5 Towards a Holistic Evaluation Concept for Personalised Learning in Flipped Classrooms

Abstract

Incorporating the student's preferences regarding pace, methods, and contents into teaching is particularly hard in today's higher education, providing courses to large numbers of students often over electronic media. Such personalised learning can be implemented via self-regulated learning approaches using the method of the flipped classroom. However, literature on the design and evaluation of such courses is scarce. Evaluation models and instruments are not adapted to the specific nature of the flipped classroom yet, combining presence and online teaching. The present paper aims at conceptualising a holistic approach towards an evaluation concept for personalised learning. Based on an overview of evaluation models in the learning sciences and ISs domains an evaluation concept is presented and applied to a course instantiation focusing on the topics of (1) fulfilment of general requirements and effects on (2) learning outcomes, (3) adoption, and (4) individual factors of the students.

Co-Author
Prof. Mareike Schoop, PhD

5.1   Evaluating Modern Teaching and Learning

Universities have long held an unrivalled position in delivering higher education. Traditionally, lectures, example classes, tutorials, laboratories, and other forms of teaching were the formats of choice. What they all have in common is that they are lecturer-centred in that the lecturer directs the learning process, the forms of interaction (if any), the teaching method(s), and the learning directions and is thus the focal point of such courses.

Nowadays, it is an acknowledged fact that students have different needs and approaches of acquiring knowledge. Heterogeneous groups of students exhibit a large variety of individual factors (e.g. educational background, cultural background, personality traits, skills, and abilities), which require personalised teaching and learning. Personalised learning aims to incorporate individual preferences into the learning process regarding pace, methods, and contents (U.S. Department of Education 2010). This is not a new idea as it has been done by lecturers and students on a daily basis. However, especially in large classes and online learning environments, personal preferences of students are difficult for the lecturer to consider. New approaches towards personalised learning, therefore, follow the constructionist learning paradigm (Kafai 2006), putting the students in charge of their own learning process. One method to do so is self-regulated learning, which emphasises the students' metacognitive abilities (i.e. knowing about one's own learning). Self-regulated learning aims to empower students to plan, implement, and reflect their learning for continuous improvement. Thereby, students should be enabled to derive learning strategies matching their individual preferences overcoming previous limitations (Zimmerman 2002).

E-learning supports the ideas of personalised learning, providing learning materials at any place and any time. Especially blended learning as the combination of presence and electronic learning opens up new avenues of learning (Garrison and Vaughan 2011). One method to implement blended learning in a meaningful way is the flipped classroom. It combines not only presence and electronic learning but also self-regulated and lecturer-moderated (rather than lecturer-centred) learning. Flipped classrooms turn around traditional lecture and tutorial teaching, requiring a distant preparation phase performed by the students themselves, consequently enabling the lecturer to discuss and apply the acquired knowledge within the lecture (Lage et al. 2000). The distant preparation phase, therefore, relies on the integration of e-learning technology (Strayer 2012).

Whilst the idea of the flipped classroom seems to be very promising for improving active, collaborative, and self-regulated learning, scientific dissemination on concrete course design, evaluation, and learning outcomes is still scarce (McNally et al. 2017; Abeysekera and Dawson 2014; Butt 2014; Bishop and Verleger 2013; Pierce, Fox 2012). This is problematic because there are various ways to implement a flipped classroom course design depending on its topic, theory-focus, assessment-focus, full-flip or partial-flip, etc., which require a generalisable approach to evaluation (Bishop and Verleger 2013). Furthermore, the students have different preferences and requirements towards a flipped classroom, which must be incorporated (McNally et al. 2017). Evaluation instruments focus either on learning or on the use of e-learning tools, often neglecting their combination. Finally, learning interventions exert a plethora of psychological, social and technological effects regarding the students, lecturers, and institutions all being related to each other.

Therefore, the research goal of this paper is to develop an evaluation concept for personalised flipped classrooms and apply it theoretically to an example course to show its feasibility. The paper thus follows a design-oriented approach comprising of a build and evaluate cycle on the evaluation concept itself as an artefact (Hevner et al. 2004). Developing an evaluation concept, we specifically focus on scientific evaluation aiming towards understanding whether and how flipped classrooms work and what effects they exert, not teaching evaluation, which is conducted due to university quality assurance.

