de resultaten van EFLA-scores openbaar worden gemaakt door commerciële learning analytics providers of door hen die hun tools publiceren als open source, zal dit de vergelijkbaarheid een stap verder brengen. De EFLA-score van een tool kan dus zowel op commercieel als op academisch niveau worden gebruikt voor promotiedoeleinden van de betreffende tool.

Ten derde levert het EFLA bewijsmateriaal. Met de toenemende behoefte om learning analytics tools meer te verankeren in leertheorieën (Jivet et al., 2017), kan het raamwerk worden gebruikt om vast te stellen of een learning analytics tool het beoogde doel heeft bereikt. Met andere woorden, of het daadwerkelijk invloed heeft op leer- en onderwijsprocessen en deze efficiënter en effectiever maakt. Hier is vanuit de community (bijvoorbeeld Ferguson and Clow (2017)) en door beleidsmakers ook een oproep voor gedaan (bijvoorbeeld ET2020 Working Group on Digital Skills and Competences (2016)). Bovendien kan er door het leveren van bewijs dat een learning analytics tool werkelijk aan het beoogde doel voldoet, worden aangetoond dat het verzamelen van gebruikersgegevens, wat gezien wordt als een van de grootste privacy-gerelateerde problemen, geoorloofd is.

Ongeacht welke van deze aspecten als belangrijkste stimulans door de 'markt' gezien wordt om het EFLA toe te passen voor de evaluatie van learning analytics tools, de input van de daadwerkelijke gebruikers van de tools is essentieel als het gaat om het ontsluiten van het volledige potentieel van het EFLA. Het is dan ook van optimaal belang om de medewerking van de gebruikers te krijgen. Alleen als zij de argumenten zien en begrijpen om de EFLA-vragenlijst (mogelijk meerdere malen achter elkaar) te beantwoorden en zelf de toegevoegde waarde hiervan ervaren, zullen ze bereid zijn mee te werken. Uiteindelijk moeten beide kanten tot een besef komen dat, zoals Baker (2016) het beschrijft, het doel van het verzamelen en analyseren van data binnen het onderwijsdomein er niet is om intelligente systemen of domme systemen te ontwikkelen, maar om intelligente en succesvolle studenten te vormen en om het onderwijs te bevorderen.

Zusammenfassung

Seit Learning Analytics 2011 ein eigenes Forschungsfeld geworden ist, waren die Erwartungen groß, dass nun das Problem der Studienabbrüche gelöst, Studentenerfolge verbessert und Lern- und Lehrprozesse unterstützt werden können. Empirische Belege dafür, dass Learning Analytics Tools tatsächlich den gewünschten Effekt erzielen, sind bis heute jedoch rar. Es herrscht immer noch eine Diskrepanz zwischen dem, was die Forschungsgemeinschaft im Bereich von Learning Analytics für möglich hält, und was bisher tatsächlich erreicht wurde (Ferguson et al., 2016a). Von der System Usability Scale (SUS) (Brooke, 1996) inspiriert, beschreibt diese Doktorarbeit den kontinuierlichen und iterativen Prozess der Erstellung, Nutzung, Evaluierung und Verbesserung des Evaluation Framework for Learning Analytics (EFLA).

Das Framework adressiert den gegenwärtigen Mangel an Evaluierungsinstrumenten, indem es ein standardisiertes Verfahren anbietet, Learning Analytics Tools zu evaluieren und den Einfluss von Learning Analytics auf die Bildungspraxis mess- und vergleichbar zu machen. Ähnlich wie die SUS, basiert das EFLA auf den subjektiven Bewertungen von Learning Analytics Tools durch die Benutzer und bietet somit eine schnelle und schlichte, jedoch gründlich entwickelte, validierte und zuverlässige Art und Weise, eine allgemeine Angabe zur Gesamtqualität eines Tools zu erhalten. Die Hauptziele dieser Doktorarbeit sind dreigeteilt: (1) Identifizierung von Qualitätsindikatoren von Learning Analytics, (2) Erstellung eines anwendbaren Evaluierungsinstruments auf Basis dieser Indikatoren, und (3) Validierung des Evaluierungsinstruments.

Die Hauptziele der Doktorarbeit werden in drei Teilen behandelt. Basierend auf dem Beitrag der Learning Analytics Community, wird in Teil I eine Group Concept Mapping Studie durchgeführt und deren Resultate dazu benutzt, die erste Version des Frameworks zu erstellen. Anschließend werden Learning Analytics Experten sowie die Ergebnisse einer Literaturstudie zu Rate gezogen, um das EFLA zu evaluieren und zu überarbeiten und so die zweite Version des EFLA zu erstellen. Teil II setzt sich mit den Fragen auseinander, ob ein Learning Analytics Widget Einfluss auf die Noten von Lernenden haben kann und ob das EFLA für die Evaluierung solch eines Widgets und die Wiedergabe von Nutzerwahrnehmungen zu unterschiedlichen Zeitpunkten genutzt werden kann. Sowohl die Ergebnisse der Widgetevaluierung als auch Resultate einer Expertenfokusgruppe werden dann verwendet, um das EFLA zu evaluieren und die dritte Version zu erstellen. Teil III der Doktorarbeit befasst sich schließlich damit, Veränderungen zwischen unterschiedlichen Versionen eines Widgets mit EFLA zu messen. Des Weiteren werden die Validitäts- und Reliabilitätsanalysen präsentiert, deren Ergebnisse zur Erstellung der vierten und finalen EFLA-Version führen.

