UK Academy for Information Systems Conference Proceedings 2019

UK Academy for Information Systems

Spring 4-10-2019

Using a Quantified-Self App to Personalise Learning -- A Comparison of Visualisations for the Evaluation of the Learning Process

Josepha Witt
University of Hohenheim, josepha.witt@uni-hohenheim.de

Philipp Melzer
University of Hohenheim, Philipp.Melzer@uni-hohenheim.de

Mareike Schoop
University of Hohenheim, m.schoop@uni-hohenheim.de

Using a Quantified-Self App to Personalise Learning – A Comparison of Visualisations for the Evaluation of the Learning Process

Josepha Witt, Philipp Melzer, Mareike Schoop
Information Systems Group
University of Hohenheim,
70599 Stuttgart, Germany
{josepha.witt, philipp.melzer, schoop}@uni-hohenheim.de

Abstract

The digitalisation of students' lives leads to the almost ubiquitous use of apps for all parts of life. The digitalisation of university learning has led to many learning management systems in use in institutions of higher education. However, it has not quite kept up with the demand for highly flexible learning at all hours and in all locations. Learning apps are not used frequently by universities to improve students' personalised learning. The paper reports on an app that combines self-regulated learning and the Quantified-Self approach to support such ubiquitous learning. When students track their learning in an app, they can later on benefit from the tracked data on an individual as well as aggregated level. Data analyses provide the potential for individual evaluation of the learning or comparison to peers. Thus, this study derives an extensive set of user stories for such app from the literature. Those user stories are the basis for evaluating the approach by turning them into visualisations that are then tested based on a mixed-method approach. The evaluation finds differences among the evaluated visualisations regarding ease of understanding, intuitive operations, visual appeal, and metacognition as well as potential for further improvement. From the findings an improved set of visualisations is generated and the results are fed back into the user stories.

Keywords: Personalised learning, self-regulated learning, quantified-self, visualisation, app design, mixed-methods

1.0 Introduction

Personalisation of learning regarding pace, content, and methods of the individual learners is usually performed by the teacher. However, with increasing class sizes and increasing degree of online learning especially in higher education, such personalisation becomes difficult. To encompass heterogeneity of the learners, self-regulated personalised learning provides a way to shift responsibility to the learners and enables them to personalise their learning themselves by selecting and using different learning tasks and tools in order to achieve the learning goals (Melzer, 2018). Self-regulated learning (SRL) provides the basis for such personalisation as it describes learning as a cyclic process encompassing (1) a preparatory phase, (2) a performance phase, and (3) an appraisal phase (Panadero, 2017). In addition to the core learning activity, the learners focus on observing and documenting their behaviour to provide a basis for reflecting and improving the learning. SRL can be performed using pen and paper for documentation; however, recent approaches combine self-regulated learning and Quantified-Self (Melzer, 2019; Neitzel & Rensing, 2017; Piotrkowicz et al., 2017). The Quantified-Self (QS) describes a movement of tracking personal data (e.g. food consumption, sports) through wearables or mobile phones. The data gathered is used to deduce self-knowledge and improve individual behaviour. In the domain of self-regulated learning, such apps provide numerous benefits in terms of ubiquitous access to the data, powerful analyses, and individual recommendations. These benefits are of vital importance for app users, as they need to compensate the often cumbersome process of manual data tracking to increase reuse intention of the apps eventually. The app "Quantified-Self in E-Learning" (QSEL) (Melzer, Schneider & Schoop, 2018) as a good example for the above approach tracks learning sessions of students w.r.t. duration, place, lecture, topic, task, tool, and satisfaction.

The research goal of the current paper is to analyse how such app should visualise the data in order to benefit its users and increase reuse intention. Focusing primarily on the evaluation of the data, the following research questions will be investigated in the paper:

1. What are relevant user stories to transmit the benefits of such an app?
2. What are possible visualisations of these user stories?
3. What are evaluation criteria for those visualisations?

The research follows a design science approach (Baskerville & Pries-Heje, 2010) deriving user stories based on the literature. These user stories are then implemented as mock-ups showing potential visualisations of the gathered learning data. Finally, those visualisations are subject to exploratory evaluation adhering to a mixed-method approach. The contribution of this paper, therefore, can be seen in generalisable user stories (meta-requirements) to communicate the benefits of a QS app to the users as well as specific implementations of these user stories which are evaluated and in turn fed back into the user stories.

2.0 Background and Approach

Considering SRL in the context of higher education, the cyclic learning process can be applied with different levels of granularity. In the following, we will consider three levels of the SRL process (cf. Figure 1): the academic learning level (which starts with selecting a degree programme and ends when achieving the degree); the term-based learning level (which starts with choosing the courses for the particular term and ends with completing the exam for each of those courses); the session-based level (which is a fine-grained process covering one learning session and ends with a reflection of the learning outcome).

The three phases of the cyclic SRL process are composed of different categories (Zimmerman & Moylan, 2009) (cf. Figure 2), which relate to QS as follows. The preparatory phase includes the analysis of the task and self-motivational beliefs. In the performance phase, the learner records his/her behaviour during self-observation and adjusts it based on tasks referred to as self-control. Afterwards, self-judgment and self-reacting include the comparison of the current performance with previous ones and the satisfaction with the performance. This has an impact on further iterations of the process. Based on the SRL process, user stories, which occur using a QS app in the context of learning, can now be developed (cf. Table 1) and be assigned to the learning levels (cf. Figure 1).