The paper is structured as follows. Section 5.2 presents the chosen design-oriented research methodology. The overview of existing evaluation models and instruments for learning interventions regarding their main constructs self-regulated learning, learning outcomes, adoption, and individual factors is explained in section 5.3. Section 5.4 describes the conceptual background of the personalised flipped classroom and its implementation in a real university course to which the evaluation concept will be applied to. Section 5.5 presents the resulting evaluation concept discussing methodological aspects as well as proposing a combination of the frameworks and instruments presented in section 5.3. Finally, section 5.6 summarises and discusses the presented evaluation concept showing future research directions.

5.2   Methodology

This study is part of a larger research project to analyse the effects of personalised learning in a holistic manner, following a design-oriented research methodology. Such a pragmatist approach is prevalent in ISs (Hevner et al. 2004) as well as in the learning sciences (Brown 1992; Collins 1992) striving to create knowledge by designing solutions to practical problems. The designed artefacts resemble complete real-life learning interventions or software, which are applied to their intended context. Thus, evaluation is not restricted to artificial scenarios but happens within a natural environment providing rich insights and continuous improvement (Anderson and Shattuck 2012). However, a design-oriented methodology is not restricted to solving a specific problem. In contrast, it aims to generalise findings beginning with instantiations of courses or software to mid-range theories eventually creating grand theories (Gregor and Hevner 2013).

In our previous work a personalised flipped classroom has been designed and implemented over a complete semester at a German university (Melzer and Schoop 2017a). The present paper aims to develop a sound evaluation concept for this course. However, the target of this evaluation is not only the instantiated course, but also its underlying PLF (Melzer and Schoop 2015) and the general requirements derived from it.

5.3   An Overview of Models and Instruments for the Evaluation of Personalised Learning

Bishop and Verleger (2013) distinguish three dimensions regarding the evaluation of flipped classrooms:

1) lecturer or student as object of analysis,
2) objective or subjective analysis, and
3) time and quantity of analyses.

To evaluate learning interventions from a student perspective investigating student engagement, Fredricks and McColskey (2012) name several measurement methods such as student self-report surveys, lecturer ratings of students, interviews, and observations. An evaluation of a learning intervention only from a student's perspective, however, would be incomplete as students cannot estimate, for example, the achievement of learning goals planned by a lecturer. Moreover, one could argue that the

institution in which the learning takes place also affects the learning and therefore could be a viable object of analysis.

While objective measurement focuses on grades and scores in various forms (Findlay-Thompson and Mombourquette 2014), subjective measurement gathers perceptions and opinions from the participants. Objective measurement in the learning sciences is usually differentiated into formative and summative measurement. Formative scores encompass for example self-control tests employed before or during lectures to test, whether students are able to follow the course. Hence, summative scores are grades or points achieved in the final exam conducted after the semester. While the literature on flipped classrooms requests formative assessment to motivate the students to prepare as well as provide a constant measure of retention (Lehmann et al. 2015; Bishop and Verleger 2013; Milman 2012), McNally et al. (2017) specifically reports on the importance of summative measurements.

Finally, there are several research designs including matched or unmatched pre-post-test designs requiring a differing number of surveys or tests. Looking at the literature, it can be stated that there are very few approaches that actually implemented and reported on flipped classroom design and evaluation. The majority of the 24 studies found by Bishop and Verleger (2013) measures subjective student perceptions whilst only two investigate objective student performance (Day and Foley 2006; Moravec et al. 2010). Only one of the studies evaluates the flipped classroom over a complete semester (Day and Foley 2006). Furthermore, Bishop and Verleger (2013) find that only half of the studies employ matched pre-post-tests whereas the other half employs post-tests only. In the following, we will therefore present theories and instruments which have been used in the domains of learning sciences and ISs to evaluate personalised learning interventions and are adaptable to flipped classrooms providing a holistic approach to evaluation.