Während der Evaluierung und Verbesserung des EFLA werden sowohl verschiedene Kurskontexte (formeller, kollaborativer Kurs vs. informeller MOOC) als auch verschiedene Widgets (Aktivitätswidget für kollaboratives Lernen vs. Aktivitäts- und Ressourcenwidgets für individuelles Lernen) sowie verschiedene Studienabläufe (Evaluierung des gleichen Widgets zu unterschiedlichen Zeitpunkten vs. Evaluierung unterschiedlicher Widgets zum gleichen Zeitpunkt) und verschiedene Studienumgebungen (Live-Studie vs. Laborstudie) eingesetzt. Aufgrund der festgestellten Validität und Reliabilität kann das Framework nun dazu genutzt werden, die Lücke zwischen der Bereitschafts- und Reifemessung von Learning Analytics in einer Einrichtung zu schließen. Es ergänzt somit die bereits stattfindenden Evaluierungen auf Institutionsebene, indem die Lernenden und Lehrenden an der Evaluierung beteiligt werden. Dadurch ist ein holistischer Blick auf die involvierten Lern- und Lehrprozesse möglich. Außerdem wird so das Verlangen des Forschungsfeldes nach Wirksamkeitsnachweisen adressiert.

Die Learning Analytics Community hat nun die Gelegenheit, die Anwendbarkeit und den Nutzen des EFLA zu bestätigen. So wie die System Usability Scale (SUS) ihren vollen Erfolg erst dann richtig geltend machen konnte, nachdem sie von der Forschungsgemeinschaft aufgegriffen und genutzt wurde (Brooke, 2013), wird auch die Anwendung des EFLA in der Praxis und in größerem Rahmen nötig sein, um das Potenzial des Frameworks voll auszuschöpfen. Online abrufbare Fragebogenvorlagen[15] sowie ein interaktives Scoring-Sheet[16] vereinfachen die Annahme des Evaluierungsframeworks und ermöglichen eine ebenso einfache und 'quick and dirty' Evaluierung wie die SUS.

Um einen vollen Erfolg des Frameworks zu gewährleisten, müssen zwei Aspekte berücksichtigt werden. Einerseits muss der 'Markt' (d.h. diejenigen, die Learning Analytics Tools konzipieren, entwerfen, implementieren, einbauen, erforschen, zur Verfügung stellen, kaufen, etc.) davon überzeugt werden, dass sie von der Nutzung des EFLA zur Evaluierung ihrer Tools profitieren. Andererseits müssen die Nutzer (d.h. diejenigen, denen die Learning Analytics Tools angeboten werden und die diese nutzen) davon übezeugt werden, den EFLA-Fragebogen auszufüllen, da ohne ihre Mitwirkung keine Evaluierung möglich ist. Für beide Gruppen gibt es mehrere Nutzungsanreize.

Erstens verschafft das EFLA Einblicke. Wenn Learning Analytics Tools mit dem Framework evaluiert werden, können Einblicke in die Sichtweise der Lernenden und Lehrenden auf die Tools gewonnen und ihre Erfahrungen mit den Tools erfasst werden. Durch den Gebrauch des Frameworks können Probleme aufgedeckt und so Möglichkeiten identifiziert werden, Lernenden eine adaptivere und weniger einheitliche Lernerfahrung zu bieten, wie von der Forschungsgemeinschaft vorgeschlagen (Teasley, 2017). Ist ein Problem identifiziert, kann es in einer aktualisierten und verbesserten Version eines Learning Analytics Tools aufgegriffen werden. Ein evaluiertes Tool kann sowohl ein ganzes Dashboard, als auch eine einzelne Visuali-

[15] `https://rebrand.ly/EFLAtemplate`
[16] `https://rebrand.ly/EFLAscoring`

sierung sein. Der gewählte Detaillierungsgrad ist von denjenigen abhängig, die die Evaluierung durchführen.

Zweitens ermöglicht das EFLA Vergleichbarkeit. Das Framework kann benutzt werden, um Learning Analytics Tools innerhalb eines Settings (z.B. zwei Widgets innerhalb eines Kurses) oder zwischen zwei unterschiedlichen Settings (z.B. Widgets und Dashboards unterschiedlicher Kurse oder auch unterschiedlicher Institutionen) zu vergleichen. Wenn man weiß, wie ein Tool in den EFLA-Dimensionen abschneidet, kann dies bei seiner Positionierung in der stetig wachsenden Sammlung vorhandener Tools helfen und eine Weiterentwicklung fördern. Würden die EFLA-Scores öffentlich zugänglich gemacht (z.B. von kommerziellen Learning Analytics Anbietern oder denjenigen, die ihre Tools als Open Source veröffentlichen), könnte die Vergleichbarkeit noch einen Schritt weiter gehen. Der EFLA-Score eines Tools könnte dann auf kommerziellem sowie akademischem Niveau zu Werbezwecken genutzt werden.

Drittens liefert das EFLA Belege. Das Verlangen, Learning Analytics Tools stärker in und mit Lerntheorien zu begründen, wächst stetig (Jivet et al., 2017). Mit dem Framework kann evaluiert werden, ob ein Learning Analytics Tool seinen beabsichtigten Zweck erfüllt, d.h., ob es tatsächlich einen Einfluss auf die Lern- und Lehrprozesse hat und diese effizienter und effektiver macht. Dies herauszufinden, fordern sowohl die Community selbst (z.B. Ferguson and Clow (2017)) als auch die politischen Entscheidungsträger (z.B. ET2020 Working Group on Digital Skills and Competences (2016)). Hinzu kommt, dass mit dem Aufzeigen von Belegen für die Zweckerfüllung von Learning Analytics Tools, die Begründung für das Sammeln von Nutzerdaten, was als einer der größten Streitpunkte im Zusammenhang mit der Privatssphärenproblematik gesehen wird, verantwortet werden kann.