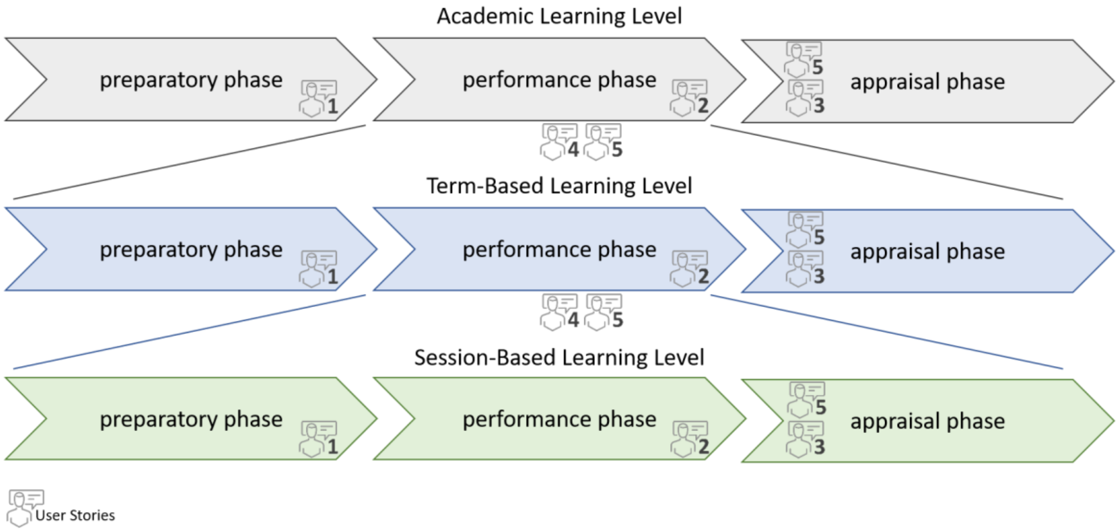

Figure 1. **SRL process of different levels of granularity during studies**

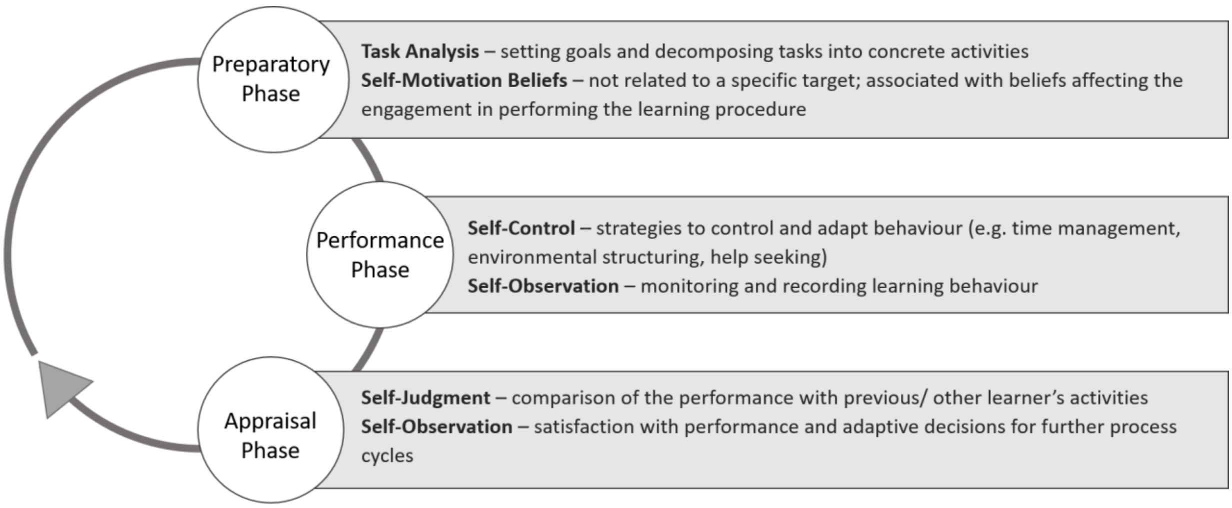

Figure 2. **Phases of the SRL process adapted from Zimmerman and Moylan (2009) and Panadero (2017)**

As one-time activities to use the app (registration, settings, etc.) are not related to that process, they are not considered. Looking at the composition of the first phase (preparatory phase), task analysis includes setting goals and decomposing tasks into concrete activities (cf. Figure 2), which can be supported using the app (cf. user story 1, Table 1). In contrast, self-motivation beliefs are not related to a specific target; rather they are associated with beliefs or feelings, affecting the engagement in the performance phase (cf. Zimmerman & Moylan, 2009). Therefore, no user story captures self-motivation beliefs.

Self-observation (which is part of the performance phase) covers the task of self-recording, i.e. tracking the behaviour using a QS app; this is captured as user story 2 (cf. Table 1). Besides gathering information regarding the preparatory phase, this user story takes place following the performance phase of a process (cf. Figure 1). In the session-based learning level, session data (e.g. learning duration) is recorded; the term-based learning level captures exam grades; the academic learning level documents degrees, certificates, or badges.

User Story 1: User can track their learning objectives and related tasks to comply with and reflect on the plan prepared in the preparatory phase.
User Story 2: User can track their behaviour to self-observe their performance during learning.
User Story 3: User can evaluate their performance based on data from recent learning lesson to reflect and perform better next time.
User Story 4: User can evaluate their performance based on aggregated data to monitor the achievement of objectives and self-control their learning performance.
User Story 5: User can compare their performance to previous individual performances or peer performances to affect further preparation and appraisal phases.

Table 1. User Stories of a Quantified-Self learning app

Using that data, QS apps usually provide two views. For the case of QS sports apps, user can evaluate their steps on a daily basis as well as average values over certain time periods. For QS learning apps, user can evaluate their recent performance after tracking data (cf. user story 3, Table 1), which can be categorised as a part of the appraisal phase (cf. Figure 1). On the other hand, user data is aggregated over a longer period of time (cf. user story 4, Table 1). As each iteration of a process generates data (cf. user story 2), it is fed back in the performance phase of the higher-level process (cf. Figure 1). The collected data can be aggregated in that phase to self-control the learning and adapt further iterations of the lower level process. Applied to the context of learning, data (e.g. learning time) is captured at the end of each learning session and can be seen right after performing the learning; meanwhile, the data over several sessions can be aggregated (e.g. average learning time a day) across a term while learning sessions are performed. Therefore, even if the aggregated analysis is assigned to the performance phase, it refers to the related appraisal phase and the preparatory phase of the next iteration.

Facing the appraisal phase, self-judgement includes the comparison of the learner's performance with previous activities or the performance of other learners' behaviour (cf. user story 5, Table 1), which can be implemented by visualising absolute as well as aggregated data (cf. Figure 1). Comparisons of the performance with fellow students can affect the appraisal and preparatory phase (Schumacher & Ifenthaler, 2018). Self-reacting is based on the learner's satisfaction with the learning performance and adaptive decisions for further process cycles. As those activities are cognitive reactions (Zimmerman & Moylan, 2009) and cannot be assessed by statistical evaluations, they are not captured as a user story. Although the visualisation of learning data, which provides the user with objective information, can affect the subjective impression of a learner, this is more likely an overall goal of that task than a user story.

The five user stories illustrated in Table 1 are developed based on the theory of SRL and QS in learning/ learning analytics in general. Facing the second research question (i.e. what are possible visualisations of these user stories), not all user stories are considered further. We decided to focus on the development of visualisations for user stories 3, 4, and 5 as they provide the highest potential for large and diverse data. User story 1, which refers to data gathered while planning (preparatory phase), and user story 2, which relates to the collection of data itself, are not included in proposing a concrete visualisation of a QS learning app through the example of QSEL.

The data collected by QSEL includes parameters regarding learning condition (location, duration, time, and type of breaks), personal condition before/after a learning session (mood, atmosphere, perceived productiveness, and concentration), and daily self-evaluation (sleep, food, sports). Data is gathered on two different occasions. Firstly, session-based data at the end of the performance phase is used. Secondly, term-based grades at the end of the performance phase are recorded. Subsequent to the performance phase of the academic learning level, no data is gathered within QSEL as the high-level process has only a few iterations (i.e. achieving a Bachelor and a Master degree). For the moment, QSEL focus on session-based and term-based data. Hence visualisations will focus on the learning data on session-based and term-based learning levels.

3.0 Visualisations

The visualisations were created by a project team of eight graduate students of information systems or management. Two design teams were formed and each developed visualisations of the learning data, which easily enables the users to follow their learning process via graphical analyses.

In the following, the resulting eight screens, representing three different applications within the session-based and term-based learning levels are presented. Originally developed and evaluated in German, the visualisations were translated into English by the authors.