5.3.1 Self-Regulated Learning

Self-regulated learning is seen as a way for students to address individual factors during the learning process (Zimmerman 2002). The student's awareness and knowledge about the learning process itself is seen as the key factor, which must be implemented appropriately. It includes tasks such as setting learning goals, deriving learning strategies, monitoring

learning performance, restructuring physical or social context, time management, self-evaluation as well as understanding its results, and finally adapting the learning accordingly. Self-regulated learning is defined to be proactive and therefore matches the core idea of the flipped classroom. Furthermore, self-efficacy and self-motivation have been found to be important predictors of learning performance (Zimmerman 2002).

A comprehensive instrument which can be used to investigate self-regulated learning is the Motivated Strategies for Learning Survey (MSLQ) analysing a student's motivation as well as specific learning strategies as its main constructs (Duncan and McKeachie 2005). It is cited over 650 times (Google Scholar 2017) and generally reported to achieve valid results (Fredricks and McColskey 2012). The construct of learning motivation comprises of several factors, namely intrinsic and extrinsic goal orientation, task value, learning control, self-efficacy, and test anxiety. Learning strategies represent general skills such as organisation, metacognition, time management, as well as specific strategies such as rehearsal, critical thinking, peer learning, and help seeking. While the MSLQ assesses self-regulated learning in traditional learning interventions, Liaw and Huang (2013) analyse predictors of self-regulated learning in e-learning environments. Although they state a comprehensive model of self-regulated learning in e-learning environments as a remaining research challenge, they show that self-regulation depends on the interactive learning environment, satisfaction, and usefulness of the environment and individual factors such as anxiety and self-efficacy. Nevertheless, self-regulated learning usually focuses on constructs from the learning sciences, ignoring an ISs perspective.

5.3.2   Learning Outcomes

Learning Outcomes are best reflected by analysing formative or summative student grades or scores as objective measures focusing on a learning intervention as a whole. As a subjective measure, perceived quality of the teaching can be used to show the success of a learning intervention.

Therefore, the Information Systems Success Model (ISSM; DeLone and McLean 1992; DeLone and McLean 2003) has been adapted to the domain of e-learning measuring information quality (i.e. accuracy, completeness, ease of understanding, and relevance of the materials), system quality (i.e. availability, ease of use, reliability, and response time), and service quality (i.e. overall support) of online courses (Chiu et al. 2007).

These constructs reflect antecedent factors of IS success, comprising of intention to use, actual use, user satisfaction, and net benefits. The ISSM, adapted to e-learning, however, is one of the very few instruments combining constructs focusing on learning itself as well as technological factors.

5.3.3   Adoption

Adoption of ISs is widely investigated using the Technology Acceptance Model (TAM; Venkatesh and Bala 2008), which poses perceived usefulness and perceived ease of use as key predictors influencing the behavioural intention to use a software as well as the usage behaviour.

TAM has already been adopted in the research area of e-learning showing the importance of perceived usefulness and ease of use in this domain. Furthermore, highlighting the importance of multimedia system functionalities, system interactivity, critical mass, computer self-efficacy, subjective norm, and content quality as additional predictors of system use (Gross et al. 2016; Sung Youl Park 2009; Pituch and Lee 2006; Lee 2006; Selim 2003). Furthermore, several studies have shown that actual usage of an e-learning tool and user satisfaction are related to each other (Sun et al. 2008; Liaw and Huang 2013).

5.3.4   Individual Factors

Individual factors are also investigated within the learning sciences as well as ISs. Such factors can be related to demographic variables such as gender, age, or job status (Lu and Chiou 2010), personal context, or institutional factors (Melzer, Schoop 2015). Self-regulated learning, for example, relies very much on the personality traits of task anxiety, self-efficacy, and learning motivation (Duncan and McKeachie 2005).