Welche dieser Punkte auch immer der Hauptanreiz für den 'Markt' sein mögen, das EFLA für die Evaluierung von Learning Analytics Tools zu verwenden, sie müssen an die Nutzer der Tools weitergeleitet werden, um deren Mitarbeit sicherzustellen. Nur wenn die Nutzer die Begründung für das (möglicherweise wiederholte) Ausfüllen des EFLA-Fragebogens nachvollziehen können und den Mehrwert für sich selbst erkennen, werden sie bereit sein, mitzumachen. Letztendlich kommt es darauf an, dass beide Seiten verstehen, dass, wie Baker (2016) es beschreibt, das Ziel des Datensammelns und -analysierens in der Bildungspraxis nicht die Schaffung intelligenter oder dummer Systeme ist, sondern die Schaffung intelligenter und erfolgreicher Lernender sowie die Förderung von Bildung.

Acknowledgements

Since I was a child, I have had a close relation to the Netherlands and to Heerlen and Maastricht especially, as we quite regularly drove there for a day, strolling through the streets, stocking up on school supplies and kaftpapier, Libelle's and Dutch goodies such as hagelslag and pindakaas. Of course, I had not heard about the Open Universiteit back then but that changed quickly when I started on my path into the academic world as a researcher at the Fraunhofer FIT in 2009. When working in the domain of technology enhanced learning, there is no way to not come across people of the OU and their TEL research during conferences, workshops and in journals. And so, what began as an external collaboration when I was looking for a promotor and a co-promotor for my PhD thesis, turned into a full time PhD position in 2014.

My first thanks therefore go to Marcus Specht and Hendrik Drachsler. Thank you both for taking me aboard the OU ship and for giving me the opportunity and putting the trust in me to do my work and research independently while still always being there for valuable discussions and providing constructive feedback. Additionally, thank you, Hendrik, for not only being a great co-promotor who continuously helped to uncover the red thread and who took notes of our discussions during car rides to meetings, but for also being a great colleague who involved me in internal and external research projects, the management of EU projects and other events within the learning analytics community from the beginning.

Many thanks also go to the co-authors of the papers included in this thesis, Slavi Stoyanov, Joop de Kraker, Karel Kreijns, Aad Slootmaker, Christian Toisoul and Stefaan Ternier, for great collaborations, valuable discussions and feedback. Further thanks go to my colleagues at the OU's Welten Institute in general and the TELI group in particular for wonderful conversations during lunch breaks, fun times in the game room, yummy food fests as well as inspirational and insightful discussions, new ideas and the sharing of experiences. A big thank you goes to Mieke Haemers for her reliability and organisational wisdom.

A very special set of thanks goes to my two paranimfen Katja Niemann and Ioana Jivet. Thank you for proofreading my manuscript and being a continuous source of support and feedback during the last phase of my PhD project. An additional heartfelt thank you goes to Katja, my former colleague and business trip companion. Thank you for being all ears since the beginning, never tiring of discussing my topic throughout the years and providing inspiration at all times as well as for being an aaaaawesome friend in Bonn, in Hamburg and anywhere in the world.

Profound thanks go to my parents Miek and Ralf. Thank you for helping me grow into the person I am today, for teaching me to look beyond my own nose, for fostering in me an open-mindedness and a curiosity to explore, for making me appreciate the emotional as well as the rational side of things, for passing on an interest in educational as well as technical aspects, for providing me with the confidence that I can achieve and accomplish my goals and for encouraging me throughout this PhD adventure of mine.

Last but in no way least, my deepest thanks and gratitude go to my husband Daniel. Thank you so much for accompanying and supporting me throughout this journey and letting me do my thing. Thank you for your patience and for always being there when I needed you and for leaving me be in my writing bubble when I needed that. Thank you for taking my mind off things at the right time in the right way and for making me laugh. Thank you for bringing so much music into my life and for nurturing my body, mind and soul with your delicious food. And above all, thank you for your love.

The research reported in Chapter 1, Chapter 2, Chapter 4 and Chapter 5 has been partially funded by the Seventh Framework Programme of the European Commission through the LACE project under grant 619424 (http://www.laceproject.eu). The research reported in Chapter 6 has been partially funded by the Competitiveness and Innovation Framework Programme of the European Commission through the ECO project under grant 621127 (http://project.ecolearning.eu).

SIKS Dissertation Series

The complete list of dissertations carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems from 1998 on can be found at `http://www.siks.nl/dissertations.php`.

2011

2011-01 Botond Cseke (RUN)
Variational Algorithms for Bayesian Inference in Latent Gaussian Models

2011-02 Nick Tinnemeier (UU)
Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language

2011-03 Jan Martijn van der Werf (TUE)
Compositional Design and Verification of Component-Based Information Systems

2011-04 Hado van Hasselt (UU)
Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference

2011-05 Bas van der Raadt (VU)
Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.