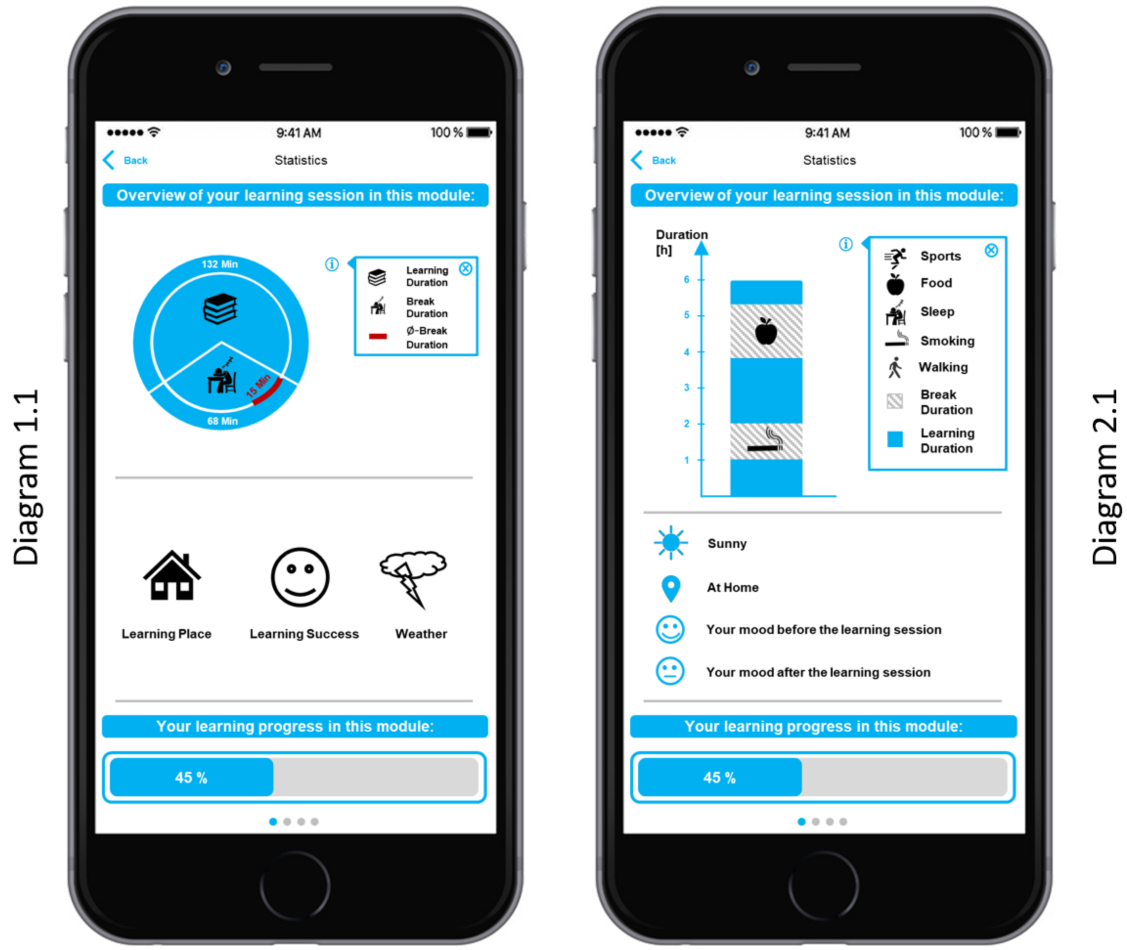

Figure 3. **Visualisation of session data: diagram 1.1 from design team 1 (on the left) and diagram 2.1 from design team 2 (on the right)**

As the session-based learning level is the most detailed level considered, there is no data of lower process iterations which can be aggregated (cf. user story 3). Therefore, only an illustration of absolute data is possible (cf. Figure 3). Figure 3 presents the screen immediately following a completed learning session to provide feedback right after entering learning data. Both diagrams 1.1 and 2.1 show learning and break duration. Additionally, the data just gathered concerning learning environment and mood is shown and the learners receive information about their progress concerning the module.

The gathered session data of each iteration of the cyclical process is fed in the performance phase of the term-based learning level, where aggregated analysis can be provided (cf. user story 4).

Figure 4. **Visualisation of aggregated session data: diagrams 1.2 and 1.3 from design team 1 (on the left) and diagrams 2.2 and 2.3 from design team 2 (on the right)**

Figure 4 and Figure 5 contrast overall data, independent of a learning session and partially module-independent. Diagrams 1.2 and 2.2 compare learning duration and learning location with success versus average mood during the sessions, grouped by weekdays. Diagrams 1.3 and 2.3 provide details on the break activities.

Diagrams 1.4 and 2.4 compare one's own learning duration across learning sessions of one module with the duration of fellow students. Diagrams 1.5 and 2.5 illustrate effective learning phases based on the learner's perception, depending on personal constitution versus the time of day. Most of the diagrams of those screens offer to select different timeframes and abstraction levels regarding modules and others, which allows the learner to customise the visualisation.

Figure 5. **Visualisation of aggregated session data: diagrams 1.4 and 1.5 from design team 1 (on the left) and diagrams 2.4 and 2.5 from design team 2 (on the right)**

At the end of the performance phase on the term-based learning level, data on grades is entered and can be checked during the appraisal phase (cf. user story 3), illustrated in Figure 6. Both design teams proposed to collect grades in addition to the current data, as it is the “best available measure of students’ actual learning outcome” (Ott, Robins, Haden, & Shephard, 2015, p. 177). Diagrams 1.6 and 2.6 provide a performance overview in comparison to fellow students whereas diagrams 1.7 and 2.7 show the grade in relation to the number of hours learned for the module exam. Especially diagrams 1.6 and 1.7 offer the additional benefit of comparing the result with the set goal (Duval, 2011; Schumacher & Ifenthaler, 2018).