Personality traits also reflect individual factors, which have been thoroughly investigated in the domain of personalised learning as learning styles (Coffield et al. 2004) and ISs as cognitive styles (Robey and Taggart 1981). Both streams of research, however, provide inconclusive findings (Scott 2010 and Huber 1983). Thus, we acknowledge weak effects of individual factors, being relevant to categorise the types of students and their reaction towards flipped classrooms. Individual factors have been found to affect different e-learning constructs such as learning outcomes (Melzer

and Schoop 2016) and satisfaction (Lu and Chiou 2010). Furthermore, individual preferences have been used as categorisation measures grouping students in a flipped classroom into flip endorsers and flip resisters explaining their behaviour (McNally et al. 2017).

5.4   A Personalised Flipped Classroom University Course

The following section explains the theoretical underpinnings of a personalised flipped classroom in the form of general requirements. Moreover, a concrete university course which serves as the basis for our evaluation concept is described.

5.4.1   The Personalised Learning Framework

Following the notion of the PLF (Melzer and Schoop 2015), we understand personalised learning as an inherently active and collaborative process. The PLF combines the COI framework (Garrison 2011) with the theory of Cognitive Fit (Vessey 1991) to explain personalised learning in blended learning environments (cf. Figure 19). The COI framework conceptualises a social, cognitive, and teaching presence. While social presence requires creating and maintaining cohesive learning groups, cognitive presence models the individual learning following the process of practical inquiry. Teaching presence eventually models design and implementation of classes, scaffolding, and facilitating discourse between students. The COI is able to personalise learning in electronic contexts following the idea of Cognitive Fit selecting learning tasks and tools according to their personal or group preferences. This, however, requires the availability of alternative tasks and tools to achieve the desired learning goals. While learning tasks can be defined according to Bloom's Taxonomy of Learning (Bloom et al. 1984; Anderson and Krathwohl 2001) into lower (remember, understand, apply) and higher (analyse, evaluate, create) order thinking skills, learning tools are defined using the notion of PLEs. PLEs are defined as a wide variety of electronic tools that facilitate learning including operating systems, office applications, as well as social media tools (Attwell 2007). The

process of personalisation is affected by several contextual and institutional moderators, namely the university's strategy, infrastructure, culture, and the student's personal goals and context.
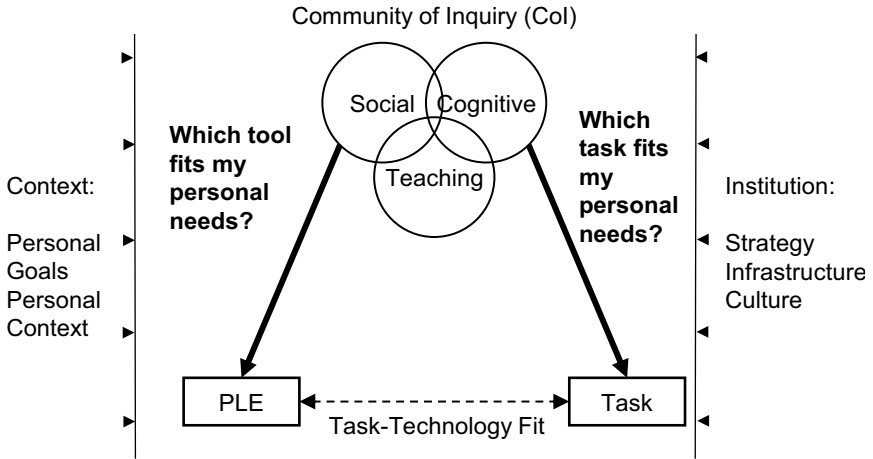

Figure 19   Personalised Learning Framework (Melzer and Schoop 2015, p. 7)

From the PLF, Melzer and Schoop (2015) derive several general requirements for personalised flipped classrooms. Firstly, personalisation must be provided. Personalisation concerns learning tasks (e.g. exercises) and tools (e.g. websites, social media tools) and communication facilities. Consequently, the lecturer must provide freedom and guidance for personalisation on a central learning platform supporting the students with reasonable IT infrastructure and support (Melzer and Schoop 2015). A VLE can be used as a central platform with links to other websites, tools, or services similar to a PLE. Establishing social presence requires open communication between participants and encouraging collaboration while cognitive presence is implemented by the model of practical inquiry facilitating an exploration-based approach to learning. Teaching presence requires the lecturer to design and organise the course, facilitate discourse, and provide direct instruction where necessary (Garrison and Arbaugh 2007).