2011-06 Yiwen Wang (TUE)
Semantically-Enhanced Recommendations in Cultural Heritage

2011-07 Yujia Cao (UT)
Multimodal Information Presentation for High Load Human Computer Interaction

2011-08 Nieske Vergunst (UU)
BDI-based Generation of Robust Task-Oriented Dialogues

2011-09 Tim de Jong (OU)
Contextualised Mobile Media for Learning

2011-10 Bart Bogaert (UvT)
Cloud Content Contention

2011-11 Dhaval Vyas (UT)
Designing for Awareness: An Experience-focused HCI Perspective

2011-12 Carmen Bratosin (TUE)
Grid Architecture for Distributed Process Mining

2011-13 Xiaoyu Mao (UvT)
Airport under Control. Multiagent Scheduling for Airport Ground Handling

2011-14 Milan Lovric (EUR)
Behavioral Finance and Agent-Based Artificial Markets

2011-15 Marijn Koolen (UvA)
The Meaning of Structure: the Value of Link Evidence for Information Retrieval

2011-16 Maarten Schadd (UM)
Selective Search in Games of Different Complexity

2011-17 Jiyin He (UVA)
Exploring Topic Structure: Coherence, Diversity and Relatedness

2011-18 Mark Ponsen (UM)
Strategic Decision-Making in complex games

2011-19 Ellen Rusman (OU)
The Mind's Eye on Personal Profiles

2011-20 Qing Gu (VU)
Guiding service-oriented software engineering - A view-based approach

2011-21 Linda Terlouw (TUD)
Modularization and Specification of Service-Oriented Systems

2011-22 Junte Zhang (UVA)
System Evaluation of Archival Description and Access

2011-23 Wouter Weerkamp (UVA)
Finding People and their Utterances in Social Media

2011-24 Herwin van Welbergen (UT)
Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior

2011-25 Syed Waqar ul Qounain Jaffry (VU)
Analysis and Validation of Models for Trust Dynamics

2011-26 Matthijs Aart Pontier (VU)
Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots

2011-27 Aniel Bhulai (VU)
Dynamic website optimization through autonomous management of design patterns

2011-28 Rianne Kaptein (UVA)
Effective Focused Retrieval by Exploiting Query Context and Document Structure

2011-29 Faisal Kamiran (TUE)
Discrimination-aware Classification

2011-30 Egon van den Broek (UT)
Affective Signal Processing (ASP): Unraveling the mystery of emotions

2011-31 Ludo Waltman (EUR)
Computational and Game-Theoretic Approaches for Modeling Bounded Rationality

2011-32 Nees-Jan van Eck (EUR)
Methodological Advances in Bibliometric Mapping of Science

2011-33 Tom van der Weide (UU)
Arguing to Motivate Decisions

2011-34 Paolo Turrini (UU)
Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations

2011-35 Maaike Harbers (UU)
Explaining Agent Behavior in Virtual Training

2011-36 Erik van der Spek (UU)
Experiments in serious game design: a cognitive approach

2011-37 Adriana Burlutiu (RUN)
Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference

2011-38 Nyree Lemmens (UM)
Bee-inspired Distributed Optimization

2011-39 Joost Westra (UU)
Organizing Adaptation using Agents in Serious Games

2011-40 Viktor Clerc (VU)
Architectural Knowledge Management in Global Software Development

2011-41 Luan Ibraimi (UT)
Cryptographically Enforced Distributed Data Access Control

2011-42 Michal Sindlar (UU)
Explaining Behavior through Mental State Attribution

2011-43 Henk van der Schuur (UU)
Process Improvement through Software Operation Knowledge

2011-44 Boris Reuderink (UT)
Robust Brain-Computer Interfaces

2011-45 Herman Stehouwer (UvT)
Statistical Language Models for Alternative Sequence Selection

2011-46 Beibei Hu (TUD)
Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work

2011-47 Azizi Bin Ab Aziz (VU)
Exploring Computational Models for Intelligent Support of Persons with Depression

2011-48 Mark Ter Maat (UT)
Response Selection and Turn-taking for a Sensitive Artificial Listening Agent

2011-49 Andreea Niculescu (UT)
Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

2012

2012-01 Terry Kakeeto (UvT)
Relationship Marketing for SMEs in Uganda

2012-02 Muhammad Umair (VU)
Adaptivity, emotion, and Rationality in Human and Ambient Agent Models

2012-03 Adam Vanya (VU)
Supporting Architecture Evolution by Mining Software Repositories

2012-04 Jurriaan Souer (UU)
Development of Content Management System-based Web Applications

2012-05 Marijn Plomp (UU)
Maturing Interorganisational Information Systems

2012-06 Wolfgang Reinhardt (OU)
Awareness Support for Knowledge Workers in Research Networks

2012-07 Rianne van Lambalgen (VU)
When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions

2012-08 Gerben de Vries (UVA)
Kernel Methods for Vessel Trajectories

2012-09 Ricardo Neisse (UT)
Trust and Privacy Management Support for Context-Aware Service Platforms

2012-10 David Smits (TUE)
Towards a Generic Distributed Adaptive Hypermedia Environment

2012-11 J.C.B. Rantham Prabhakara (TUE)
Process Mining in the Large: Preprocessing, Discovery, and Diagnostics

2012-12 Kees van der Sluijs (TUE)
Model Driven Design and Data Integration in Semantic Web Information Systems

2012-13 Suleman Shahid (UvT)
Fun and Face: Exploring non-verbal expressions of emotion during playful interactions

2012-14 Evgeny Knutov (TUE)
Generic Adaptation Framework for Unifying Adaptive Web-based Systems

2012-15 Natalie van der Wal (VU)
Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.