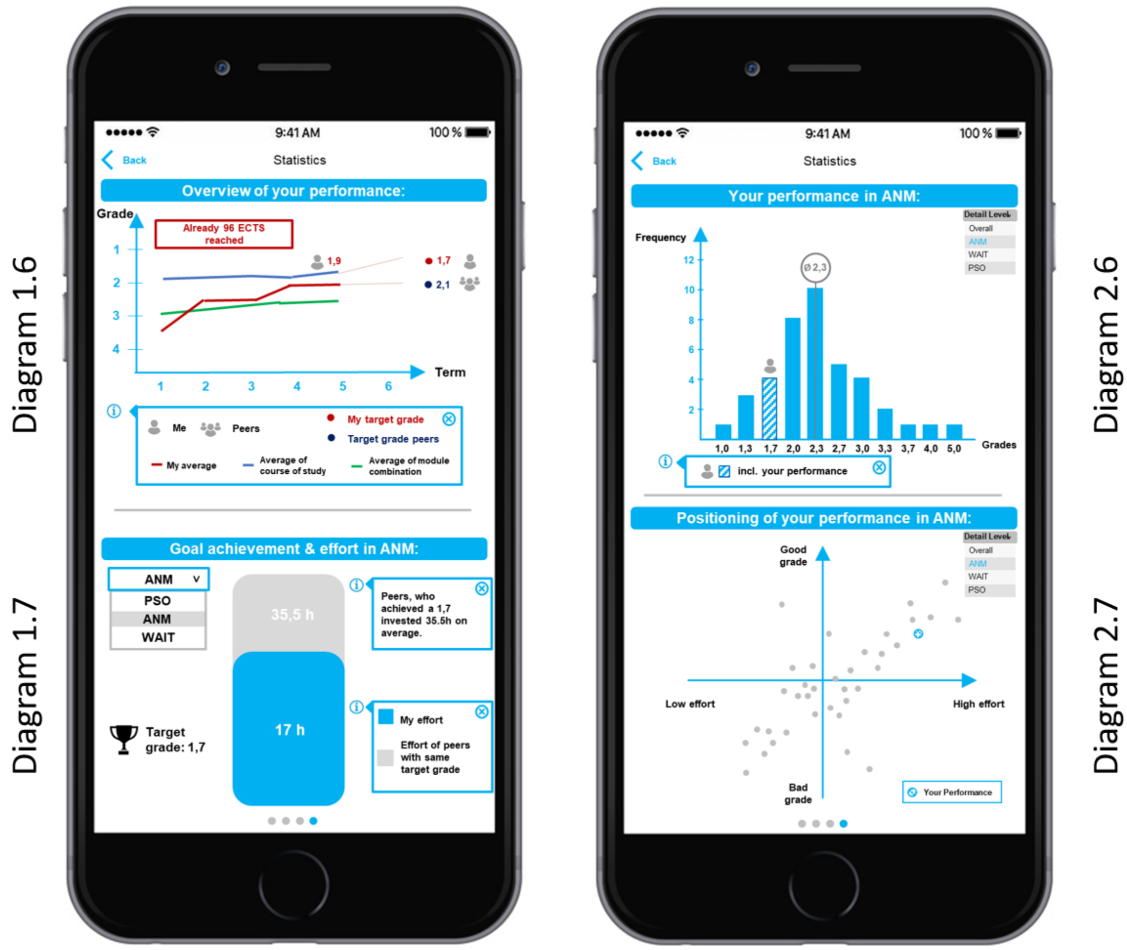

Figure 6. **Visualisation of absolute term-based data: diagrams 1.6 and 1.7 from design team 1 (on the left) and diagrams 2.6 and 2.7 from design team 2 (on the right)**

The different diagrams are evaluated in the following to propose a final result as to how the visualisation of the learning data could be visualised for the purpose of learner support.

4.0 Evaluation

The visualisation shown above will be evaluated following a mixed-method approach aiming to compare the visualisation by design team 1 (i.e. diagrams 1.x) to those by design ream 2 (i.e. diagrams 2.x).

The evaluation started with the qualitative part. Expert interviews were conducted to broaden the designers' own view (particularly as they were graduate students). In parallel, a quantitative online survey was conducted to assess the quality of the visualisations with students as potential future users. Afterwards the results of both methods were integrated.

4.1 A Research Model for Visualisations of the Learning Process

As a theoretical basis for the evaluation, WebQual (Loiacono, Watson, & Goodhue, 2007) is used. WebQual is an adaptation of the Technology Acceptance Model (Davis, Bagozzi, & Warshaw, 1989) and other related theories especially for the evaluation of web-based systems. Based on the WebQual instrument, a diverse range of constructs can be analysed focusing on the concepts of usefulness, ease of use, and entertainment, which eventually increase reuse intention. As our study is concerned with the domain of learning, we decided to add the concept of perceived benefit to the model involving the constructs of metacognition and learning effectiveness (Nussbaumer, Hillemann, Gütl, & Albert, 2015). Using the combined research model depicted in Figure 7, this study evaluates the treatments regarding the different constructs proposed by WebQual.

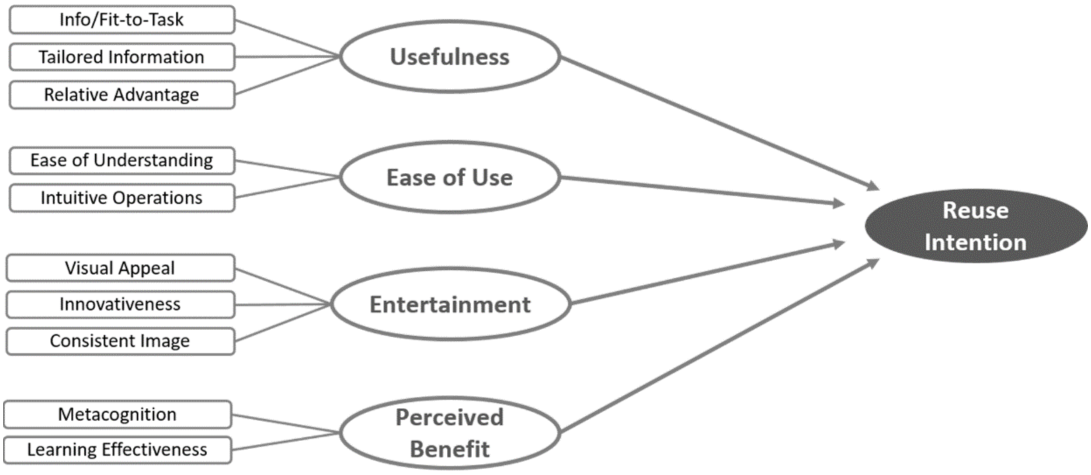

Figure 7. Research Model adapted from Loiacono et al. (2007) and Nussbaumer et al. (2015)

4.2 Qualitative Interviews

The qualitative investigation was conducted as an expert evaluation of the visualisations. A diverse group of three experts in the fields of e-learning and usability engineering (interviewee 1), information design and innovative learning (interviewee 2), and app development (interviewee 3) could be acquired. Each of the participants is an active researcher and practitioner in their domain with several years of experience. A semi-structured interview was conducted with each of the experts asking them to evaluate the diagrams of both treatments. The interview guideline was derived from the research model to structure the questions. The interviews were recorded and transcribed for qualitative content analysis following an open coding approach. The goal of these interviews was to generate diverse feedback on the visualisations and concrete potential for further improvement. To reflect commonalities and links among the codes, categories were generated from the coding. In a second step more generalisable categories were developed which include one or more first-order codes, classifying them thematically into thirteen second-order-themes. In the last step we identified three core dimensions by abstracting the individual second-order-themes.

4.3 Quantitative Survey

For the quantitative evaluation, a survey was constructed based on the research model. Survey items were adopted from the WebQual model (Loiacono et al., 2007) and from Nussbaumer et al. (2015) translated into German to fit the target population and the domain of self-tracking in the learning domain. Additionally, control variables were added regarding experience of the participants using mobile devices and motivation (Gimpel, Nißen, & Görlitz, 2013). A total of 460 students participated in the online survey divided into 14 randomized treatment groups (group size 30 – 37 participants), who each evaluated two to three visualisations separately.
They were required to fill in the survey over social media without any further incentive. 24 datasets had to be removed due to missing or inconclusive data leading to a final dataset of 436 participants.

5.0 Results

In the following, the results will be presented beginning with the interviews, followed by the survey.

5.1 Expert Interviews

The open coding of the interviews led to the first-order categories, second-order categories, and core dimensions shown in Figure 8.
The dimension *change in behaviour* consists of three underlying categories, namely *recommendations*, *goals* and *comparisons*, which can induce a change in students' learning habits. The interviewees often request recommendations to improve the learning e.g.:

> "[...] if you want people to use your app, you just have to offer the people a benefit. And perhaps also offer such help as: when is it useful to take breaks? So you know when it makes sense. So 30 minutes is e.g. a good time and then five minutes break and then work again." (Interviewee 2, translated from German by the authors)

Hereby the interviewees refer to the provision of recommendations which, on the one hand, are based on theoretical findings (e.g. recommendations regarding sleep duration or the ratio of learning and breaks) and, on the other hand, generated by analysing existing learning data.

Furthermore, setting goals and comparisons to other learners or to oneself can help to achieve a behavioural change. It must be noticed that according to the experts, comparisons can lead to either positive or negative behavioural changes in a way that learners tend to get motivated or also discouraged if there are e.g. large discrepancies between one's own performance and that of others.