5.4.2   From a Traditional Lecture to a Personalised Flipped Classroom

The course to be transformed into a personalised flipped classroom is associated to a business and ISs curriculum of several master programmes including about 120 to 150 students per year at a German university. The course is recommended for attendance in the first semester and comprises weekly lectures, five negotiation journal entries, and a final exam. The final grade comes from the exam result (50%) and the grades on the journal entries (50%). The course is taught in English focusing on planning, conducting, and evaluating negotiations in business contexts using traditional face-to-face as well as electronic negotiation media. The lecture aims to provide knowledge from an ISs perspective on the topics of negotiation basics, negotiation planning, communication aspects, decision and negotiation analytics, electronic negotiations, dispute resolution, and culture in negotiations. It leaves plenty of time for applying this knowledge in practical tasks such as discussions, role-plays, and case studies. Further application and reflection tasks are performed in the negotiation journal, where students have to negotiate in real-life with other people, analyse negotiations as well as assess their own negotiation behaviour.

For the winter term of 2016, this course has been transformed into a personalised flipped classroom (Melzer and Schoop 2017a). Combining and adapting existing approaches from self-regulated learning (Zimmerman 2002) and flipped classrooms (Oeste et al. 2014; Bishop and Verleger 2013), the course is organised in three phases:

1) preparation (i.e. self-regulated preparation of theoretical knowledge in groups performing personalisable learning tasks over personalisable learning tools);
2) lecture (i.e. the lecturer focuses on student-centred discussions and guides applications of the previously learned knowledge to broaden and deepen the knowledge acquisition);
3) reflection (i.e. individual reflection on preparation and lecture performing sophisticated learning tasks writing the negotiation journal over an online course).

This process model is implemented using the VLE ILIAS (ILIAS e.V. 2016) as a central platform for preparation, communication, materials, and reflection. Personalisation is thus implemented providing alternative learning

tasks and tools to achieve learning goals within the preparation (e.g. read a paper vs. watch a video to understand a topic) as well as within lectures (e.g. discuss vs. perform a role-play to apply a topic) and reflection (e.g. through broad task descriptions and multimedia support facilitating reflection). Students are encouraged to prepare in groups according to their individual preferences and synthesise the acquired knowledge before respectively in the lecture.

5.5   An Evaluation Concept for Personalised Learning in Flipped Classrooms

The following section presents the developed evaluation concept for personalised learning. Starting with its methodological underpinnings, the previously presented evaluation models are arranged to fit personalised learning in flipped classrooms.

5.5.1   *Methodology*

The main goal of the present work is to design a general evaluation concept for personalised learning, which will be applied to the personalised flipped classroom described in section 5.4. Its first step should thus be to check whether the general requirements derived from the PLF have been implemented correctly and are noticed by the students. Furthermore, learning outcomes, adoption, and individual factors have to be assessed. Figure 20 displays relevant constructs as well as their relationships. These constructs will be assessed using a mixed-method approach as suggested for the holistic evaluation of learning interventions by the design-oriented methodology (Anderson and Shattuck 2012) as well as previous studies on flipped classroom evaluation (Fredricks and McColskey 2012; McNally et al. 2017).
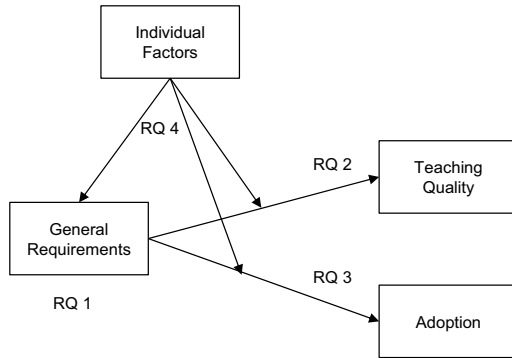
Figure 20   Underlying Constructs for the Evaluation of Personalised Flipped Classrooms