2012-16 Fiemke Both (VU)
Helping people by understanding them - Ambient Agents supporting task execution and depression treatment

2012-17 Amal Elgammal (UvT)
Towards a Comprehensive Framework for Business Process Compliance

2012-18 Eltjo Poort (VU)
Improving Solution Architecting Practices

2012-19 Helen Schonenberg (TUE)
What's Next? Operational Support for Business Process Execution

2012-20 Ali Bahramisharif (RUN)
Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing

2012-21 Roberto Cornacchia (TUD)
Querying Sparse Matrices for Information Retrieval

2012-22 Thijs Vis (UvT)
Intelligence, politie en veiligheidsdienst: verenigbare grootheden?

2012-23 Christian Muehl (UT)
Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction

2012-24 Laurens van der Werff (UT)
Evaluation of Noisy Transcripts for Spoken Document Retrieval

2012-25 Silja Eckartz (UT)
Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application

2012-26 Emile de Maat (UVA)
Making Sense of Legal Text

2012-27 Hayrettin Gurkok (UT)
Mind the Sheep! User Experience Evaluation Brain-Computer Interface Games

2012-28 Nancy Pascall (UvT)
Engendering Technology Empowering Women

2012-29 Almer Tigelaar (UT)
Peer-to-Peer Information Retrieval

2012-30 Alina Pommeranz (TUD)
Designing Human-Centered Systems for Reflective Decision Making

2012-31 Emily Bagarukayo (RUN)
A Learning by Construction Approach for Higher Order Cognitive Skills Improvement
Building Capacity and Infrastructure

2012-32 Wietske Visser (TUD)
Qualitative multi-criteria preference representation and reasoning

2012-33 Rory Sie (OUN)
Coalitions in Cooperation Networks (COCOON)

2012-34 Pavol Jancura (RUN)
Evolutionary analysis in PPI networks and applications

2012-35 Evert Haasdijk (VU)
Never Too Old To Learn
On-line Evolution of Controllers in Swarm- and Modular Robotics

2012-36 Denis Ssebugwawo (RUN)
Analysis and Evaluation of Collaborative Modeling Processes

2012-37 Agnes Nakakawa (RUN)
A Collaboration Process for Enterprise Architecture Creation

2012-38 Selmar Smit (VU)
Parameter Tuning and Scientific Testing in Evolutionary Algorithms

2012-39 Hassan Fatemi (UT)
Risk-aware design of value and coordination networks

2012-40 Agus Gunawan (UvT)
Information Access for SMEs in Indonesia

2012-41 Sebastian Kelle (OU)
Game Design Patterns for Learning

2012-42 Dominique Verpoorten (OU)
Reflection Amplifiers in self-regulated Learning

2012-43 Withdrawn

2012-44 Anna Tordai (VU)
On Combining Alignment Techniques

2012-45 Benedikt Kratz (UvT)
A Model and Language for Business-aware Transactions

2012-46 Simon Carter (UVA)
Exploration and Exploitation of Multilingual Data for Statistical Machine Translation

2012-47 Manos Tsagkias (UVA)
Mining Social Media: Tracking Content and Predicting Behavior

2012-48 Jorn Bakker (TUE)
Handling Abrupt Changes in Evolving Time-series Data

2012-49 Michael Kaisers (UM)
Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions

2012-50 Steven van Kervel (TUD)
Ontologogy driven Enterprise Information Systems Engineering

2012-51 Jeroen de Jong (TUD)
Heuristics in Dynamic Sceduling; a practical framework with a case study in elevator dispatching

2013

2013-01 Viorel Milea (EUR)
News Analytics for Financial Decision Support

2013-02 Erietta Liarou (CWI)
MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing

2013-03 Szymon Klarman (VU)
Reasoning with Contexts in Description Logics

2013-04 Chetan Yadati (TUD)
Coordinating autonomous planning and scheduling

2013-05 Dulce Pumareja (UT)
Groupware Requirements Evolutions Patterns

2013-06 Romulo Goncalves (CWI)
The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience

2013-07 Giel van Lankveld (UvT)
Quantifying Individual Player Differences

2013-08 Robbert-Jan Merk (VU)
Making enemies: cognitive modeling for opponent agents in fighter pilot simulators

2013-09 Fabio Gori (RUN)
Metagenomic Data Analysis: Computational Methods and Applications

2013-10 Jeewanie Jayasinghe Arachchige (UvT)
A Unified Modeling Framework for Service Design.

2013-11 Evangelos Pournaras (TUD)
Multi-level Reconfigurable Self-organization in Overlay Services

2013-12 Marian Razavian (VU)
Knowledge-driven Migration to Services

2013-13 Mohammad Safiri (UT)
Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly

2013-14 Jafar Tanha (UVA)
Ensemble Approaches to Semi-Supervised Learning Learning

2013-15 Daniel Hennes (UM)
Multiagent Learning - Dynamic Games and Applications

2013-16 Eric Kok (UU)
Exploring the practical benefits of argumentation in multi-agent deliberation

2013-17 Koen Kok (VU)
The PowerMatcher: Smart Coordination for the Smart Electricity Grid

2013-18 Jeroen Janssens (UvT)
Outlier Selection and One-Class Classification

2013-19 Renze Steenhuizen (TUD)
Coordinated Multi-Agent Planning and Scheduling

2013-20 Katja Hofmann (UvA)
Fast and Reliable Online Learning to Rank for Information Retrieval

2013-21 Sander Wubben (UvT)
Text-to-text generation by monolingual machine translation

2013-22 Tom Claassen (RUN)
Causal Discovery and Logic

2013-23 Patricio de Alencar Silva (UvT)
Value Activity Monitoring

2013-24 Haitham Bou Ammar (UM)
Automated Transfer in Reinforcement Learning

2013-25 Agnieszka Anna Latoszek-Berendsen (UM)
Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System