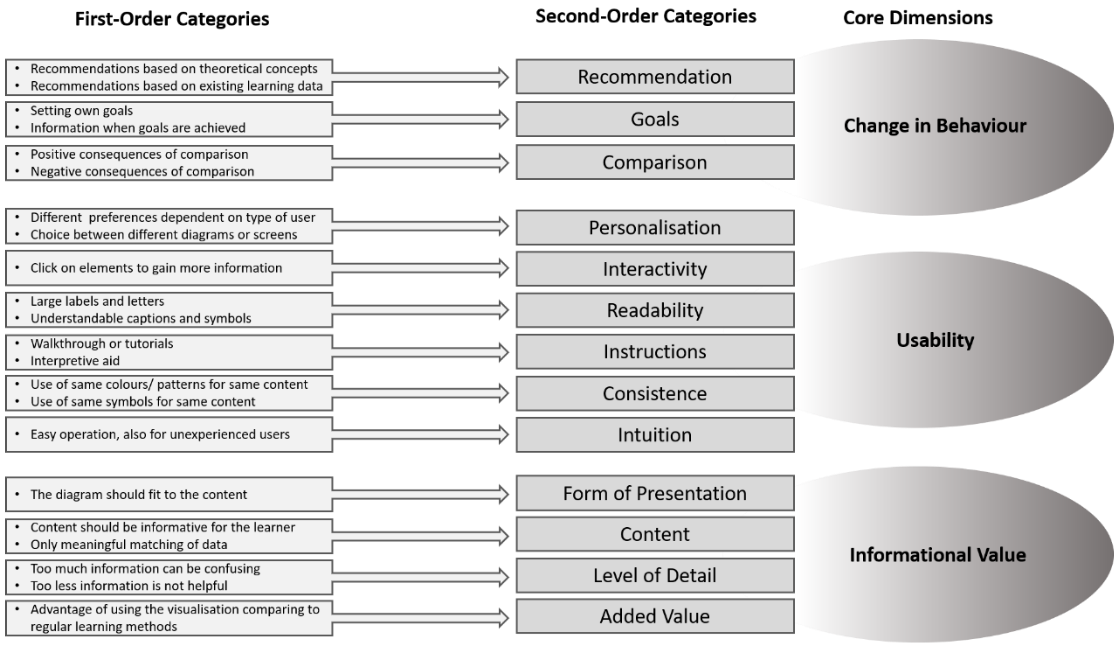

Figure 8. Results of the interviews

The usability dimension to which the visualisations belong contains *personalisation*, i.e. it should be possible to customise the data according to personal preferences.

Moreover, *readability* in terms of large labels such as understandable captions, symbols and *instructions* such as tutorials or interpretive aids are mentioned. A consistent design is deemed to ensure e.g. a uniform use of colours or symbols and to support an intuitive operation (*intuition*) are important aspects, too.

The third core dimension, i.e. *informational value*, focuses on the data used. According to the respondents it is essential to choose a *form of presentation* which is adequate to the content conveyed otherwise information could get distorted or lost.

Moreover, only a meaningful matching of the available data leads to useful *content* which can be informative for the learner. The presented content can vary in its *level of detail*. Hereby, it is important to choose an adequate depth of information within the continuum of providing enough information and being stressed by too much information. This also alludes to the category *added value* which is the benefit a learner receives by using the visualised data

> "[…] because I just get the same data displayed differently. There the advantage of the app to paper is surely huge." (Interviewee 1, translated from German by the authors)

Comparing the treatments to each other, the interviewees had different opinions. Interviewee 1 generally rated diagrams 2.x higher due to his perception of a more intuitive design. Only for the visualisation of session data (cf. Figure 3) he preferred treatments 1.1 and 1.2 because of the higher informational content. In contrast, the second interviewee, preferred the first four diagrams of treatment 1, as he put more emphasis on the amount of the provided information which is higher in those. Nevertheless, he noted that these visualisations would need further instructions. The last three visualisations and especially the grading scale of treatment 2 (Figure 6, diagram 2.6) were preferred by this expert as they are much more intuitive and offer higher benefit to learners. In contrast, the third interviewee preferred the grading scale of treatment 1 (Figure 6, diagram 1.6) due to the higher amount of content. However, the third expert did not favour one treatment in total and often had different favourites. Summing up, we can say that the experts had different opinions concerning the treatments and diagrams and, therefore, had different preferences. Table 2 provides a summary of advantages and disadvantages reported by the interviewees.

Diagram	Treatment 1	Treatment 2
1	+ More information + More intuitive - Explanation required	
2	+ More intuitive - Explanation required	+ More intuitive
3	+ More intuitive - Explanation required	+ More intuitive
4	+ More intuitive - Explanation required	+ More intuitive
5		+ More intuitive + High perceived benefit
6		+ More intuitive + High perceived benefit
7		+ More intuitive + High perceived benefit

Table 2. Advantages and disadvantages of the visualisations retrieved from the interviews.

5.2 User Survey

The data collected from the user survey include 57.6% females and 42.4% male students. The sample exhibits an average age of 23.9 (SD=3.32) years. Participants were mostly German students with 5% international students in the sample being enrolled in a Bachelor (52%) or Master (44.8%) programme. Participants revealed mediocre experience with self-tracking (M=4.56, SD=1.52). The randomised generation of treatment groups lead to treatment 1 having a significantly higher experience compared to treatment 2 (M(T1)=4.781, M(T2)=4.404, t(422)=-2.615, p=0.009, r=0.126). The remaining control variables self-entertainment, self-design, and self-discipline showed no effects on the treatment groups.

While most constructs produce acceptable results regarding Cronbach's alpha being between 0.800 and 0.930, learning effectiveness has to be discarded due to conflicting items (Nunnally & Bernstein, 1994). Over all visualisations, the constructs relative advantage, ease of understanding, and intuitive operations are rated the highest, while info/fit-to-task, tailored information, and innovativeness are rated the lowest (cf. Table 3).

No. of Items	Construct	Cronbach's Alpha	Mean	Standard Deviation
3	Info/Fit-to-Task	0.828	5.03	1.18
3	Tailored Information	0.851	4.83	1.20
3	Relative Advantage	0.865	5.75	1.18
3	Ease of Understanding	0.8973	5.74	1.10
3	Intuitive Operations	0.850	5.64	1.04
3	Visual Appeal	0.930	5.45	1.27
3	Innovativeness	0.874	4.74	1.29
3	Consistent Image	0.881	5.33	1.12
2	Metacognition	0.800	5.12	1.27
2	Learning Effectiveness	-1.421	4.21	0.83

Table 3. Validity and descriptives over all visualisations. All items measured on a 7-point Likert scale from 1 "totally disagree" over 4 "neutral" to 7 "totally agree" (N=436).

Figure 9 reports the mean values per criterium and diagram separately (low values in red, high values in green). It confirms, that treatment 2 is on average higher rated than treatment 1. While this effect is mainly visible for info/fit-to-task (e.g. diagrams 1.1., 1.3, 1.5.), regarding the criterium of tailored information is only met by diagram 1.7. For diagram 1.5, rather low means are reported for the criteria ease of understanding, intuitive operations, visual appeal, and consistent image. Regarding innovativeness, low means are observed in general, only seeing an exception for diagram 1.7 having a mediocre value. Finally, regarding metacognition good values are observed for diagrams 1.7 and 2.2., while the lowest mean is observed for diagram 1.3.

Diagram	N	Info/Fit-to-task	Tailored Information	Relative Advantage	Ease of Understanding	Intuitive Operations	Visual Appeal	Innovativeness	Consistent Image	Metacognition	Total
1,1	30	4,7111	4,7444	5,7556	5,9111	5,7222	5,6667	4,3778	5,2667	5,0667	**5,2469**
1,2	30	4,8778	4,6444	5,4556	5,5333	5,4333	5,0889	4,8000	5,0222	4,9833	**5,0932**
1,3	30	4,5889	4,4444	5,4444	5,9778	5,6778	5,7778	4,8333	5,0333	4,4167	**5,1327**
1,4	30	4,9222	4,9222	5,8778	5,6000	5,6333	5,2000	4,4556	5,2222	5,1500	**5,2204**
1,5	30	4,5444	4,9111	5,5889	5,0778	4,8111	4,8333	4,7111	4,7667	4,8000	**4,8938**
1,6	30	5,0000	4,8111	5,6889	5,4556	5,3444	5,0000	5,0000	5,2667	4,9833	**5,1722**
1,7	30	5,3000	5,1556	5,8444	5,7333	5,6667	5,3667	5,3111	5,4667	5,5667	**5,4901**
2,1	31	5,3548	4,7634	5,8495	6,2366	5,9462	5,8817	4,5806	5,6022	5,3548	**5,5078**
2,2	31	5,3118	5,0538	5,8387	5,6882	5,7634	5,5591	4,6989	5,3871	5,6935	**5,4438**
2,3	31	4,9677	4,8387	5,6989	5,7634	5,8817	5,6989	4,6559	5,3978	5,1290	**5,3369**
2,4	34	4,9608	4,4706	6,0196	5,8922	5,7157	5,4902	4,5882	5,3333	5,0294	**5,2778**
2,5	33	5,1919	5,0303	5,7980	5,6162	5,7172	5,4848	4,6869	5,4242	5,3030	**5,3614**
2,6	33	5,2929	4,7576	5,7071	6,1212	5,8990	5,7273	4,5455	5,7273	4,8788	**5,4063**
2,7	33	5,2828	5,0404	5,8485	5,7374	5,6768	5,5152	5,0909	5,6061	5,3182	**5,4574**
Total	**436**	**5,0275**	**4,8272**	**5,7477**	**5,7431**	**5,6399**	**5,4541**	**4,7370**	**5,3295**	**5,1216**	**5,2920**

Figure 9. Mean values for WebQual criteria per diagram

To compare the created treatments a one-way ANOVA is performed. The analysis shows significant effects from the treatments regarding ease of understanding ($F(13, 422)=2.220$, $p=0.008$, $\omega=0.035$), intuitive operations ($F(13, 422)=2.404$, $p=0.004$, $\omega=0.040$), visual appeal ($F(13, 422)=1.906$, $p=0.028$, $\omega=0.026$), and metacognition ($F(13, 422)=2.031$, $p=0.017$, $\omega=0.030$), while no differences can be found for info/fit-to-task, tailored information, relative advantage, and innovativeness.

As a follow-up analysis, contrasts are defined to compare especially those diagrams with each other, which display the same contents. For ease of understanding, diagrams 2.5 ($M(T1)=5.078$, $M(T2)=5.616$, $t(422)=1.970$, $p=0.049$, $r=0.095$) and 2.6 ($M(T1)=5.456$, $M(T2)=6.121$, $t(422)=2.436$, $p=0.015$, $r=0.118$) are evaluated significantly better compared to the corresponding diagrams of treatment 1. A similar effect can be shown regarding intuitive operations, where diagram 2.5 is again evaluated superior ($M(T1)=4.811$, $M(T2)=5.717$, $t(422)=3.098$, $p=0.003$, $r=0.416$) whilst diagram 2.6 is evaluated considerably better for visual appeal ($M(T1)=5.000$, $M(T2)=5.727$, $t(422)=2.280$, $p=0.027$, $r=0.304$). Finally, regarding metacognition, diagrams 2.2 ($M(T1)=4.983$, $M(T2)=5.694$, $t(422)=2.208$, $p=0.028$, $r=0.107$) and 2.3 ($M(T1)=4.417$, $M(T2)=5.129$, $t(422)=2.215$, $p=0.027$, $r=0.107$) are evaluated significantly higher compared to their counterparts from treatment 1. While the control variables mainly do not exert effects on the treatment groups, a small significant effect is revealing that participants evaluating diagrams 1.4, 1.5, 1.6, and 1.7 had a higher self-discipline than their counterparts in treatment 2 (M(T1)=5.622, M(T2)=4.922, t(52.662)=-2.134, p=0.038, r=0.282). The results are summarised in Table 4.

Diagram	Treatment 1	Treatment 2
1	-	-
2	$M_{metacognition}$=4.983*	$M_{metacognition}$=5.694*
3	$M_{metacognition}$=4.417*	$M_{metacognition}$=5.129*
4	-	-
5	$M_{ease\_of\_understanding}$=5.077* $M_{intuitive\_operations}$=4.811**	$M_{ease\_of\_understanding}$=5.616* $M_{intuitive\_operations}$=5.717**
6	$M_{ease\_of\_understanding}$=5.456* M_{visual_appeal}=5.000*	$M_{ease\_of\_understanding}$=6.121* M_{visual_appeal}=5.727*
7	-	-

Table 4. Comparisons between treatments **p<0.01, *p<0.05

6.0 Discussion

In the following section, the results from the qualitative as well as from the quantitative data analysis are summarised and combined.

On the one hand, the survey participants (as potential future users of the app) rate the visualisations in general to be rather easy to understand, intuitive to operate, and eventually providing a relative advantage for them. This is consistent to the interviewees, which find the diagrams mostly informative and intuitive. Contrasting the easiness to understand the diagrams, the interviewees requested further explanation for some of them. On the other hand, they rate learning effectiveness, innovativeness, and tailored information to be low. While the low rating of learning effectiveness might be due to low validity of the measurement items (cf. Table 3), low innovativeness and tailored information are consistent to the feedback obtained in the interviews. Interviewees suggested improving the diagrams including tooltips as well as the possibility for customisation (e.g. selecting variables and dimensions to visualise dynamically), which offers additional benefit to the user (cf. Duval, 2011; Keim, 2002).

	Quantitative Choice	Expert 1	Expert 2	Expert 3	Summarised Evaluation
Diagram 1	MU2	MU1	MU1	MU1	MU2
Diagram 2	MU2	MU2	Indifferent	MU2	MU2
Diagram 3	MU2	MU2	MU1	MU2	MU2
Diagram 4	MU2	MU2	MU1	Indifferent	MU2
Diagram 5	MU2	MU2	MU2	MU2	MU2
Diagram 6	MU2	MU2	MU2	Indifferent	MU2
Diagram 7	MU1	MU2	MU2	MU2	MU1

[1] Absolute increase/decrease in mean value of reuse intention by choosing MU2 instead of MU1 (on 7-point Likert scale)

Table 5. Results of analysis

Comparing the treatments, the interviewees favoured diagrams 1 – 4 of treatment 1 and diagrams 5 – 7 of treatment 2, albeit discussing different advantage and disadvantages (cf. Table 5). However, survey participants did not confirm this result, as they rather favoured treatment 2 (especially diagrams 2, 3, 5, 6) over treatment 1. While diagrams 2.2 and 2.3, showing aggregated session data, were favoured due to higher potential for metacognition, diagrams 2.5 and 2.6, showing aggregated term-based data, were preferred due to their ease of understanding, intuitive operations, and visual appeal. Differing results might be explained by the possibility of the interviewees to ask for an explanation of the diagrams, more time to assess the diagrams, as well as high expertise in related areas. Nevertheless, combining those findings, a stable trend could be formulated, that diagram 1.1 is slightly dominating its counterpart. Regarding absolute session data, users obviously favour as much information as possible. Diagrams 2-4 provide more information and, thus, more benefit to the users, however they are harder to understand, while the diagrams 2.5 – 2.7 are equally informative but more intuitive than their counterparts. In general, research question 3 can be answered, as the chosen criteria for the evaluation of the visualisations were useful and provided a valid and precise measure, except learning effectiveness, which suffered from low validity.

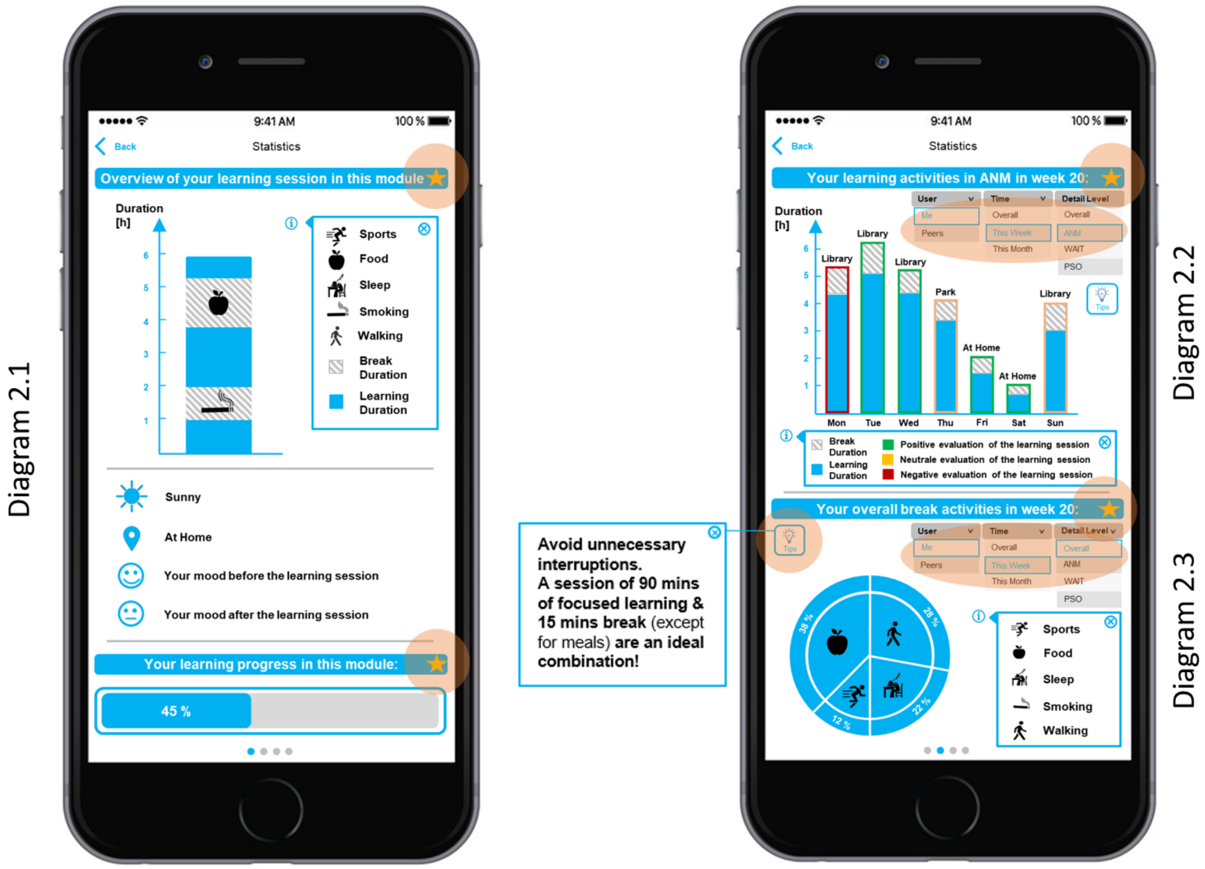

Figure 10. **Visualisation capturing diagram 1, 2 and 3**

Combining the results of the qualitative and quantitative data analyses and improvements stated in the interviews the consolidated visualisations depicted in Figure 10 and Figure 11 have been developed in order to answer research question 2. Based on those diagrams which were favoured by the survey participants, improvements of the interviewees have been implemented, highlighted in red. Thus, the visualisations were extended by filtering features to select specific variables and dimensions to be visualised, as well as providing tooltips and guidance to the users. Furthermore, an additional visualisation (cf. Figure 12) is added, providing a list of frequently accessed analyses and diagrams to the users to add favourite visualisations.

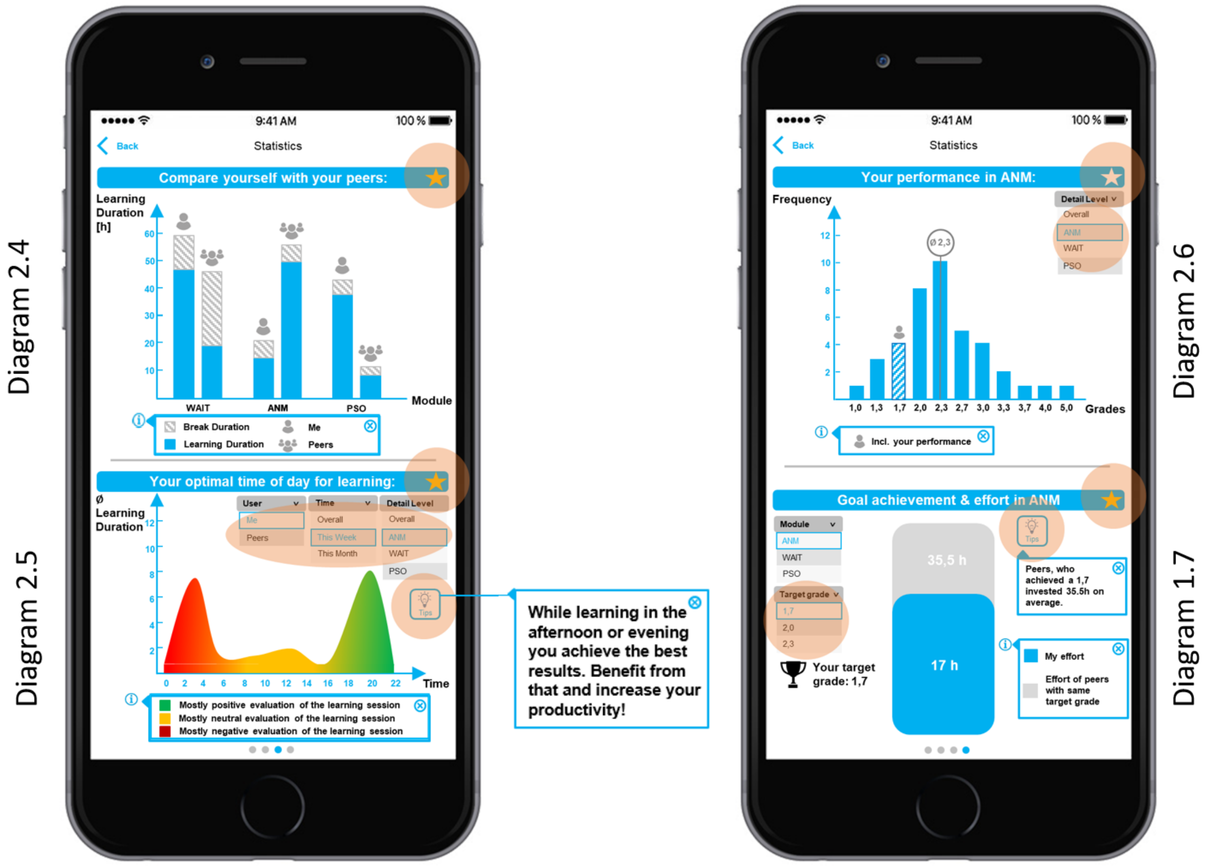

Figure 11. **Visualisation capturing diagram 4, 5, 6 and 7**

Finally, answering research question 1, the results from evaluating the visualisations can be generalised towards the formulated user stories. As stated in the interviews, the visualisations provided the possibility for goal-setting and evaluation on the basis of individual as well as peer-based comparisons. The visualisations are found to provide the necessary benefits to induce the change in behaviour required for the quantified-self.

This study is subject to several limitations. First of all, extensive data analysis of the survey data revealed problems with data quality, indicated by low discriminant validity between the constructs. This may be due to the missing incentives for survey participants and the broad distribution of the survey over social media. Despite the large sample size, the data does not fulfil constraints to perform more complex analyses such as confirmatory factor analysis or structural equation modelling. Furthermore, the construct learning effectiveness had to be removed from descriptive analyses.

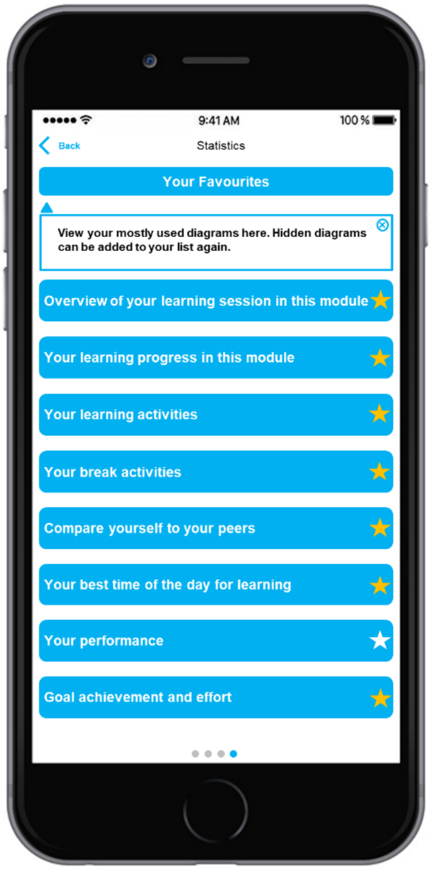

Figure 12. Additional visualisation to choose favourite visualisations

The control variables (experience, self-discipline, self-design, and self-reflection) have been asked for after the participants had evaluated their mock-ups. Thus, we cannot rule out that the treatments have an effect on the control variables. Therefore, they to be interpreted with great care.

The next research step will be the thorough implementation of the visualisations based on the evaluation for extensive field testing. As the research introduced in the current paper was performed by a project group of eight Master students, both the implementation of alternatives, the planning of a large roll-out for testing purposes, and the conduct and analyses of the tests will be the subject of various Master theses and PhD research.

References

Baskerville, R. L. and Pries-Heje, J. (2010) *Explanatory Design Theory*, Business & Information Systems Engineering, 2(5) 271–282.

Davis, F. D., Bagozzi, R. P. and Warshaw, P. R. (1989) *User Acceptance of Computer Technology: A Comparison of Two Theoretical Models*, Management Science, 35(8) 982–1003.

Duval, E. (2011) *Attention please! Learning analytics for visualization and recommendation*, In Proceedings of International Conference on Learning Analytics and Knowledge (LAK), ACM, Banff, Canada.

Gimpel, H., Nißen, M. and Görlitz, R. (2013) *Quantifying the Quantified Self: A Study on the Motivations of Patients to Track Their Own Health*, In Proceedings of International Conference on Information Systems (ICIS), AIS, Milano, Italy.

Keim, D. A. (2002) *Information visualization and visual data mining*, IEEE Transactions on Visualization and Computer Graphics, 8(1) 1–8.

Loiacono, E., Watson, R. and Goodhue, D. (2007) *WebQual: An Instrument for Consumer Evaluation of Web Sites*, International Journal of Electronic Commerce, 11(3) 51–87.

Melzer, P. (2018) A Conceptual Framework for Personalised Learning: Influence Factors, Design, and Support Potentials, Springer Gabler, Wiesbaden.

Melzer, P., Schneider, B. and Schoop, M. (2018) *Die Unterstützung von selbstregulierten personalisierten Lernprozessen durch den Quantified-Self-Ansatz in Theorie und Anwendung*, In Proceedings of E-Learning Fachtagung Informatik (DeLFI).(Eds, Krömker, D., Schroeder, U.) Gesellschaft für Informatik, Frankfurt/Main, Germany.

Neitzel, S. and Rensing, C. (2017) *Automatische Sammlung von Aktivitäten Lernender in offenen Lernumgebungen und deren Nutzung in einer Lerntagebuchanwendung*, In Proceedings of E-Learning Fachtagung Informatik (DeLFI).(Eds, Igel, C., Ullrich, C., Wessner, M.) Gesellschaft für Informatik, Chemnitz, Germany.

Nunnally, J. C. and Bernstein, I. H. (1994) Psychometric theory, 3. ed., McGraw-Hill, New York.

Nussbaumer, A., Hillemann, E.-C., Gütl, C. and Albert, D. (2015) *A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments*, Journal of Learning Analytics, 2(1) 101–133.

Ott, C., Robins, A., Haden, P. and Shephard, K. (2015) *Illustrating Performance Indicators and Course Characteristics to Support Students' Self-Regulated Learning in CS1*, Computer Science Education, 25(2) 174–198.

Panadero, E. (2017) *A Review of Self-regulated Learning: Six Models and Four Directions for Research*, Frontiers in psychology, 8:422.

Piotrkowicz, A., Dimitrova, V., Treasure-Jones, T., Smithies, A., Harkin, P., Kirby, J. and Roberts, T. (2017) *Quantified Self Analytics Tools for Self-regulated Learning with myPAL*, In Proceedings of Workshop on Awareness and Reflection in Technology Enhanced Learning co-located with European Conference on Technology Enhanced Learning (EC-TEL 2017).(Eds, Kravcik, M., Mikroyannidis, A., Pammer-Schindler, V., Prilla, M.) Tallinn, Estonia.

Schumacher, C. and Ifenthaler, D. (2018) *Features students really expect from learning analytics*, Computers in Human Behavior, 78 397–407.

Zimmerman, B. J. and Moylan, A. R. (2009) *Self-regulation: Where metacognition and motivation intersect*, In Handbook of Metacognition in Education.(Eds, Hacker, D., Dunlosky, J., Graesser, A.) Routledge, New York, US, pp. 299–315.