Design and Evaluation of an IT-based Formative Feedback Tool to Foster Student Performance

Completed Research Paper

Roman Rietsche
University of St.Gallen
Müller-Friedberg-Str. 8,
CH-9000 St.Gallen
roman.rietsche@unisg.ch

Kevin Duss
University of St.Gallen
Müller-Friedberg-Str. 8,
CH-9000 St.Gallen
kevin.duss@student.unisg.ch

Jan Martin Persch
University of Kassel
Pfannkuchstrasse 1, D-34121 Kassel
persch@uni-kassel.de

Matthias Söllner
University of St.Gallen
Müller-Friedberg-Str. 8,
CH-9000 St.Gallen
matthias.soellner@unisg.ch &
University of Kassel
Henschelstr. 4, D-34121 Kassel
soellner@uni-kassel.de

Abstract

Today's world is changing faster than ever before. Students need to constantly acquire new knowledge and the required skills shift more and more to higher order thinking skills. One major approach to support the students in this shift is through formative feedback, which initiates self-regulated learning processes. However, given the constant rise of student numbers while public spending remains constant, one of the major challenges is how to effectively provide formative feedback in large-scale lectures. The aim of our project is to develop an IT-tool which provides students with formative feedback to increase their performance in large-scale lectures. Our results show that through the use of the IT-tool the objective and perceived student performance can be significantly increased. By documenting our design knowledge according to the eight components of Gregor and Jones (2007), we take a first step towards creating a design theory for formative feedback tools.

Keywords: Formative feedback, IT-tool, large-scale lectures, student performance, design theory

Introduction

Today's world is changing faster than ever before. Students graduating from university find a world which is highly volatile. Decisions need to be made under high uncertainty. Problems to solve are often complex and interdependent. Large amount of data leads to ambiguity during decision making (Bennett and Lemoine 2014). As a consequence the question is: "[how to prepare] students for jobs that don't yet exist, using technologies that haven't been invented, in order to solve problems we don't even know are problems yet?" (Riley 2004). To prepare students for these challenges, a shift in the compositions of skills and knowledge is needed. In the past, students could mostly rely on knowledge they learned once for a long

time. However, today's students need to constantly acquire new knowledge and the required skills shift more and more to higher order thinking skills such as critical thinking, collaboration or problem solving (Fadel et al. 2015).

One major approach to achieve this goal is through formative feedback (Hattie and Timperley 2007; Kluger and DeNisi 1996; Price et al. 2010; Shute 2008). Hattie and Timperley (2007) conducted a meta-analysis with 800 studies, which revealed that formative feedback belongs to the top influential factors for student performance. Formative feedback initiates self-regulated learning processes, which are a prerequisite to achieve higher order thinking skills. The processes are triggered through unveiling the gap between the student's self-assessment of the current state of knowledge and their actual performance (Shute 2008).

However, given the constant rise of student numbers, one of the major challenges is how to provide formative feedback in large-scale lectures effectively (Fortes and Tchantchane 2010). This challenge affects both campus universities as well as MOOC providers for distance learning. One solution might be to increase the teaching resources by scaling the number of teaching assistants and applying a tutor feedback model such as the one proposed by Marjanovic (2016). However, for most universities, this is not possible due to their financial resource constraints and frequently even decreasing budgets. This development is mirrored in numbers provided by the Organisation for Economic Co-operation and Development (OECD). According to them, the number of students at universities rose from 2005 to 2014 by 15 percentage points in the US and 29 in Germany while public spending for education decreased in the same period by 7 percentage points in the US and 1 in Germany (OECD 2016).

As a solution, we leverage the potentials of modern IT to support lecturers in providing formative feedback. Hence, the aim of our project is to develop an IT tool which provides formative feedback to increase student performance in large-scale lectures. To achieve our goal, we follow the design science research (DSR, Hevner 2007) approach and aim to answer the following two research questions.

RQ1: What are the requirements that should be considered when designing an IT tool that aims at providing students with formative feedback during the learning process in large-scale lectures?

RQ2: To what extent does the IT tool usage increase student performance?

To answer these research questions, the presented paper is structured as follows. In the next section we describe the theoretical background. Afterwards, we describe the research methodology and how we followed the DSR process to design the IT tool (artifact). In particular, it reveals how we derived design principles and features for the IT tool from scientific literature and expert interviews. Further, it is explained, how we evaluated the alpha version and derived additional design principles and features for the beta version. We followed the structure of Gregor and Hevner (2013) and moved the evaluation and discussion in separate sections. The evaluation section shows in detail the outcomes of the one-group field test to evaluate the beta version. The following section discusses the results and presents a documentation of our design knowledge in accordance with Gregor and Jones (2007). In the final section, we conclude our paper, discuss limitations of the study and provide an outlook for future research.

Theoretical Background

The Importance of Feedback in the Context of Learning

To define the term feedback we use the definition of Hattie and Timperley (2007), who have stated that feedback is conceptualized as information provided by an agent, for example a lecturer providing corrective information. Hence, feedback is provided based on student performance. The outcome of feedback is an information specifically relating to the task or process of learning that fills a gap between what is understood and what is aimed to be understood (Sadler 1989). According to Hattie and Timperley (2007) feedback in higher education must answer three major questions. The first question aims at defining what the goals are about. The second question asks which progress is being made towards meeting the goal. The third question asks what activities need to be undertaken in order to achieve better student performance. More precisely, the first question addresses the learning goals related to the task or performance. The judgement concerning the achievement of a learning goal may occur on many dimensions, such as directly "passing a test" or "completing an assignment" (Shute 2008). The second question involves providing information in relation to a task or performance goal. Usually, this is related to prior performance and/or to success or

failure in a particular task. Feedback is effective when it consists of information regarding the progress and/or on how to proceed with a certain task (Black and Wiliam 2009). The last question helps guiding students by providing advice that can lead to greater possibilities for learning. This could involve more self-regulation in the learning process, greater fluency and automaticity, deeper understanding, more strategies and processes regarding the completion of tasks, and more information about what is and what is not understood (Hattie and Timperley 2007). Feedback, as we understand it, is based on a preceding self-assessment (SA) and a computer-based Assessment (CbA) that judges a student's performance (Thelwall 2000). The feedback involves the presentation and direct comparison of the results of these two assessments. Thus, it points out the students' gap between their self-assessed and computer-based assessed knowledge regarding a learning outcome.

Feedback can either be formative or summative, depending on its timing as well as purpose and effect. Formative feedback aims to improve the competences and skills of a student (Yorke 2003). By triggering self-regulated learning processes, it provides students with the possibility to adapt their learning habits. Summative feedback is a final evaluation at the end of study and summarizes a student's performance (Sadler 1989). In its purest form, summative assessment only involves feedback limited to marks, grades or scores. Even if offered earlier, this type of feedback has turned out to be less effective than descriptive feedback (Brown et al. 2016). Therefore, our focus lies on formative feedback.

Self-Assessment

Self-assessment (SA) can be defined as the involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they actually meet these criteria and standards (Boud 1991). Previous research shows that reflection based on SA contributes to student performance (Nicol and Macfarlane-Dick 2006). For example, McDonald and Boud (2003) showed that training in how to carry out SA can improve student performance in the final exam. According to Sadler (1989) students carrying out SA of their own work contribute to improving learning in the course being studied and provide a foundation for lifelong learning. However, SA by itself can lead to an overestimation of ones' skills. Kruger and Dunning (1999) found out, that overestimation often goes along with a lack of metacognitive skills among less skilled participants. Hence, SA should be paired with objective assessment to foster awareness and understanding of how to control the own learning and consequently train metacognitive skills which are essential for the development of higher order thinking skills (Nicol and Macfarlane-Dick 2006).

Computer-based Assessment

Computer-based Assessment (CbA) is an objective judgement of the student performance. Communicating the results of the CbA is determined as feedback (Thelwall 2000). In higher education, diverse types of assessment are realized computer-based, because of the resulting didactical benefits. CbA and feedback allow immediate interventions in the learning process with the objective of closing the gap between the current and the pursued understanding. Students are therefore able to timely adapt their learning habits to reach their aimed performance goal (van der Kleij et al. 2012). Feedback based on CbA remains unbiased and is accurate if it is carefully designed at the beginning. Moreover, this kind of feedback is nonjudgmental and ignores irrelevant characteristics of students (Mason and Bruning 2001). One main advantage of CbA is the efficiency, which saves the lecturers time and the university costs. Once the initial programming is completed, computers are basically able to provide students with unlimited feedback (Thelwall 2000). This circumstance makes CbA interesting for large-scale lectures.

Student Performance

There are different definitions of student performance that depend on the purpose of teaching and learning. The purpose of our research is to evaluate if student performance can be increased through the use of an IT tool. Since we rely on an IT tool, we use the learning goals definition and its four categories specified by Gupta et al. (2010). Learning goals which represent student performance, are categorized into skill-based, cognitive, affective and metacognitive goals. Skill-based goals focus on the ability of students to use learned methods such as a business process modelling language. Cognitive goals focus on the mental awareness and judgment of students. Do the students understand the concept of the lecture and are able to choose the right tool for the given business problem? Affective goals focus on the emotional aspects of the students' behavior

such as motivational knowledge (usefulness of the IT tool) or satisfaction with the learning process (Söllner et al. 2017). Metacognitive or self-regulated learning knowledge refers to the students' knowledge regarding their own learning and information processing (Gupta et al. 2010). The goals defining student performance can be considered from two different perspectives. The final exam of a course provides insight into the objective student performance, while a survey allows to gather information about their perceived student performance. We analyze skill and cognitive goals in terms of both objective and perceived performance. Affective and metacognitive goals, however, are analyzed only regarding perceived performance because they cannot be measured by evaluating exam results (Chemers et al. 2001).

Cognitive Dissonance as a Kernel Theory

From an educational perspective, cognitive dissonance can be defined as an uncomfortable internal state occurring when the presented unfamiliar or contradictory information conflicts with existing knowledge (Festinger 1962). The theory of cognitive dissonance holds that individuals have a motivational drive to resolve this dissonance by either changing their beliefs, attitudes and behaviors, or rationalizing their beliefs, attitudes, and behaviors.

Presenting students the results of their SA and CbA could lead to such cognitive dissonance. Empirical studies of cognitive dissonance confirm that the need to resolve this dissonance is extremely motivating for students and activates cognitive processes until the dissonance is resolved (Elliot and Devine 1994). One of the key aspects of the cognitive dissonance theory (CDT) is the relationship between the level of cognitive dissonance and the motivation to resolve it. Students might be quickly bored with a level of dissonance that is too easily resolved, but can be frustrated with a level of dissonance that is too high (Festinger 1962). According to Piaget (1985), the dissonance is seen as an essential trigger for the learning process resulting that students engaged in problem-solving activities and constructing new knowledge structures. As an added benefit to the learning process, the motivational aspects of resolving cognitive dissonance create an environment where students are continually exposed to content-relevant information facilitating deeper processing.

Research Methodology

To achieve our research goal, we used the three cycle DSR approach of Hevner (2007). We followed this approach because we wanted to a) use a scientific method to solve a set of practical problems that researchers and practitioners experience in their own practice and b) contribute to the existing body of knowledge by designing and evaluating a new research artifact and documenting our design knowledge according to Gregor and Jones (2007). Moreover, to ensure that our IT tool addressed all important aspects to increase student performance, we followed a theory-driven design approach by grounding our research on the formative feedback theory (Black and Wiliam 2009) and CDT (Festinger 1962). In this paper, we present the details of all phases advocated by Hevner (2007) for the development of the IT tool. Figure 1 shows the steps that have been carried out as well as their order and the corresponding subsections which contain detailed information. The first step includes the problem formulation. The practical driven motivation of the problem was discussed in detail in the introduction section. Based on the problem definition, in the second and third step, we derived user stories and requirements for the design of the artifact from a) scientific literature and b) six semi-structured interviews with lecturers and students. The length of the interviews was timed between 40 and 60 minutes. In the fourth step, we derived design features based on the requirements for the alpha version and implemented them. In the fifth step the evaluation of the alpha version was conducted. In the sixth step, the design features of step four had been revised based on the evaluation results of step five. The outcome of this step was the beta version. In the seventh step, the beta version was evaluated in a one-group field test with 194 students of a master course in management information systems. In the eighth step, the design knowledge is documented according to the eight components for documenting a design theory (Gregor and Jones 2007). To improve readability, we discuss step seven and eight in separate sections. Hence, step seven "evaluation of the beta-version" can be found in section five *Evaluation of Beta Version* and step eight "documentation of design knowledge" in section six Discussion. For the evaluation of our artifact we followed the evaluation framework proposed by Venable et al. (2016). We followed this process as it describes a systematic approach evaluating each step of the artifact design process. We performed an artificial ex ante evaluation of the design features of the alpha version and a naturalistic ex post evaluation of the design features of the beta version.

**Figure 1. Three Cycle Design Science Research Process by Hevner (2007)**
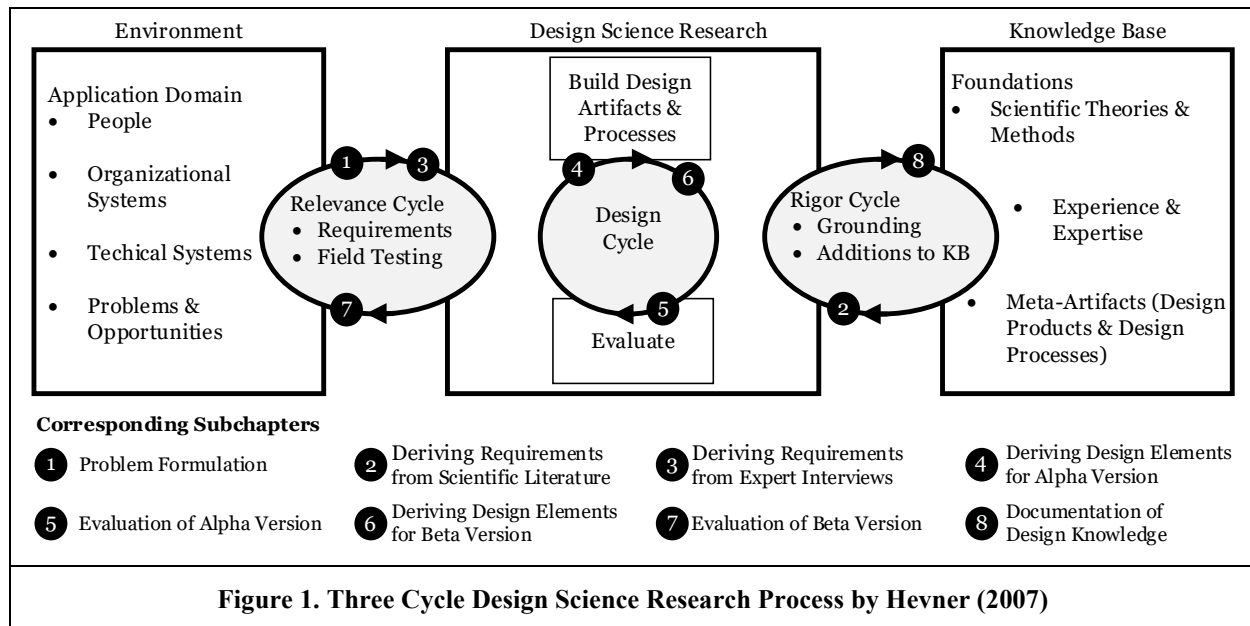
Designing and Evaluating the Formative Feedback Tool

Based on the problem formulation (step one) in the introduction section, this section discusses how we derived the (meta-)requirements as well as the design principles and features, relevant for the development of the formative feedback tool (FFT). All insights we gathered, are summarized in Figure 2.

The basic structure of the FFT includes different levels such as courses, learning units (LU), learning outcomes (LO) and computer-based assessments (CbA). Every course in FFT can have multiple LUs. A LU normally represents one lesson and the topics covered in it. In turn, every LU can have multiple LOs. A LO defines an objective pursued by the students within a LU and is usually formulated in one sentence. Every LO can involve multiple CbAs checking the knowledge of students. However, a CbA is not necessarily assigned exclusively to one LO. It can also be integrated in other LOs which may belong to different LUs or courses. The lecturers weight CbAs when assigning them to a LO by allocating a number between 10 and 100 percent.

Step 2: Deriving Meta-Requirements from Scientific Literature

In this section, we gather meta-requirements from scientific literature for the fundamental development of the FFT (Gregor and Jones 2007). The first meta-requirement (MR1) deals with the three questions feedback should answer (Hattie and Timperley 2007) which we discussed in the subsection The Importance of Feedback in the Context of Learning. The second meta-requirement (MR2) deduced from scientific literature deals with the observation that if students receive feedback often and regularly, it promotes better monitoring and self-regulation of their learning progress (Bitzer et al. 2016; Gibbs and Simpson 2005). Hence, the FFT should enable students to get feedback at any time and any place. Moreover, students need the possibility to actively monitor their performance over time (MR3). The FFT should provide an overview of the student performance aggregated at LO, LU and course level. The last meta-requirement (MR4) addresses the possibility of students to compare their performance with the performance of fellow students to resolve uncertainties about their own abilities (Festinger 1954). However, the FFT should provide the possibility to disable social comparison mechanisms if a student's demand for competitiveness is low. According to Nicol and Macfarlane-Dick (2006), feedback on the student performance leads to an improved awareness.

Figure 2. Overview of the Design Principles and Features Derived from the Defined (Meta-)Requirements
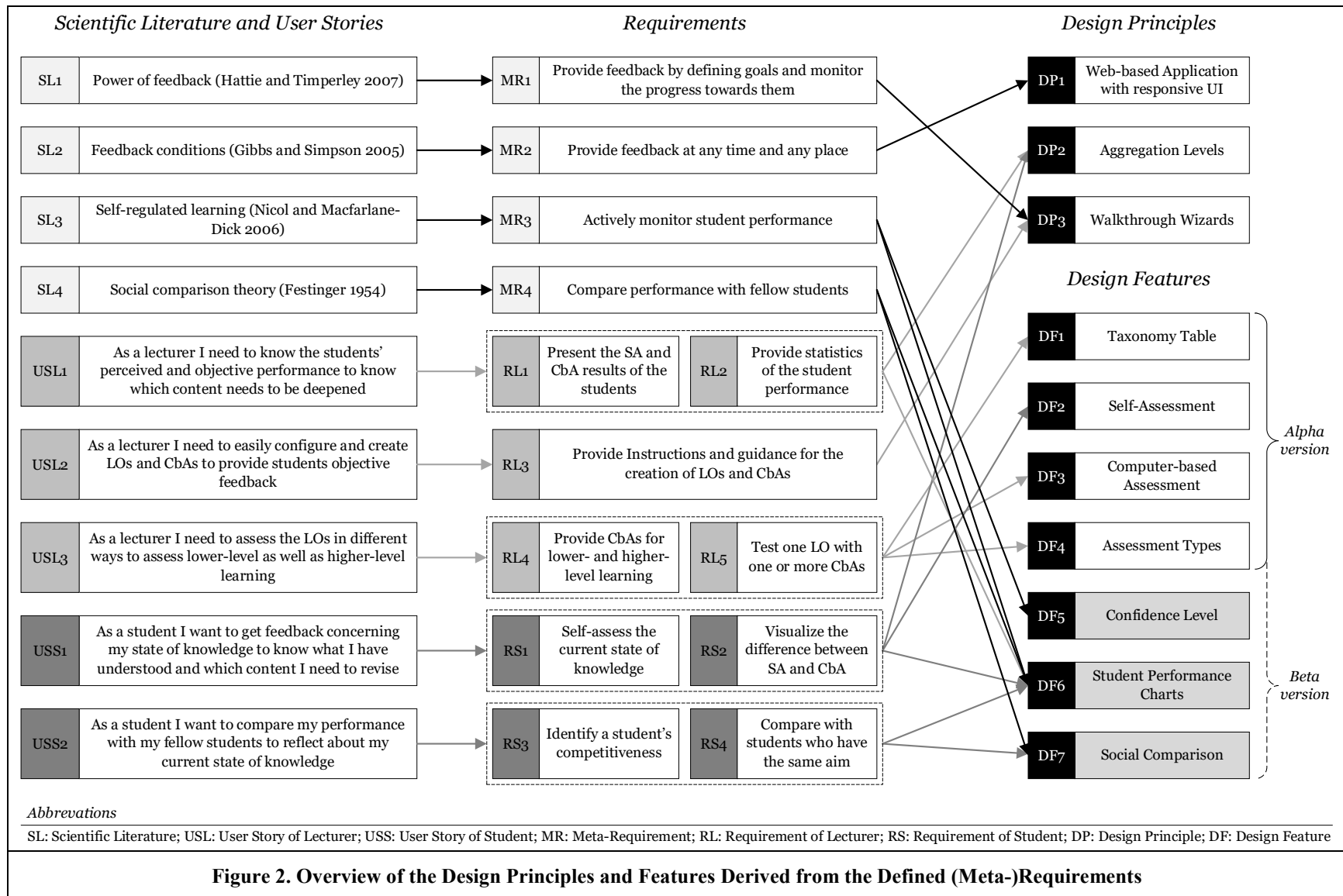
Step 3: Deriving Requirements from Expert Interviews

After defining a set of meta-requirements, we identified user stories and requirements originating from specific field problems of lecturers and students. The first user story of lecturers (USL1) reflects the lack of information of the student performance during the semester. Lecturers have the requirement that the FFT should provide statistical information of the student performance throughout the semester (RL1 and RL2). The second user story (USL2) represents the fact that the FFT should be easy to use and self-explanatory. Hence, the FFT should provide wizards and instructions to minimize the time for adopting the tool (RL3). The third user story (USL3) addresses the fact that lecturers want to develop the students' action competences. Therefore, the FFT should provide the possibility to create LOs and assign multiple weighted CbAs (RL5) for lower-level as well as higher-level learning (RL4).

The first user story of students (USS1) represents their need to know what their current state of knowledge is, and which content needs to be deepened. Hence, the FFT should provide the possibility to carry out SA (RS1) and CbA to the LOs and graphically present the student performance as well as differences between SA and CbA (RS2). The second user story (USS2) indicates the demand of students to compare their student performance with the student performance of fellow students. FFT should provide the possibility to compare the results of individual students to the class average (RS4) based on their competitiveness (RS3).

Step 4: Deriving Design Principles and Features for the Alpha Version

The first design principle (DP) is overarching and must be considered in the implementation of every other DP and design feature (DF). It defines that the emerging tool should be a web-based application with a responsive user interface (UI) in order to be available for lecturers and students from anywhere at any time independent of the operating system in use (DP1). To meet this requirement, we implemented the FFT as a web-based application accessible via browser. The front-end of the FFT was developed with the latest web technologies including Hypertext Markup Language version 5 (HTML5), Cascading Style Sheets (CSS) and JavaScript (JS) as well as the JS library jQuery. Moreover, students should be able to efficiently use the FFT from any web-enabled device regardless of its screen size. To achieve a fully responsive user interface (UI), we made use of the front-end web framework Bootstrap 3 which offers a fluid grid system. As a result, users can access the FFT with different devices such as notebooks, tablets or smartphones running on different operating systems such as Windows, macOS, iOS or Android. Figure 3 shows the front-end of the FFT on a
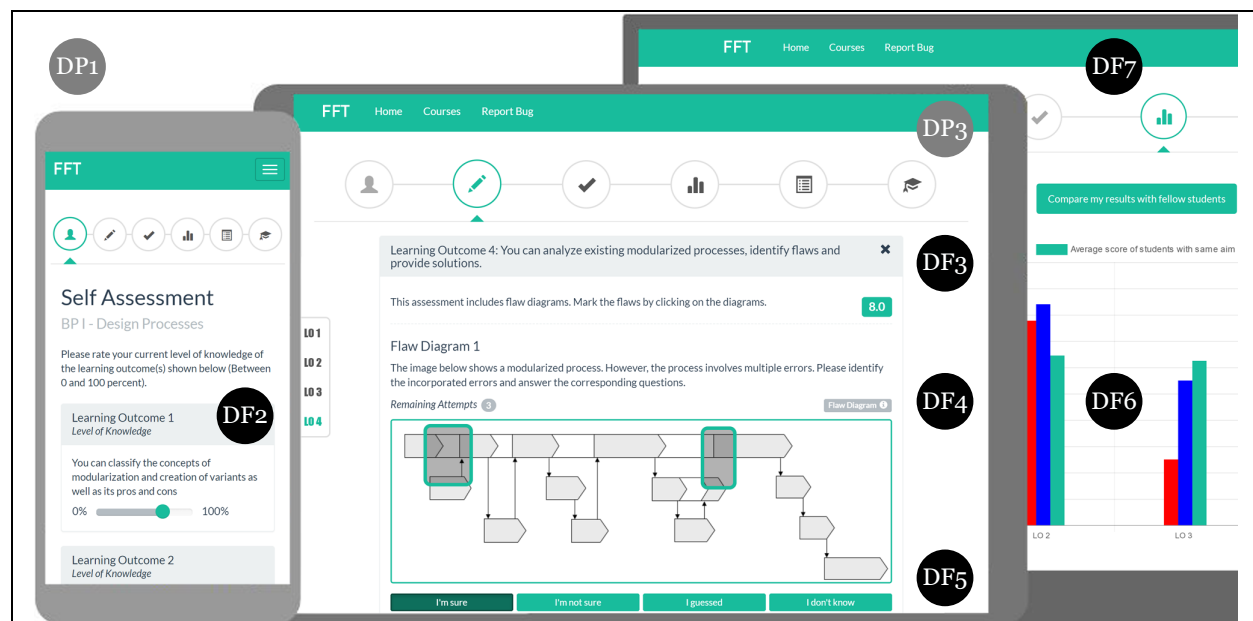

Figure 3. FFT's Student View (Multiple Devices) Including Selected Design Principles and Features

notebook, tablet and smartphone. Students face this view when conducting CbAs to a LU. The FFT presents assessment results at LO, LU and course level. This leads to the benefit that lecturers can analyze student performance at different aggregation levels (DP2) to evaluate if certain content was understood or needs to be repeated. Students can do the same individually for themselves and simultaneously compare their performance to the class average in order to disclose potential knowledge gaps. For the user roles lecturer and student, we incorporate walkthrough wizards (DP3) in the FFT. The FFT automatically starts a wizard when lecturers create a new LO. This wizard aims to introduce lecturers to the structure of FFT offering them a simple way to effectively use the tool (Rietsche et al. 2017). It guides the lecturers from the creation of the LO, through the assignment of CbAs, up to the creation of new or the modification of existing CbAs. These three visualized steps building on one another provide a simple way to use FFT and help lecturers to better understand and manage the relationship between LOs and CbAs. For students, the FFT's assessment process involves multiple steps. To maintain a clear structure, to allow for an intuitive use of FFT and to reduce the cognitive load of students, a wizard guides the students through the steps of the process. Specifically, the process consists of SA, CbA, feedback, social comparison, review and completion.

Formative feedback plus CbA can address learning objectives on various levels. Bloom's Taxonomy (Bloom et al. 1984) provides hierarchical models which allow the categorization of learning objectives according their complexity. In the cognitive domain, the Taxonomy defines six major categories of learning. These categories are ordered from easy to difficult while the expertise of each subjacent category is essential for the understanding of the next higher one (Krathwohl 2002). In the revision of Bloom's Taxonomy, Anderson and Krathwohl (2001) redefined these six categories of learning and show how they intersect with distinct levels of knowledge. Therefore, the revised Taxonomy is a two-dimensional framework. The two dimensions "Cognitive Process" and "Knowledge" and their intersections can be visualized within the Taxonomy Table (see section 1 in Figure 4). The Taxonomy Table allows the classification of assessments and the goal-oriented provision of feedback. Moreover, it reveals untapped potential in the way knowledge is transferred.

When creating a LO, lecturers use the Taxonomy Table (DF1) to define which objectives they want to achieve. For lecturers unfamiliar with the revised Taxonomy, FFT offers an assistant that determines the right placements in the table by posing a few cumulative questions. In the process of assigning CbAs to a LO, the lecturers then continuously see in the Taxonomy Table if the selected CbAs correspond with the defined objectives. If the user tries to create a LO without meeting all the objectives, a warning is being displayed. The highlighted area in the Taxonomy Table (see section 1 in Figure 4) represents the learning objectives the lecturer wants to achieve with the LO that is being created. The black check marks in the same table indicate which learning objectives are actually achieved with the assigned CbAs (see section 2 in Figure 4). If there are pursued objectives that remain uncovered by the assigned CbAs, FFT warns the lecturer by filling the affected cells with an orange cross. The classification of objectives, which a CbA is covering, depends to a certain degree on the question type included in the CbA. In the context of FFT, we use MCQs primarily to address the cognitive process dimensions "remember" and "understand" and, thus, to target lower-level learning (Attali and van der Kleij 2017).

In flaw diagrams, items of type "Substitution/Correction" (Scalise and Gifford 2006), students are asked to find and revise errors. According to Krathwohl (2002) these cognitive processes mainly take place within the dimensions "analyze" and "evaluate" and target higher-level learning. Those classifications are set as default when adding a new assessment. However, the specific classification within the Taxonomy Table can vary from CbA to CbA even if they are of the same type. Therefore, lecturers can adjust the default classification of individual CbAs based on the characteristics of the involved questions. In the FFT, feedback consisting of SA and CbA is based on the LUs which are covered throughout the semester. That way, students have enough time to use the formative feedback provided by FFT to improve themselves before the final exam takes place. SA (DF2), the first step of the assessment process, forces students to assess themselves in terms of their level of knowledge. One slider for each LO allows the students to define their current understanding of the corresponding learned content within a range from 0 to 100 percent. After the students assessed themselves, they conduct the CbAs (DF3). In the student view, all CbAs assigned to the LOs of a LU are presented to the user collectively. In other words, the FFT shows all the separately created CbAs within a LU in a single page to the students. The students then can choose their own way through the CbAs. The CbAs consist either of the assessment type (DF4) FD or MCQ. When creating a MCQ, lecturers enter a question and offer multiple response options for the students to choose from.

Figure 4. Taxonomy Table During the Process of Assigning CbAs to a Learning Outcome

Depending on the form of the question selected by the lecturers, either one or multiple response options can be correct. A FD basically consists of an image with incorporated errors. Students need to find these errors within a limited number of attempts and answer associated questions. In order to enable this, lecturers upload and crop flawed images and then mark the areas containing the flaws. To each of those flaws, the lecturer can add an MCQ or free text question.

Step 5: Evaluation of Alpha Version

The first evaluation of our artifact was carried out ex ante using an artificial evaluation setup (Venable et al. 2016). The purpose of this evaluation was to measure the completeness of the DPs and DFs. To reach this goal, we conducted a focus group workshop with four lecturers and thirteen students. In the workshop, the participants tested the user stories and created feedback reports based on the criteria a) logic, b) usability, c) usefulness and d) efficiency.

The evaluation of the workshop results led to revisions of the FFT's requirements. We investigated the click paths of the students. The results showed that the students often switched between multiple possible answers. Hence, a further requirement was to offer students the possibility to provide feedback in terms of how confident they were while answering the assessment questions. Further, the results showed that the usability needs to be improved. In the alpha version, the student performance was presented as a line diagram with a consistent line between the first and the last LU. This visualization was irritating. Additionally, the students and lecturers only received the results at LO level. The requirement emerged to revise the visualization of the results and show the student performance at different aggregation levels. During the workshop, we observed that students shared their results amongst each other in order to evaluate their performance. Based on the discussion with the students, a further requirement was requested. The students should be able to compare their performance with the performance of their fellow students. In the following section, the DFs derived from these requirements will be discussed in more detail.

Step 6: Deriving Design Features for the Beta Version

Based on the results of step five, the DFs of the artifact were revised and extended. In Figure 2, the DFs added for the beta version are highlighted in grey. When answering the items of a CbA, the FFT requests the students to select their confidence level (DF5). The levels consist of "I'm sure", "I'm not sure", "I guessed" and "I don't know". The monitoring of the students' confidence levels over time, may help students

to improve their performance (Nicol and Macfarlane-Dick 2006). Neither answering the questions within the CbAs nor stating the confidence level is mandatory for students. However, if students answer a question, a confidence level needs to be selected as well to continue the assessment process.

The results of the SA and CbAs are presented to students on an individual basis and to lecturers in aggregated form within student performance charts (DF6). FFT uses individual student results of the SA and CbAs aggregated either at LO or LU level to generate a bar chart in the students' overview. One bar represents the SA results and the other represents the aggregated CbA results expressed as a percentage of the maximum points. Consequently, the y-axis of the chart has a range from 0 to 100 percent and the x-axis displays, depending on the aggregation level, all LOs of a LU or all LUs of a course. While the FFT visualizes the individual results for students, it summarizes the results of all students of a course for lecturers. Lecturers may observe a chart showing the average SA and CbA results of their class expressed in percentage and aggregated at LO or LU level. However, lecturers do not have access to the results of individual students.

In the context of social comparison (DF7), students can compare their results with fellow students who resemble themselves in terms of competitiveness. More specifically, they can activate a third bar within the chart that reveals the average CbA score of students with the same performance goal. Festinger (1954) claimed, that people possess a drive to evaluate their opinions and abilities. There are two types of comparisons: upward and downward comparison. Upward comparison can be defined as the unidirectional drive upward, meaning that people prefer to compare themselves to others whose performance or abilities are slightly better. The definition of downward comparison, according to Wills (1981), is that people enhance their own subjective well-being by comparing themselves with less fortunate others. However, past research in the academic sector shows that upward comparison leads to better results than downward comparison. Blanton et al. (1999) observed in their experiment that students who compared themselves with well performing students increased their own academic performance the most during semester.

Evaluation of Beta Version

This section describes the naturalistic ex post evaluation of the beta version (Venable et al. 2016). We tested the beta version in a one-group field test to answer our second research question: Does the FFT usage increase student performance. Therefore, we wanted to investigate whether the students who used FFT more often performed better in the final exam. Before discussing the evaluation of beta version in detail, we will briefly present how the artifact is embedded in the didactic concept of our lecture.

The lecture was a mandatory course for all management information systems students. Each week the students had a traditional lecture called LU. Each LU consisted of a *pre-*, *during-* and *post-*phase. The artifact could have been used in each of these phases. In the *pre* phase, the artifact could have been used in a flipped classroom concept as pre entry assessment. Hence, the students would have to carry out the assessment before the class is taking place. The aim would have been to ensure that each student is well prepared for the discussion. In the *during* phase, the artifact could have been used in the last 15 minutes of the lecture. The aim would have been to deepen the newly learned materials. In the *post* phase, the artifact could have been used after the lecture took place. The aim would have been to deepen the learned materials to transfer the knowledge to the long-term memory. In our lecture, the artifact was used four times in the *during* phase and four times in the *post* phase. The *during* phase was 15 minutes at the end of the LU and the *post* phase was the time between the end of the LU and the begin of the next LU, usually 7 days.

Data Collection

A total of 189 students participated in the one-group field test, used the FFT at least one time, and participated in the final exam (dataset 1). To control for mastery approach and self-efficacy and measure perceived student performance the N decreased to 92 students (dataset 2). The N in dataset 2 decreased due to not all students filled out the pre and post questionnaire. Descriptive statistics for both datasets can be found in Table 1. For the participation in the study we raffled coupons for the merchandising shop of our university.

The design of the field test was as follows: Before the FFT was used the first time, the students carried out a pre-questionnaire to measure their mastery approach and self-efficacy. Each student had a personalized token to connect the data in FFT, exam and the survey questions.

Table 1. Descriptive Statistics (Every Student in Dataset 2 is Part of Dataset 1)				
	Dataset 1: Objective Student Performance		Dataset 2: Objective and Perceived Student Performance	
Sample Size	N=189		N=92	
Male	69.31 % (n=131)		73.91 % (n=68)	
Female	30.69 % (n=58)		26.09 % (n=24)	
	Mean	SD	Mean	SD
Age	25.90 years	1.92 years	25.92 years	1.65 years

In each LU the students followed a given four-step process. In the first step, the students carried out the SA, hence rating their current state of knowledge for each LO of the current LU. In the second step, the students started the CbA and answered the MCQs and flaw diagrams. In the third step the students received feedback about which questions were answered correctly and which were not, as well as how many points they received for their answer. In the fourth step, the feedback was presented to the students in the form of a bar chart. The students saw on the one hand the percentage of the SA and on the other hand the percentage of the CbA. After completing the last LU, the students carried out a post-questionnaire to measure their perceived student performance. In the final step, the students wrote the final exam of the lecture. The final exam consisted of retention (MCQs) and transfer tasks (e.g. applying learned methods) (Shute 2008).
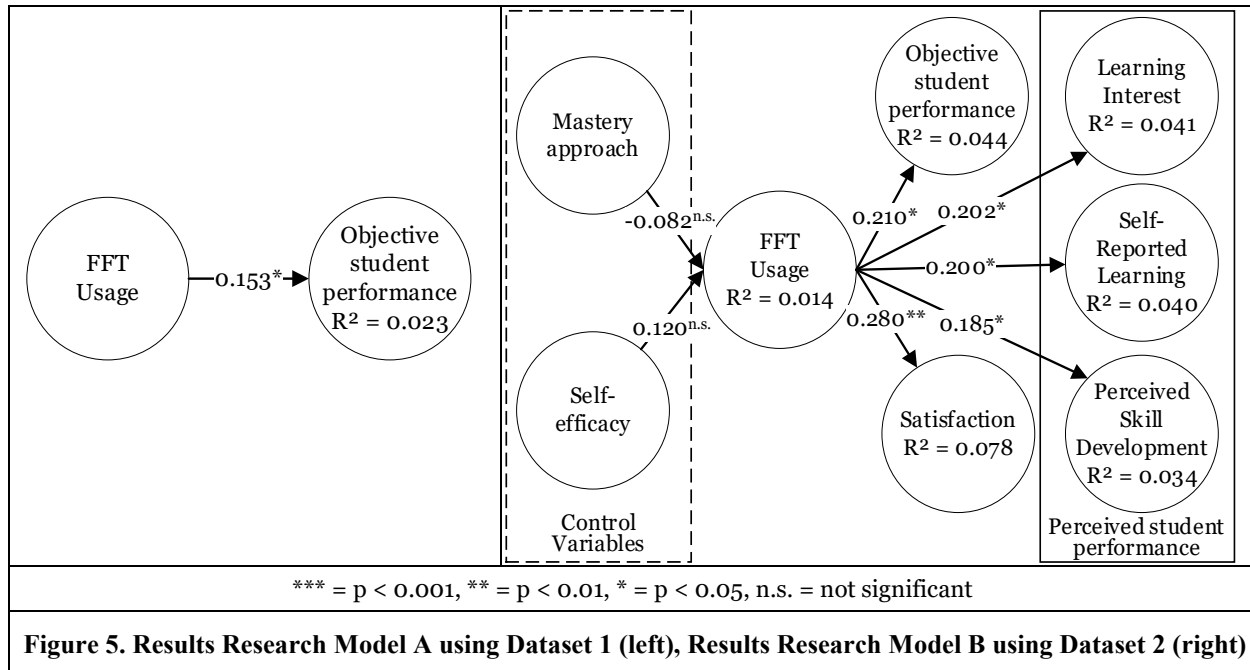
Data preparation

Before analyzing the datasets, we brought the survey responses into the right data structure. In R, we cleaned the data originating from the surveys by removing entries of students who did not complete the questionnaire. To prevent straight lining from affecting the survey results, we excluded students who selected the same response to different statements more than 20 times in a row (Millsap and Maydeu-Olivares 2009). Also, columns involving no student data at all were excluded from the data sets. Regarding the data originating from the FFT, we applied the boxplot method to identify and exclude outliers in terms of the usage duration at LU level. Thereby, every duration that exceeds 1.5 times the interquartile range (the middle 50% of the data) above the upper and below the lower quartile is considered an outlier (Backhaus et al. 2015). Outliers could occur, when students left the computer for hours while answering the questions or did not finish the assessment. After removing the outliers, we enriched FFT's data with the total usage duration and the average usage duration of each student.

Measurement of Constructs

To measure whether the *FFT usage* increases student performance, we describe the operationalization of the measurement constructs (see research model in Figure 5). The independent construct *FFT usage* can be measured according Straub et al. (1995) in different ways. We decided to measure this latent construct with two items. Firstly, frequency of use, which in our model is measured in terms of how many assessments the student has carried out (Ginzberg 1981). The second item is duration, which is in our case measured in terms of the timespan between starting with the first assessment task and finishing the last task (Srinivasan 1985). We decided to use these items based on the fact, that both are objective measurements and can be calculated based on database queries. The dependent construct *objective student performance* is measured as a manifest variable by evaluating how many points the students achieved in the final exam. The range is between 0 and 40 points. We decided to use points rather than grades, since points represent the student performance in higher precision.

One of the goals in education is the development of critical thinking and problem solving skills (Trilling and Fadel 2009). Hence, we argue that using the FFT should increase the students perceived skill development. In addition to the skill development, it is important that the students perceived that new content was learned. Hence, the learning of new factual knowledge and identifying concepts. Empirical studies found out, that there is a positive connection between student interest and student performance (Schiefele 1991). If students think about the topics, read further information or talk with fellow students, they strengthen the

Figure 5. Results Research Model A using Dataset 1 (left), Results Research Model B using Dataset 2 (right)

learned knowledge. Therefore, we argue that student performance not only consists of *objective student performance*, but rather of the combination of objective and perceived student performance (Söllner et al. 2017). To measure the dependent latent construct perceived student performance we used the widely known and often used measurements by Alavi (1994): *perceived skill development (PSD)*, *self-reported learning (SRL)* and *learning interest (LI)*. Example items are: *I did additional reading beyond the required assignments (PSD), I increased my ability to critically analyze issues (SRL)* and *I gained a good understanding of the basic concepts of the material (LI)*. According to Arbaugh (2001), satisfaction plays additionally an important role for student performance, such as if the students are satisfied using the FFT and would recommend it to fellow students (example item: *I am very satisfied with the course*). Additionally, we checked for the students' mastery approach (MA) and self-efficacy (SE). Example items are: *I enjoy challenging and difficult tasks at work where I'll learn new skills* (MA), *I am sure that I can solve the exercises and tasks of the course very well* (SE). To measure the students' MA we used the construct developed Baranik et al. (2007) and for SE the construct of Hardin et al. (2014).

We measured all latent construct variables with reflective indicators. For this purpose, we evaluated the measurement instrument with regards to its suitability to measure the constructs in a reflective manner. This was done by checking the reflective constructs according to the guidelines of Jarvis et al. (2003). We used a 7-point Likert response format that ranged from 1 ("strongly disagree") on the left to 7 ("strongly agree") on the right with 4 as a neutral point to assess the indicators.

Data Analysis

To analyze if the FFT usage increases student performance, we used a quantitative statistical method named structure equation model (SEM). This approach suits our research for three reasons (Backhaus et al. 2015). First, the nature of our research is rather confirmatory than exploratory. Second, the research model involves independent, dependent and control constructs and, thus, is rather complex. Third, the constructs of the model are latent, meaning they are not tangible or directly measurable.

As software for running the SEMs we used SmartPLS version 3.2.7 developed by Ringle et al. (2015). We used partial least squares (PLS) modeling instead of covariance-based modeling, because the objective of our research is not to evaluate a rigorous theoretical model, but rather to assess the effects of *FFTs usage* on the *objective* and *perceived student performance* (Mertens et al. 2017). Missing values in SmartPLS were handled by mean replacement. We ran the PLS algorithm with weighting scheme path and 300 iterations (anything else left default). The settings for the bootstrapping algorithm were 5000 subsamples

and individual sign changes (anything else left default). We measured all common quality criteria for the research model A and B (Backhaus et al. 2015). The factor loadings of all constructs exceeded 0.6. Cronbach´s alpha for all constructs is above 0.8, and the AVE is for all constructs is higher than 0.62. The composite reliability is for all constructs exceeds 0.83, and the VIF values are for all constructs is below 4.2. Hence, our research model fulfils all necessary quality criteria (Hair et al. 2014). Figure 5 shows the results achieved by running the PLS and Bootstrapping algorithms. We ran each algorithm for both datasets. Research model A with dataset 1 includes 189 students of the 194 students who wrote the final exam.

We included this dataset to show that students who used the FFT performed in general significantly better in the final exam. As stated in the theoretical background, to measure student performance holistically, objective and perceived student performance should be combined. Research model B shows that *FFT usage* has a significant effect on both, *objective* and perceived student performance consisting of *learning interest*, *self-reported learning,* and *perceived skill development* and *satisfaction*. We controlled for mastery approach and self-efficay of the students. Both constructs had no significant influence on the FFT usage.

We are aware that the R^2 values of the dependent variables are not high. This indicates that FFT usage alone does not explain a high magnitude of variation in our dependent variables. However, we want to stress that the goal of our evaluation was to assess whether FFT usage has a significant effect on important outcomes. In our case different measures of student performance as well as satisfaction.

Discussion

To communicate our theoretical contribution, we document our design knowledge according to the eight components of a design theory by Gregor and Jones (2007) (see Table 2).

This systematic approach helps to structure the design knowledge and learning of our project to make a first step towards developing a design theory for FFTs. Following Gregory and Muntermann (2014), we suggest that systematically developed design knowledge satisfies the criteria for scientific knowledge and provides a legitimate academic contribution.

The type of theory we are contributing to, is a "design and action" theory (Gregor and Jones 2007). The purpose of the theory is to provide prescriptions of principles of form and function for constructing artifacts. Hence, the description implies that following this recipe, the outcome is an instantiation of the class of FFTs (Gregor and Jones 2007). Furthermore, according to Walls et al. (1992), a design theory can never be purely explanatory. Hence, our design theory should also predict. With our field test we were able to show that following the recipe for the design of the artifact, we can predict that through the usage of the artifact, the a) objective student performance and b) perceived student performance significantly ($p < 0.05$) increase.

The purpose of a design theory (point 1, Table 2) is to give explicit prescriptions about the requirements for the development of FFTs. As outlined above, formative feedback building SA and CbA is one of the most influential factors for student performance. Through formative feedback, the students are able to adapt their learning process before the final exam. However, to the best of our knowledge, there is no study that has formalized design knowledge on FFT that combines SA and CbA. To close this gap, the goal of the first step is to contribute to a design theory, by explicitly formulating prescriptions for a class of FFTs that provide formative feedback using SA and CbA (see DSR process step one).

Constructs are representations of the entities of interest in a design theory (point 2, Table 2). In our project the constructs are represented by design features (DF). We derived four meta-requirements from scientific literature and nine requirements from expert interviews with lecturers and students. Based on the derived requirements we built seven DFs, which an instantiation of a FFT should have (Walls et al. 1992). The key DFs are, students receive feedback which is based on SA and CbA. Through the performance chart the students can actively monitor their performance at any time and any place. The performance of oneself can additionally be compared to fellow students for social comparison (see step two and three).

The principles of form and function are the essence of the design theory (point 3, Table 2). Since, we are designing a socio-technical system (Mumford 2006), one of the main principles is learner-centricity. To reach this goal we implemented three DPs with the purpose of enhancing usability and lowering cognitive load while using the FFT. The FFT is as web-based application with responsive UI to enhance usefulness (DP1). The FFT helps lecturers and students to interpret the assessment results by presenting them at different aggregation levels (DP2).

Table 2. Documentation of our Design Knowledge Adapted from Gregor and Jones (2007)	
1) Purpose and scope	The purpose of the FFT is to provide students with formative feedback during the semester concerning their current state of knowledge. The focus lies on higher education at campus university or distance learning providers such as MOOCs
2) Constructs	DF1) Taxonomy table, DF2) SA, DF3) CbA, DF4) Assessment type, DF5) Confidence level, DF6) Student performance charts, DF7) Social comparison
3) Principles of form and function	DP1) Usefulness through responsive UI, DP2) Reduce cognitive load through aggregation levels, DP3) Reduce cognitive load through walkthrough wizards
4) Artifact mutability	Core design features might be adapted to special contexts, e.g. a) content of the class and assessment types or b) social comparison based on the student culture
5) Testable propositions	TP1) *FFT usage* increases *objective student performance*, TP2) *FFT usage* increases *learning interest* (perceived student performance), TP3) *FFT usage* increases *self-reported learning* (perceived student performance), TP4) *FFT usage* increases *perceived skill development* (perceived student performance), TP5) *FFT usage* increases *satisfaction*
6) Justificatory knowledge	Scientific Literature on formative feedback (Hattie and Timperley 2007), SA (Boud 2013), CbA (Boud and Falchikov 2007), CDT (Festinger 1962)
7) Principles of implementation	PI1) LOs need to be linked to course content, PI2) Assessment items need to be linked to course content and LOs, PI3) Assessment items should allow students to carry out the assessment in a reasonable amount of time
8) Expository instantiation	We developed the FFT over the last 1.5 years and used it in large-scale lectures at our university

That way, the FFT enables lecturers and students to decide which content was understood and which needs to be repeated. Moreover, wizards guide lecturers and students through the front-end of the FFT in order to allow for an effective usage of the tool (DP3) (see step four and six).

Artifact mutability defines the possible variability, which may occur during the theory's instantiation (point 4, Table 2). While a design theory provides a general blueprint of a FFT design, the appearance of a corresponding concrete IT artifact is dependent on the context in which it will be instantiated. Especially, the instantiation of the theory´s core design features (e.g. assessment types and social comparisons) are highly depended on the context, e.g. content of the lecture or student culture. Specific content can only be assessed with a suitable assessment type. Social comparison might be lowering student performance in cultures with high power distances (Buholzer et al. 2018).

There are different ways to create testable propositions (TP) (point 5, Table 2). On the hand testing individual design features and on the other hand testing the effects of the tool usage in general (Gregor and Jones 2007). We decided in favor of the latter since we first want to find out if the FFT design works in general. Therefore, we created five TPs: The *FFT usage* increases *objective student performance* (TP1), *learning interest* (perceived student performance) (TP2), *self-reported learning* (perceived student performance) (TP3), *perceived skill development* (perceived student performance) (TP4) and *satisfaction* (TP4) (see step seven).

Justificatory knowledge represents insights from literature that inform, explain and validate our design decisions (point 6, Table 2). Feedback is an important aspect to improve student performance (Hattie and Timperley 2007; Kluger and DeNisi 1996; Price et al. 2010; Shute 2008). In the meta-analysis of Hattie (2015) the effect sizes of feedback and SA in higher education belong to the top 15 effect sizes. Moreover, previous research shows that SA is able to enhance student performance (Nicol and Macfarlane-Dick 2006). McDonald and Boud (2003) have shown that training students how to carry out SA can improve their performance in final exams. Furthermore, the CDT developed by Festinger (1962) provides a theoretical lens on the psychological coherences when a cognitive dissonance occurs by showing SA, CbA and optionally social comparison results to students (see Theoretical Background).

We suggest three principles of implementation (PI) (point 7, Table 2). First, LOs need to be closely related to the course content. LOs are on a continuum between very generic to very specific. Hence, the lecturer needs to define LOs which are balanced between generic and specific LOs (PI 1). Second, the assessment items need to reflect the content of the lecture and need to be linked to the LOs. The assessment items are on a continuum between very easy and very difficult. The lecturer needs to find a mix between easy and difficult items (PI 2). Third, the number of items per assessment should allow students to carry out the assessment in a reasonable amount of time (PI 3).

Expository instantiation (point 8, Table 2): We implemented the artifact over the last 1.5 years and used it in the real world setting of our university (see step four and six, Figure 3).

Conclusion

The aim of the presented research has been to a) examine what requirements should be considered when designing an IT tool that aims at providing formative feedback in large-scale lectures and b) to what extend the IT tool usage increases student performance. To design our artifact, we used the DSR approach. We derived three design principles and seven design features from the meta-requirements of the scientific literature and requirements from expert interviews with lectures and students. The results of our one-group field test have shown that for the whole class, we can show that the FFT usage has a significant influence on the objective student performance. Furthermore, for a subset of the class, who filled out the pre-/posttest, we can show that the FFT usage significantly increases objective and perceived student performance. Hence, students using our tool received more points in the final exam and perceived higher learning interest, skill development and understanding regarding the concepts and content discussed in lecture. By documenting our design knowledge, we made a first step in contributing to a design theory for FFTs.

A limitation of this study is that we tested the requirements for one specific context: a large-scale management information systems masters university class. Furthermore, this study was limited by the absence of a real experiment setting with a control group.

For future research, we suggest two points. First, a further experiment with e.g. eye tracking to investigate each design principle and feature individually. Hence, which of the principles and features have the strongest effect to the students' performance. Second, based on the CDT, our next step is to investigate how the SA and CbA results change over time when students using the FFT. Furthermore, we want to investigate the long-term effects of a high or low cognitive dissonance on student performance.

References

Alavi, M. 1994. "Computer-Mediated Collaborative Learning: An Empirical Evaluation," *MIS quarterly* (18:2), pp. 159–174.

Anderson, L. W., and Krathwohl, D. R. (eds.) 2001. *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*, New York: Longman.

Arbaugh, J. B. 2001. "How instructor immediacy behaviors affect student satisfaction and learning in web-based courses," *Business Communication Quarterly* (64:4), pp. 42–54.

Attali, Y., and van der Kleij, F. 2017. "Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving," *Computers & Education* .

Backhaus, K., Erichson, B., and Weiber, R. (eds.) 2015. *Fortgeschrittene Multivariate Analysemethoden*: *Eine anwendungsorientierte Einführung*, Berlin, Heidelberg: Springer Gabler.

Baranik, L. E., Barron, K. E., and Finney, S. J. 2007. "Measuring goal orientation in a work domain: Construct validity evidence for the 2× 2 framework," *Educational and Psychological Measurement* .

Bennett, N., and Lemoine, J. 2014. "What VUCA really means for you," .

Bitzer, P., Söllner, M., and Leimeister, J. M. 2016. "Design principles for high-performance blended learning services delivery," *Business & Information Systems Engineering* (58:2), pp. 135–149.

Black, P., and Wiliam, D. 2009. "Developing the theory of formative assessment," *Educational Assessment, Evaluation and Accountability(formerly: Journal of Personnel Evaluation in Education)* (21:1), p. 5.

Blanton, H., Buunk, B. P., Gibbons, F. X., and Kuyper, H. 1999. "When Better-Than-Others Compare Upward: Choice of Comparison and Comparative Evaluation as Independent Predictors of Academic Performance," *Journal of personality and social psychology* (76:3), pp. 420–430.

Bloom, B. S., Krathwohl, and Masia, B. B. 1984. *Bloom taxonomy of educational objectives:* Allyn and Bacon, Boston, MA. Copyright (c) by Pearson Education.

Boud, D. 1991. *Implementing student self-assessment:* HERDSA Green Guide, 2nd edn (Sydney, Higher Education Research and Development Society of Australia).

Boud, D. 2013. *Enhancing learning through self-assessment:* Routledge.

Boud, D., and Falchikov, N. 2007. *Rethinking assessment in higher education: Learning for the longer term:* Routledge.

Brown, G. T. L., Peterson, E. R., and Yao, E. S. 2016. "Student conceptions of feedback: Impact on self-regulation, self-efficacy, and academic achievement," *British Journal of Educational Psychology* (86:4), pp. 606–629.

Buholzer, F., Rietsche, R., and Söllner, M. 2018. "Knowing what Learners Like ? Developing a Cultural Sensitive Peer Assessment Process in MOOCs," *Multikonferenz Wirtschaftsinformatik (MKWI). Lüneburg, Germany* .

Chemers, M. M., Hu, L.-t., and Garcia, B. F. 2001. "Academic self-efficacy and first year college student performance and adjustment," *Journal of educational psychology* (93:1), p. 55.

Elliot, A. J., and Devine, P. G. 1994. "On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort," *Journal of personality and social psychology* (67:3), p. 382.

Fadel, C., Bialik, M., and Trilling, B. 2015. *Four-dimensional education*: *The competencies learners need to succeed.*

Festinger, L. 1954. "A Theory of Social Comparison Processes," *Human Relations* (7:2), pp. 117–140.

Festinger, L. 1962. *A theory of cognitive dissonance:* Stanford University Press.

Fortes, P. C., and Tchantchane, A. 2010. "Dealing with Large Classes: A Real Challenge," *Procedia - Social and Behavioral Sciences* (8), pp. 272–280.

Gibbs, G., and Simpson, C. 2005. "Conditions under which assessment supports students' learning," *Learning and teaching in higher education* (1), pp. 3–31.

Ginzberg, M. J. 1981. "Early diagnosis of MIS implementation failure: promising results and unanswered questions," *Management Science* (27:4), pp. 459–478.

Gregor, S., and Hevner, A. R. 2013. "Positioning and presenting design science research for maximum impact," *MIS quarterly* (37:2), pp. 337–355.

Gregor, S., and Jones, D. 2007. "The Anatomy of a Design Theory," *Journal of the Association for Information Systems* (8:5), 312.

Gregory, R. W., and Muntermann, J. 2014. "Research Note—Heuristic Theorizing: Proactively Generating Design Theories," *Information Systems Research* (25:3), pp. 639–653.

Gupta, S., Bostrom, R. P., and Huber, M. 2010. "End-user training methods," *ACM SIGMIS Database* (41:4), p. 9.

Hair, J. F., Hult, T. M., Ringle, C. M., and Sarstedt, M. 2014. *A primer on partial least squares structural equation modeling (PLS-SEM)*, Thousand Oaks, CA, USA: Sage.

Hardin, A. M., Looney, C. A., and Fuller, M. A. 2014. "Self-efficacy, learning method appropriation and software skills acquisition in learner-controlled CSSTS environments," *Information Systems Journal* (24:1), pp. 3–27.

Hattie, J. 2015. "The applicability of Visible Learning to higher education," *Scholarship of Teaching and Learning in Psychology* (1:1), pp. 79–91.

Hattie, J., and Timperley, H. 2007. "The power of feedback," *Review of Educational Research* (77:1), pp. 81–112.

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

Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 2003. "A critical review of construct indicators and measurement model misspecification in marketing and consumer research," *Journal of consumer research* (30:2), pp. 199–218.

Kluger, A. N., and DeNisi, A. 1996. "The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory," *Psychological Bulletin* (119:2), p. 254.

Krathwohl, D. R. 2002. "A revision of Bloom's taxonomy: An overview," *Theory into practice* (41:4), pp. 212–218.

Kruger, J., and Dunning, D. 1999. "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments," *Journal of personality and social psychology* (77:6), p. 1121.

Marjanovic, O. 2016. "Designing Innovative Education through Action Design Research: Method and Application for Teaching Design Activities in Large Lecture Environments," *JITTA: Journal of Information Technology Theory and Application* (17:2), p. 22.

Mason, B. J., and Bruning, R. 2001. "Providing Feedback in Computer-based Instruction: What the Research Tells Us," (15).

McDonald, B., and Boud, D. 2003. "The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations," *Assessment in Education: Principles, Policy & Practice* (10:2), pp. 209–220.

Mertens, W., Pugliese, A., and Recker, J. 2017. *Quantitative data analysis*: *A companion for accounting and information systems research*.

Millsap, R. E., and Maydeu-Olivares, A. 2009. *The SAGE handbook of quantitative methods in psychology:* Sage Publications.

Mumford, E. 2006. "The story of socio-technical design: Reflections on its successes, failures and potential," *Information Systems Journal* (16:4), pp. 317–342.

Nicol, D. J., and Macfarlane-Dick, D. 2006. "Formative assessment and self-regulated learning: a model and seven principles of good feedback practice," *Studies in Higher Education* (31:2), pp. 199–218.

OECD 2016. *Education at a Glance 2016*: *OECD Indicators*, Paris: OECD.

Piaget, J. 1985. *The equilibration of cognitive structures : the central problem of intellectual development*, Chicago: University of Chicago press.

Price, M., Handley, K., Millar, J., and O'Donovan, B. 2010. "Feedback: All that effort, but what is the effect?" *Assessment & Evaluation in Higher Education* (35:3), pp. 277–289.

Rietsche, R., Söllner, M., and Seufert, S. 2017. "Digital formative learning assessment tool - Towards helping students to take ownership of their learning," in *European Conference on Information Systems (ECIS),* Guimarães, Portugal.

Riley, R. 2004.

Ringle, C. M., Wende, S., and Becker, J.-M. 2015. *SmartPLS 3:* SmartPLS GmbH.

Sadler, D. R. 1989. "Formative assessment and the design of instructional systems," *Instructional science* (18:2), pp. 119–144.

Scalise, K., and Gifford, B. 2006. "Computer-based assessment in e-learning: A framework for constructing" intermediate constraint" questions and tasks for technology platforms," *The Journal of Technology, Learning and Assessment* (4:6).

Schiefele, U. 1991. "Interest, Learning, and Motivation," *Educational Psychologist* (26:3-4), pp. 299–323.

Shute, V. J. 2008. "Focus on formative feedback," *Review of Educational Research* (78:1), pp. 153–189.

Söllner, M., Bitzer, P., Janson, A., and Leimeister, J. M. 2017. "Process is king: Evaluating the performance of technology-mediated learning in vocational software training," *Journal of Information Technology* (33:3), pp. 233–253.

Srinivasan, A. 1985. "Alternative measures of system effectiveness: associations and implications," *MIS quarterly* , pp. 243–253.

Straub, D., Limayem, M., and Karahanna-Evaristo, E. 1995. "Measuring system usage: Implications for IS theory testing," *Management Science* (41:8), pp. 1328–1342.

Thelwall, M. 2000. "Computer-based assessment: a versatile educational tool," *Computers & Education* (34:1), pp. 37–49.

Trilling, B., and Fadel, C. 2009. *21st century skills*: *Learning for life in our times:* John Wiley & Sons.

van der Kleij, F. M., Eggen, T. J.H.M., Timmers, C. F., and Veldkamp, B. P. 2012. "Effects of feedback in a computer-based assessment for learning," *Computers & Education* (58:1), pp. 263–272.

Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: a framework for evaluation in design science research," *European Journal of Information Systems* (25:1), pp. 77–89.

Walls, J. G., Widmeyer, G. R., and El Sawy, O. A. 1992. "Building an information system design theory for vigilant EIS," *Information Systems Research* (3:1), pp. 36–59.

Wills, T. A. 1981. "Downward comparison principles in social psychology," *Psychological Bulletin* (90:2), pp. 245–271.

Yorke, M. 2003. "Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice," *Higher Education* (45:4), pp. 477–501.