2013-26 Alireza Zarghami (UT)
Architectural Support for Dynamic Homecare Service Provisioning

2013-27 Mohammad Huq (UT)
Inference-based Framework Managing Data Provenance

2013-28 Frans van der Sluis (UT)
When Complexity becomes Interesting: An Inquiry into the Information eXperience

2013-29 Iwan de Kok (UT)
Listening Heads

2013-30 Joyce Nakatumba (TUE)
Resource-Aware Business Process Management: Analysis and Support

2013-31 Dinh Khoa Nguyen (UvT)
Blueprint Model and Language for Engineering Cloud Applications

2013-32 Kamakshi Rajagopal (OUN)
Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development

2013-33 Qi Gao (TUD)
User Modeling and Personalization in the Microblogging Sphere

2013-34 Tjin-Kam-Jet (UT)
Distributed Deep Web Search

2013-35 Abdallah El Ali (UvA)
Minimal Mobile Human Computer Interaction

2013-36 Than Lam Hoang (TUe)
Pattern Mining in Data Streams

2013-37 Dirk Börner (OUN)
Ambient Learning Displays

2013-38 Eelco den Heijer (VU)
Autonomous Evolutionary Art

2013-39 Joop de Jong (TUD)
A Method for Enterprise Ontology based Design of Enterprise Information Systems

2013-40 Pim Nijssen (UM)
Monte-Carlo Tree Search for Multi-Player Games

2013-41 Jochem Liem (UVA)
Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning

2013-42 Léon Planken (TUD)
Algorithms for Simple Temporal Reasoning

2013-43 Marc Bron (UVA)
Exploration and Contextualization through Interaction and Concepts

2014

2014-01 Nicola Barile (UU)
Studies in Learning Monotone Models from Data

2014-02 Fiona Tuliyano (RUN)
Combining System Dynamics with a Domain Modeling Method

2014-03 Sergio Raul Duarte Torres (UT)
Information Retrieval for Children: Search Behavior and Solutions

2014-04 Hanna Jochmann-Mannak (UT)
Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation

2014-05 Jurriaan van Reijsen (UU)
Knowledge Perspectives on Advancing Dynamic Capability

2014-06 Damian Tamburri (VU)
Supporting Networked Software Development

2014-07 Arya Adriansyah (TUE)
Aligning Observed and Modeled Behavior

2014-08 Samur Araujo (TUD)
Data Integration over Distributed and Heterogeneous Data Endpoints

2014-09 Philip Jackson (UvT)
Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language

2014-10 Ivan Salvador Razo Zapata (VU)
Service Value Networks

2014-11 Janneke van der Zwaan (TUD)
An Empathic Virtual Buddy for Social Support

2014-12 Willem van Willigen (VU)
Look Ma, No Hands: Aspects of Autonomous Vehicle Control

2014-13 Arlette van Wissen (VU)
Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains

2014-14 Yangyang Shi (TUD)
Language Models With Meta-information

2014-15 Natalya Mogles (VU)
Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare

2014-16 Krystyna Milian (VU)
Supporting trial recruitment and design by automatically interpreting eligibility criteria

2014-17 Kathrin Dentler (VU)
Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability

2014-18 Mattijs Ghijsen (UVA)
Methods and Models for the Design and Study of Dynamic Agent Organizations

2014-19 Vinicius Ramos (TUE)
Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support

2014-20 Mena Habib (UT)
Named Entity Extraction and Disambiguation for Informal Text: The Missing Link

2014-21 Kassidy Clark (TUD)
Negotiation and Monitoring in Open Environments

2014-22 Marieke Peeters (UU)
Personalized Educational Games - Developing agent-supported scenario-based training

2014-23 Eleftherios Sidirourgos (UvA/CWI)
Space Efficient Indexes for the Big Data Era

2014-24 Davide Ceolin (VU)
Trusting Semi-structured Web Data

2014-25 Martijn Lappenschaar (RUN)
New network models for the analysis of disease interaction

2014-26 Tim Baarslag (TUD)
What to Bid and When to Stop

2014-27 Rui Jorge Almeida (EUR)
Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty

2014-28 Anna Chmielowiec (VU)
Decentralized k-Clique Matching

2014-29 Jaap Kabbedijk (UU)
Variability in Multi-Tenant Enterprise Software

2014-30 Peter de Cock (UvT)
Anticipating Criminal Behaviour

2014-31 Leo van Moergestel (UU)
Agent Technology in Agile Multiparallel Manufacturing and Product Support

2014-32 Naser Ayat (UvA)
On Entity Resolution in Probabilistic Data

2014-33 Tesfa Tegegne (RUN)
Service Discovery in eHealth

2014-34 Christina Manteli (VU)
The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems.

2014-35 Joost van Ooijen (UU)
Cognitive Agents in Virtual Worlds: A Middleware Design Approach

2014-36 Joos Buijs (TUE)
Flexible Evolutionary Algorithms for Mining Structured Process Models

2014-37 Maral Dadvar (UT)
Experts and Machines United Against Cyberbullying

2014-38 Danny Plass-Oude Bos (UT)
Making brain-computer interfaces better: improving usability through post-processing.