Figure 21 summarises the presented evaluation concept incorporating qualitative and quantitative analysis of subjective perceptions and quantitative analysis of objective performance of the students and the corresponding measures. The quantitative survey on subjective student perceptions and objective performance is complemented by qualitative data gathered using journal entries, interviews, and observations. Therefore, rich insights can be obtained refining and extending the previously identified constructs (Venkatesh et al. 2013).
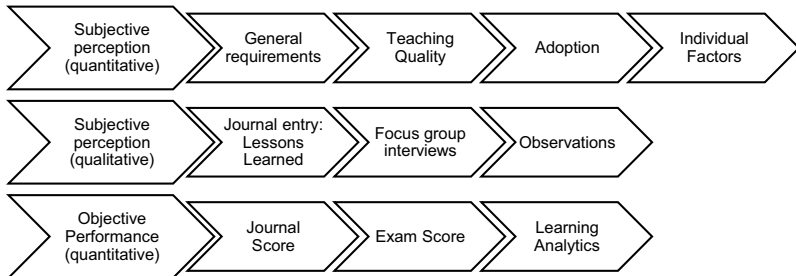

Figure 21   Subjective and Objective Evaluation Measures

5.5.2   Application of Measures

Subjective perception regarding our constructs of interest is measured using a post-test survey. First of all, it investigates whether the general requirements regarding the flipped classroom course have been fulfilled,

namely personalisation of tasks and tools, inclusion of institutional and contextual factors as well as implementation of the COI framework involving social, cognitive, and teaching presence. Regarding the requirements of personalisation, new survey items have to be created. Additionally, the MSLQ instrument is used to assess, whether self-regulated learning occurs during the course as a precursor for personalisation at the same time assessing learning strategies (Duncan and McKeachie 2005). To investigate the social, cognitive, and teaching presences defined in the COI framework, we employ the corresponding pre-validated COI instrument (Arbaugh et al. 2008). This instrument has been used in several studies, which evaluated online and blended learning courses employing the COI framework (Lambert and Fisher 2013; Shea and Bidjerano 2010) and is able to indicate the quality of the social, cognitive, and teaching presences as well as their relationships.

Secondly, learning outcomes are evaluated following the ISSM (DeLone and McLean 2003) adapted to the e-learning context focusing on information quality, system quality, service quality (Chiu et al. 2007) and satisfaction (Liaw and Huang 2013).

Adoption – especially of the course's online parts – is investigated using the TAM constructs perceived ease of use, perceived usefulness, perceived self-efficacy, and perceived anxiety (Venkatesh and Bala 2008). These constructs – together with perceived satisfaction and interactive learning – are of particular relevance since they have been found to be predictors for the level of self-regulatedness of an e-learning environment (Liaw and Huang 2013).

As individual factors, we operationalise demographics, contextual and institutional factors (Melzer and Schoop 2015), and learning styles (Honey and Mumford 1992).

Regarding the qualitative analysis, semi-structured focus group interviews with volunteering students are conducted to investigate important factors that emerged in the survey and clarify or complement the findings. The aim is to have three focus groups with at least four students each mixing the courses of study and the individual factors to achieve sufficient heterogeneity within the focus groups facilitating discussion (McLafferty 2004). To encompass not only the student perspective but also the lecturer perspective, observational notes are documented by the lecturer and the supporting research assistant observing the presence lectures. These notes focus on their experiences teaching the course and providing online

and offline clarifications, explanations, and content support regarding all of the course topics.

Objective performance of the students is evaluated analysing the negotiation journal entries (which represent the end of each flipped classroom cycle) showing the final learning outcomes. Thus, scores of these journal entries can be used to estimate learning outcomes within the corresponding units. Furthermore, scores of the summative exam written at the end of the semester can be analysed representing ultimate learning outcomes. Usually, single exam tasks correspond to specific units or topics covered in the course. Therefore, exercise scores can be analysed to investigate learning outcomes of specific units. However, these objective performance indicators have to be handled with care, since there is no valid direct comparison possible due to the change in teaching methods and in the student sample at the same time. Nevertheless, they can be analysed with reference to the respective course units focusing on specific topics, learning tasks, and learning tools.

Finally, measures provided by the learning analytics features of the VLE ILIAS are investigated. Although, ILIAS only provides limited data, forum and mailing statistics can be used to complement the measures on personalisation of communication (i.e. which communication media are used?), social presence and group cohesion (i.e. how many posts are written per medium compared to other courses), and adoption (i.e. access data or ratings of specific preparation pages, learning tasks, and learning tools).

5.6   Discussion and Outlook

This present paper reports on a holistic, mixed-method evaluation concept for personalised flipped classroom university courses following a design-oriented methodology. The described personalised flipped classroom is grounded in the PLF. Due to its novelty and the scarce literature on similar approaches (McNally et al. 2017; Abeysekera and Dawson 2014), a holistic evaluation concept is required combining models from the learning sciences and ISs. In particular, the evaluation concept elicits whether

1) the theoretical requirements regarding personalisation have been fulfilled from a student perspective,
2) the learning outcomes have been improved,

3) the adoption of e-learning elements has been successful, and
4) individual factors acting as moderators can be explicated.

Compared to the scarce literature on flipped classroom creation and evaluation, our concept is unique in several ways. Firstly, it provides a holistic perspective on flipped classroom evaluations and personalised learning following the PLF employing quantitative and qualitative methods. Furthermore, it incorporates all relevant factors present in a real-life university course to create a comprehensive picture.

The presented approach, however, also has some limitations, which mainly stem from its design-oriented approach. Since we want to evaluate a real-life university course over a complete semester providing real grades, it would be unethical and even more effort to instantiate a control group that would receive different teaching. Due to the one-semester character of the implemented flipped classroom, knowledge is gained during the semester. A pre-post-test design investigating learning outcomes would therefore be self-fulfilling and has been discarded. Finally, there are numerous variables that have been identified for investigation of our constructs leading to a high number of constructs for our survey. Therefore, we decided to split the survey into several parts assessing different constructs and topics at different points in time during the course to keep the participant's effort small. Stable constructs such as individual factors are assessed in the beginning of the course, while fulfilment of general requirements, learning outcomes, and adoption are assessed at the end. An evaluation after the exam, when the students have performed the complete course, would not be meaningful either, since the time-span between the last lecture and the exam is usually several weeks. Impressions would have been faded out too much. Furthermore, the proposed interviews may be subject to bias involving volunteering students. Following the results of McNally et al. (2017) there are different groups of students including flip endorsers and flip resisters. Students volunteering for an interview before the final exam probably are mostly flip endorsers.

The quantitative evaluation is partly redundant to the standard student evaluation of teaching prescribed by the university for quality assurance. However, both surveys are necessary, since the student evaluation of teaching does only partly address our constructs of interest (e.g. assessing the quality and adoption of e-learning, self-regulated personalisation, or

individual factors). With regard to section 5.3, we only found very few instruments capable of addressing learning interventions comprising of presence and online learning. Instruments rather investigate either presence or online learning not being able to analyse their interrelationships. Furthermore, the question remains whether exam grades are a truly objective measure of learning outcomes. On the one hand, the lecturers are the ones grading the exam and creating the evaluation standards. On the other hand, a qualitative analysis of graded journal entries might not reveal the students' true opinion, since some might focus on what they think the lecturer wants to read to achieve more points. Besides the lecturing effort, also the efforts to employ the abovementioned holistic evaluation concept is rather high. Although many studies analysing flipped classrooms or learning in general incorporate mixed-method approaches (e.g. comprising of a survey and interviews Lambert and Fisher 2013) the majority focuses on single survey post-test analyses (Bishop and Verleger 2013). Thus, an ex-post analysis of the value of each evaluation method has to be performed refining the presented approach by extending, omitting, or altering specific methods. Overall, we do indeed provide a novel evaluation approach as we have provided a novel teaching approach, both calling for further research in the field of personalised learning.