2014-39 Jasmina Maric (UvT)
Web Communities, Immigration, and Social Capital

2014-40 Walter Omona (RUN)
A Framework for Knowledge Management Using ICT in Higher Education

2014-41 Frederic Hogenboom (EUR)
Automated Detection of Financial Events in News Text

2014-42 Carsten Eijckhof (CWI/TUD)
Contextual Multidimensional Relevance Models

2014-43 Kevin Vlaanderen (UU)
Supporting Process Improvement using Method Increments

2014-44 Paulien Meesters (UvT)
Intelligent Blauw. Met als ondertitel: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden.

2014-45 Birgit Schmitz (OUN)
Mobile Games for Learning: A Pattern-Based Approach

2014-46 Ke Tao (TUD)
Social Web Data Analytics: Relevance, Redundancy, Diversity

2014-47 Shangsong Liang (UVA)
Fusion and Diversification in Information Retrieval

2015

2015-01 Niels Netten (UvA)
Machine Learning for Relevance of Information in Crisis Response

2015-02 Faiza Bukhsh (UvT)
Smart auditing: Innovative Compliance Checking in Customs Controls

2015-03 Twan van Laarhoven (RUN)
Machine learning for network data

2015-04 Howard Spoelstra (OUN)
Collaborations in Open Learning Environments

2015-05 Christoph Bösch (UT)
Cryptographically Enforced Search Pattern Hiding

2015-06 Farideh Heidari (TUD)
Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes

2015-07 Maria-Hendrike Peetz (UvA)
Time-Aware Online Reputation Analysis

2015-08 Jie Jiang (TUD)
Organizational Compliance: An agent-based model for designing and evaluating organizational interactions

2015-09 Randy Klaassen (UT)
HCI Perspectives on Behavior Change Support Systems

2015-10 Henry Hermans (OUN)
OpenU: design of an integrated system to support lifelong learning

2015-11 Luo (TUE)
Designing algorithms for big graph datasets: A study of computing bisimulation and joins

2015-12 Julie M. Birkholz (VU)
Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks

2015-13 Giuseppe Procaccianti (VU)
Energy-Efficient Software

2015-14 Bart van Straalen (UT)
A cognitive approach to modeling bad news conversations

2015-15 Klaas Andries de Graaf (VU)
Ontology-based Software Architecture Documentation

2015-16 Changyun Wei (UT)
Cognitive Coordination for Cooperative Multi-Robot Teamwork

2015-17 André van Cleeff (UT)
Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs

2015-18 Holger Pirk (CWI)
Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories

2015-19 Bernardo Tabuenca (OUN)
Ubiquitous Technology for Lifelong Learners

2015-20 Lois Vanhée (UU)
Using Culture and Values to Support Flexible Coordination

2015-21 Sibren Fetter (OUN)
Using Peer-Support to Expand and Stabilize Online Learning

2015-22 Zhemin Zhu (UT)
Co-occurrence Rate Networks

2015-23 Luit Gazendam (VU)
Cataloguer Support in Cultural Heritage

2015-24 Richard Berendsen (UVA)
Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation

2015-25 Steven Woudenberg (UU)
Bayesian Tools for Early Disease Detection

2015-26 Alexander Hogenboom (EUR)
Sentiment Analysis of Text Guided by Semantics and Structure

2015-27 Sándor Héman (CWI)
Updating compressed colomn stores

2015-28 Janet Bagorogoza (TiU)
Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO

2015-29 Hendrik Baier (UM)
Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains

2015-30 Kiavash Bahreini (OU)
Real-time Multimodal Emotion Recognition in E-Learning

2015-31 Yakup Koç (TUD)
On the robustness of Power Grids

2015-32 Jerome Gard (UL)
Corporate Venture Management in SMEs

2015-33 Frederik Schadd (TUD)
Ontology Mapping with Auxiliary Resources

2015-34 Victor de Graaf (UT)
Gesocial Recommender Systems

2015-35 Jungxao Xu (TUD)
Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction

2016

2016-01 Syed Saiden Abbas (RUN)
Recognition of Shapes by Humans and Machines

2016-02 Michiel Christiaan Meulendijk (UU)
Optimizing medication reviews through decision support: prescribing a better pill to swallow

2016-03 Maya Sappelli (RUN)
Knowledge Work in Context: User Centered Knowledge Worker Support

2016-04 Laurens Rietveld (VU)
Publishing and Consuming Linked Data

2016-05 Evgeny Sherkhonov (UVA)
Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers

2016-06 Michel Wilson (TUD)
Robust scheduling in an uncertain environment

2016-07 Jeroen de Man (VU)
Measuring and modeling negative emotions for virtual training

2016-08 Matje van de Camp (TiU)
A Link to the Past: Constructing Historical Social Networks from Unstructured Data

2016-09 Archana Nottamkandath (VU)
Trusting Crowdsourced Information on Cultural Artefacts

2016-10 George Karafotias (VUA)
Parameter Control for Evolutionary Algorithms

2016-11 Anne Schuth (UVA)
Search Engines that Learn from Their Users

2016-12 Max Knobbout (UU)
Logics for Modelling and Verifying Normative Multi-Agent Systems

2016-13 Nana Baah Gyan (VU)
The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach

2016-14 Ravi Khadka (UU)
Revisiting Legacy Software System Modernization

2016-15 Steffen Michels (RUN)
Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments

2016-16 Guangliang Li (UVA)
Socially Intelligent Autonomous Agents that Learn from Human Reward

2016-17 Berend Weel (VU)
Towards Embodied Evolution of Robot Organisms

2016-18 Albert Meroño Peñuela (VU)
Refining Statistical Data on the Web

2016-19 Julia Efremova (Tu/e)
Mining Social Structures from Genealogical Data

2016-20 Daan Odijk (UVA)
Context Semantics in News Web Search

2016-21 Alejandro Moreno Célleri (UT)
From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground

2016-22 Grace Lewis (VU)
Software Architecture Strategies for Cyber-Foraging Systems

2016-23 Fei Cai (UVA)
Query Auto Completion in Information Retrieval

2016-24 Brend Wanders (UT)
Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach

2016-25 Julia Kiseleva (TU/e)
Using Contextual Information to Understand Searching and Browsing Behavior

2016-26 Dilhan Thilakarathne (VU)
In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains

2016-27 Wen Li (TUD)
Understanding Geo-spatial Information on Social Media

2016-28 Mingxin Zhang (TUD)
Large-scale Agent-based Social Simulation - A study on epidemic prediction and control

2016-29 Nicolas Höning (TUD)
Peak reduction in decentralised electricity systems - Markets and prices for flexible planning

2016-30 Ruud Mattheij (UvT)
The Eyes Have It

2016-31 Mohammad Khelghati (UT)
Deep web content monitoring

2016-32 Eelco Vriezekolk (UT)
Assessing Telecommunication Service Availability Risks for Crisis Organisations

2016-33 Peter Bloem (UVA)
Single Sample Statistics
exercises in learning from just one example

2016-34 Dennis Schunselaar (TUE)
Configurable Process Trees: Elicitation, Analysis, and Enactment

2016-35 Zhaochun Ren (UVA)
Monitoring Social Media: Summarization, Classification and Recommendation

2016-36 Daphne Karreman (UT)
Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies

2016-37 Giovanni Sileno (UvA)
Aligning Law and Action - a conceptual and computational inquiry

2016-38 Andrea Minuto (UT)
Materials that Matter - Smart Materials meet Art Interaction Design

2016-39 Merijn Bruijnes (UT)
Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect

2016-40 Christian Detweiler (TUD)
Accounting for Values in Design

2016-41 Thomas King (TUD)
Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance

2016-42 Spyros Martzoukos (UVA)
Combinatorial and Compositional Aspects of Bilingual Aligned Corpora

2016-43 Saskia Koldijk (RUN)
Context-Aware Support for Stress Self-Management: From Theory to Practice

2016-44 Thibault Sellam (UVA)
Automatic Assistants for Database Exploration

2016-45 Bram van de Laar (UT)
Experiencing Brain-Computer Interface Control

2016-46 Jorge Gallego Perez (UT)
Robots to Make you Happy

2016-47 Christina Weber (UL)
Real-time foresight - Preparedness for dynamic innovation networks

2016-48 Tanja Buttler (TUD)
Collecting Lessons Learned

2016-49 Gleb Polevoy (TUD)
Participation and Interaction in Projects. A Game-Theoretic Analysis

2016-50 Yan Wang (UVT)
The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

2017

2017-01 Jan-Jaap Oerlemans (UL)
Investigating Cybercrime

2017-02 Sjoerd Timmer (UU)
Designing and Understanding Forensic Bayesian Networks using Argumentation

2017-03 Daniël Harold Telgen (UU)
Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines

2017-04 Mrunal Gawade (CWI)
Multi-core Parallelism in a Column-store

2017-05 Mahdieh Shadi (UVA)
Collaboration Behavior

2017-06 Damir Vandic (EUR)
Intelligent Information Systems for Web Product Search

2017-07 Roel Bertens (UU)
Insight in Information: from Abstract to Anomaly

2017-08 Rob Konijn (VU)
Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery

2017-09 Dong Nguyen (UT)
Text as Social and Cultural Data: A Computational Perspective on Variation in Text

2017-10 Robby van Delden (UT)
(Steering) Interactive Play Behavior

2017-11 Florian Kunneman (RUN)
Modelling patterns of time and emotion in Twitter #anticipointment

2017-12 Sander Leemans (TUE)
Robust Process Mining with Guarantees

2017-13 Gijs Huisman (UT)
Social Touch Technology - Extending the reach of social touch through haptic technology

2017-14 Shoshannah Tekofsky (UvT)
You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior

2017-15 Peter Berck (RUN)
Memory-Based Text Correction

2017-16 Aleksandr Chuklin (UVA)
Understanding and Modeling Users of Modern Search Engines

2017-17 Daniel Dimov (UL)
Crowdsourced Online Dispute Resolution

2017-18 Ridho Reinanda (UVA)
Entity Associations for Search

2017-19 Jeroen Vuurens (UT)
Proximity of Terms, Texts and Semantic Vectors in Information Retrieval

2017-20 Mohammadbashir Sedighi (TUD)
Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility

2017-21 Jeroen Linssen (UT)
Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)

2017-22 Sara Magliacane (VU)
Logics for causal inference under uncertainty

2017-23 David Graus (UVA)
Entities of Interest — Discovery in Digital Traces

2017-24 Chang Wang (TUD)
Use of Affordances for Efficient Robot Learning

2017-25 Veruska Zamborlini (VU)
Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search

2017-26 Merel Jung (UT)
Socially intelligent robots that understand and respond to human touch

2017-27 Michiel Joosse (UT)
Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors

2017-28 John Klein (VU)
Architecture Practices for Complex Contexts

2017-29 Adel Alhuraibi (UvT)
From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT

2017-30 Wilma Latuny (UvT)
The Power of Facial Expressions

2017-31 Ben Ruijl (Uvt)
Advances in computational methods for QFT calculations

2017-32 Thaer Samar (RUN)
Access to and Retrievability of Content in Web Archives

2017-33 Brigit van Loggem (OU)
Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity