Development of an Evaluation for Flipped Classroom Courses

Vogelsang, Kristin; Hoppe, Uwe,

1 University of Osnabrueck, Germany
Kristin.vogelsang@uos.de
Uwe.hoppe@uos.de

Abstract. Digital learning has become more than just a trend in the modern world. Blended learning concepts have proven their usability. A further development is the so called flipped classroom. This approach repurposes class time to focus on application and discussion. The acquisition of knowledge will happen at home, enabled by virtual lectures. In the past, many researchers and lecturers demonstrated the advantages of flipped classroom concepts within case studies and self-developed assessment. Still a standardized evaluation of flipped classroom concepts is missing. However, it is necessary to learn from the past and improve future education. This paper presents a generic evaluation of flipped classroom concepts. On the example of a case, we will demonstrate which elements of a flipped classroom concept can be assessed and which measures may serve to improve the acceptance, the learning outcomes and applicability for students and teachers.

Keywords: Evaluation, Assessment, Flipped Classroom, Online Tutorials, Technology Acceptance, Learning Success

1 Introduction

The rising digitalization of economy, society and everyday life also affects educational institutions [1]. Teachers and students are no longer bound to traditional ex-cathedra lectures. They may also attend pure virtual lectures or blends [2]. One possible blended learning approach is the so called flipped or inverted classroom. A flipped classroom (FC) is defined as a concept where interactive, group based learning activities occur inside the classroom, while the individual learning of basic knowledge occurs outside the classroom [3]. Many studies prove the success of FC constructs [4, 5]. Although the construction of flipped classrooms is quite popular, there is still a lack of substantial empirical evidence on learning outcomes [6]. Researchers worry about the deficiency of rigor [3, 7]. Furthermore, scientists identify an absence of comprehension regarding the impact on students’ approaches on learning [4]. The quality of lectures has been a point of controversial discussions in the past [8]. Evaluations help to ensure a constant and transparent quality of lectures and learning outcomes [9]. Still, there is no standardized method for the evaluation of flipped classroom classes available [3]. We
contribute to this research area by introducing a generalizable multi-method approach to evaluate flipped classroom concepts. Two types of evaluation can be distinguished: formative, which aims at the improvement and learning quality and summative, which assess the learning effects [9]. Learning quality is strongly related to learning time, learning success and acceptance [10]. Today the learning time is covered by the workload of the students and given by the so called Bologna-Declaration [11]. Thus, we will concentrate on the learning success and the acceptance. The measurement of learning success and the system acceptance are built on applicable and proven metrics such as technology acceptance model (TAM) [12] and the learning taxonomy of Krathwohl [13, 14]. Based on the existing theory, we will develop a model of a formative, multi-method evaluation of flipped classroom concepts.

In the following, we will give a brief update on shortcomings in flipped classroom assessment research. Afterwards we will introduce the underlying theories and align them to the research topic. We present a model for the evaluation of flipped classroom and use the model on the hand of a case study. Afterwards we will discuss the results. Furthermore, we give a brief summary on limitations and further research.

2 Definitions

2.1 Shortcomings in Flipped Classroom Assessment Research

A flipped classroom is an inversion of traditional classroom and at-home activities [7, 15]. This educational technique consists of two parts: a computer-based individual knowledge acquisition at home and an interactive group based learning environment inside the classroom [3, 16, 17]. The learners receive the basic contents via online tutorials. For the part of training and applying the knowledge, ideally in problem solving situations, they meet vis-à-vis in small groups guided by the lecturer [3]. The individual time of knowledge acquisition is transferred from the group session in classroom to the individual learning environment. The guided knowledge acquisition time can be used to work with the knowledge, to gain and to deepen the skills [16].

The majority of articles about flipped classroom deals with the conceptualization of FC-courses [e.g. 18, 19] for a specific context (e.g. education in health management [20]). Some also include assessments of the presented approaches [21–23]. Findings from existing meta-analysis [3, 24–26] show that the evaluation of FC-concepts includes two major fields: FC-acceptance and the measurement of the learning success. Regarding the acceptance of FC there is a focus on learner acceptance rather than on teacher acceptance of the concepts. Student’s acceptance is often stated as students’ perception or opinion and based on the students’ subjective responses. The items for the measurement often show a lack of the scientific basis [3, 27]. In most studies, acceptance is used as a measure to compare the approval of a FC in comparison to traditional lectures [28–30]. Some studies consider the use frequency of the video tutorials and the use habits as well [31]. The FC acceptance of the teacher is rarer. In this area studies on cost and effort estimations are prevailing [31, 32].
We found less articles dealing with the learning success in FC rather than those dealing with acceptance. Most studies are based on individual statements and test-results [19, 22, 28]. Some researcher conducted tests in order to compare the learning level of a FC-group with those from a group following traditional lectures. Critique came up about comparing test results as a measure for the learning success [3].

Many studies on FC show a lack of rigor [3, 7]. The evaluations were often built as a response to the situation [7]. Most researchers used self-developed scales (e.g. [33, 34]) without a systematic derivation. The findings show a research gap. There is a need for a standardized evaluation of FC classes. Uniform classroom evaluations enable the systematic analysis in order to improve the future learning quality [35, 36]. We will contribute to the discussion and present a multi-method instrument to evaluate FC-courses. The instrument is based on proven theories. The metrics will aim at the basic parameters of learning quality.

2.2 Measuring Learning Success

Knowledge acquisition and the capability to perform on the basis of this knowledge describe the learning process. However, both are not directly observable. The measurement of the learning success can be enabled only by workarounds [27]. “The learning outcomes represent the goal assessment or measures for determining the success of a learning program” [37]. Learning outcomes are collected and categorized in learning taxonomies [27]. The cognitive objectives of the taxonomy of Anderson and Krathwohl [38] can be used to assess the reached level of the learning outcomes. This taxonomy is based on the taxonomy of Bloom [39]. The cited model consists of two dimensions: a cognitive dimension and a knowledge dimension. The first dimension implies the knowledge adoption and the corresponding cognitive processes: remember, understand, apply, analyze and create. The complexity of knowledge adoption increases. This means that the basic learning objectives can be ranked in a hierarchical manner. Basic targets have to be attained before higher levels like apply, analyze and create can be reached. The second dimension distinguishes certain kinds of knowledge like factual, conceptual, procedural and metacognitive knowledge. The taxonomy enables the development of learning objectives. They include the cognitive processes as well as a topic of interest. Thus, the taxonomy helps to compare different teaching approaches. The achievement of higher levels enables higher learning success.

A flipped classroom model is more an expansion of the curriculum, rather than a mere re-arrangement of activities [3]. As shown above, the complex structure of learning success makes it difficult to measure. In the past taxonomies were mainly used to prove the summative effects of FC concepts [21, 26]. A detailed derivation of the learning success is rare. The assessment of learning-success should consider the range of learning outcomes. Well-developed learning objectives are the basis for the comparison of the outcomes. They could be ranked objectively by the attainment of higher learning levels. This makes the taxonomy of Krathwohl an impartial instrument to measure learning-success.
2.3 Acceptance Research

The use of a certain technology often depends on the acceptance of the systems. This relation was proven and established by Davis [12] in the technology acceptance model (TAM) and is still recognized until nowadays. This clear and easy to understand relation, made the TAM one of the most frequently cited models in Information Systems research. The model and its key components are still explored [40]. The basic idea of the TAM is that the intention to use impacts the real use of a technology. The intention is influenced by different factors like the perceived usefulness (PU) of the respective technology and their perceived ease of use (PEU) as well as factors from the fields of social environment, attitude or work-related factors. TAM was strongly influenced by the theory of reasoned action [41]. Usually the core components of TAM are operationalized by different items. The item relations are measured with the help of questionnaires including 7-point-likert-scales. The standardized measures allow to predict the acceptance and to measure the strongest impacts on acceptance. Furthermore, the uniform use enables the comparison of different TAM-studies [42].

By using TAM, we concentrate on the technology level. The behavior of the lecturer and students’ perceptions about the FC-course will be evaluated in a further sequence. The standardized items and the independency from the technology used make the TAM an applicable instrument. In the context of our study the TAM is used to discover the influence factors that impact the decision to use the online tutorials. The examination of the weights of the impact factors can be used to improve the tutorials for future courses. Using the TAM to evaluate learning concepts is not new, though not common. Different studies can be identified, which use TAM for the evaluation of blended learning [43]. In FC-research acceptance is often measured by similar but not standardized constructs [44].

3 Development of the Flipped Classroom Assessment

The assessment model presented aims at the formative assessment of the learning quality. Learning quality can be described by learning time, learning success and the acceptance of teaching tools [10]. These aspects are very different and can hardly be evaluated at once. The content related evaluation takes place a priori and aims at the capabilities of the lecturer. The assessment of the acceptance will happen during the class and focusses on the perception of the students. This means the evaluation has to be proceeded at different times (Table 1).

<table>
<thead>
<tr>
<th>t</th>
<th>What</th>
<th>Measure</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-</td>
<td>Learning level</td>
<td>Learning objectives and levels</td>
<td>Learning goals must be defined and ranked in accordance with the learning levels. Higher learning levels should be enabled.</td>
</tr>
<tr>
<td>classroom</td>
<td>Acceptance</td>
<td>TAM</td>
<td>High value for UI and PU are favourable.</td>
</tr>
<tr>
<td></td>
<td>Perception</td>
<td>Binominal scale</td>
<td>Positive values for individual speed, teacher contact and knowledge acquisition are favourable.</td>
</tr>
</tbody>
</table>

Table 1. Evaluation time table
The interplay from analysis and formative interference of the concept using independent measures, enables the long-time improvement of the course. In the run-up of the class the lecturer has to ensure adequate levels of learning objectives. During the term the objective criterion of the evaluation shifts from the concept of the lecture to the students. Using the TAM, the acceptance of the technical components can be assessed. Although TAM is known as a standard instrument in order to predict future use, the use situation differs. In order to face cohort effects and reduce factors that have a considerable external influence on the results, we built the model with the help of well-considered and limited key variables. This means the model was built to make it easy to use and to understand. Furthermore, we considered major basic conditions aiming at reliability and validity [45] (Table 2).

**Table 2. Conditions for acceptance measurement**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use proven and accepted measures [40, 46]</td>
<td>Using measures from original TAM or proven operationalisations of variables</td>
</tr>
<tr>
<td>Limit the number of factors to guarantee the</td>
<td>Reduction to few, well proven measures</td>
</tr>
<tr>
<td>statistical proof even by small group sizes</td>
<td>Checking the reliability</td>
</tr>
<tr>
<td>[47]</td>
<td>Exclusion of variables used for special technologies (e.g. perceived playfulness for game based learning)</td>
</tr>
<tr>
<td>Independency of measures from the lecture</td>
<td></td>
</tr>
<tr>
<td>objectives [48]</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3. Items and measurement of the acceptance**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Definition</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEU</td>
<td>the degree to which a person believes, that using a system would be free of effort</td>
<td>Interacting with the system does not require a lot of my mental effort. I find the system to be easy to use. … easy to get the system to do what I want it to do.</td>
</tr>
<tr>
<td>PU</td>
<td>the degree to which a person believes that using a particular system would enhance his or her job performance</td>
<td>Using the system improves my performance in my course. … in my course increases my productivity. … enhances my effectiveness in my course. I find the system to be useful in my course.</td>
</tr>
<tr>
<td>SN</td>
<td>the persons perception that most people who are important to him think he should or should not perform the behavior in question</td>
<td>People who influence my behaviour think that I should use the system. People who are important to me think that I should use the system.</td>
</tr>
<tr>
<td>PEn</td>
<td>the extent to which the computer is perceived to be enjoyable […]</td>
<td>7 Point Likert scales between the opposites: Fun/Frustrating; negative/positive; pleasant/unpleasant; pleasurable/painful; exciting/dull; foolish/wise; enjoyable/unenjoyable</td>
</tr>
</tbody>
</table>

We used the key constructs of TAM: PU and PEU [49] and the explanatory constructs subjective norm (SN) [41] and perceived enjoyment (PEn) [12, 50]. These factors aim at the social implications of the tutorial use like social influences and attitude towards
the tutorials. The construct of PEn is expected to play an important role to explain the IU. Hedonistic aspects on software usage become more and more important, especially in learning oriented usage situations. Table 3 shows the operationalization of the constructs. The assessment can be proceeded during the semester.

Furthermore, we evaluated the students’ perception on the influence of the concept and the learning habits and the perception of the lecturer. We developed the scale on the basis of former research in the field of FC and deduced measures for the evaluation. three fields were mainly evaluated to assess the attitude of the students towards the FC concept: individual organization, knowledge acquisition and lecturer contact. The research field is dominated by the assessment of the knowledge acquisition. The evaluation of the influence of the lecturer is a rare. For the development of our evaluation model, we concentrate on FC-courses. Table 4 shows the measures, the focus and the related work.

<table>
<thead>
<tr>
<th>Question</th>
<th>Focus</th>
<th>Related Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can determine my own speed of learning. I can work self-contained.</td>
<td>individual speed and learning time</td>
<td>Evaluation of self-regulated learning and perception about the self-paced time  [21, 51]</td>
</tr>
<tr>
<td>I work continuously over the whole semester.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I receive a good overview on the topic. I can manage the learning matter.</td>
<td>knowledge acquisition</td>
<td>Attitude towards FC in regard to pass the exams and to emphasize learning success  [18, 19, 28, 29, 52, 53]</td>
</tr>
<tr>
<td>I feel well prepared for the tests.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The lecturer is dedicated. The lecturer takes notice of me.</td>
<td>lecturer contact</td>
<td>Assessment of the awareness of the lecturer  [20, 29, 34, 54]</td>
</tr>
</tbody>
</table>

In order to present the applicability of the evaluation concept, we will demonstrate a use case of the evaluation in the next chapter. The case will show, how the evaluation can be proceeded and how the results can influence the conceptualization of further classes.

4 Application Example of the Evaluation

The case presented is a master class with about 100 participants. The subject of the course is “Project Management”. The FC-concept is built of three different bricks: the basic learning content taught via online video tutorials, face to face reading seminars and guest lectures from practitioners. The video tutorials enable the gain of fundamental knowledge during the individual learning time [2]. The tutorials are based on sets of slides, that contain the basic content. The slides were verbally described and complemented by examples and comprehensive case descriptions. The sets of slides all include the fundamental knowledge in accordance with the taxonomy used like definitions, rules and concepts of the field. 25 videos with an individual length between 10 and 40 minutes are provided. The division into sections enables an individual learning style [25, 26, 55, 56]. All videos have a total length of 307 minutes. The second
brick is presented by guest lectures. The third brick is the so-called “reading seminar”. The seminar is made up for a maximum of 20 learners each. The attendance of this seminar is voluntary. Though the students receive a reward for the participation in form of bonus points for the final exams. Students are confronted with articles from specific journals, that correspond with the topics of the video-tutorials. The students are asked to prepare the texts at home using a questionnaire as guideline. During the reading seminar they have to present their results, discuss them with their peers and align them to the basic contents of the class. Table 5 shows the achievable learning levels for the online tutorial (Lecture 8 “Project Execution) and the corresponding reading class (Lindner, F.; Wald, A.: Success factors of knowledge management in temporary organizations, in: International Journal of Project Management, 29 (2011)).

<table>
<thead>
<tr>
<th>Level</th>
<th>Online tutorial</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>create</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>assess</td>
<td>-</td>
<td>You are able to assess the knowledge management methods presented. You can assess the results in regard of the context.</td>
</tr>
<tr>
<td>analyze</td>
<td>-</td>
<td>You understand which methods can be used and choose the right one regarding the context.</td>
</tr>
<tr>
<td>apply</td>
<td>You are able to use methods of quality control</td>
<td>You are able to use the methods demonstrated.</td>
</tr>
<tr>
<td>understand and remember</td>
<td>You are able to define the term „product quality“</td>
<td>You may understand the hypotheses set up in the text and the results of the hypothesis testing.</td>
</tr>
<tr>
<td></td>
<td>You receive an overview about the topic of project quality</td>
<td>You are able to describe the key concepts of the text.</td>
</tr>
</tbody>
</table>

Table 5 gives an overview of learning levels on the example of the case. We realize a gap at the level of “create”, which has to be patched.

During the term, we asked the students to fill out a questionnaire in order to evaluate the acceptance and the perception of the students. We received 68 completed questionnaires from the students. Cronbach’s alpha is for all constructs higher than 0.8, which is regarded as a significant value for the reliability of the items [47].

![Figure 1. Technology Acceptance Model for the Flipped Classroom](image)

We proceeded a linear regression analysis to measure the impact of the constructs on the use intention (UI). Figure 1 shows that the model is able to explain about 60% of the Use Intention (adjusted R²=0,595). This can be regarded as a comparatively high
value in order to explain the intention of the students to use the system [40]. The main predictors of the UI are PU and Perceived Ease of Use. PU can be explained to more than 50% (adjusted R²=0.565). It is mainly influenced by PEn, rather than by SN and EU. This means the intrinsic motivation to perform is higher than the influence by the peer group. It is important to mention that the software-use situation can be described as voluntarily. Although highly recommended by the teacher in order to pass the class, PEn and Subjective Norm have no significant direct impact on UI. This can be explained by the utilitarian use situation. Hedonistic and social aspects become less important, when the use-situation is work-related [50]. EU impacts the UI directly as well as indirectly via PU. The students regard the access to the system as easy and uncomplicated.

Table 6. Results regarding the Students’ Perception.

<table>
<thead>
<tr>
<th>Question</th>
<th>FC-Concept</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can determine my own speed of learning.</td>
<td>62</td>
<td>67</td>
</tr>
<tr>
<td>I can work self-contained.</td>
<td>60</td>
<td>66</td>
</tr>
<tr>
<td>I receive a good overview on the topic.</td>
<td>56</td>
<td>68</td>
</tr>
<tr>
<td>I can manage the learning matter.</td>
<td>56</td>
<td>65</td>
</tr>
<tr>
<td>I feel well prepared for the tests.</td>
<td>54</td>
<td>61</td>
</tr>
<tr>
<td>The lecturer is dedicated.</td>
<td>30</td>
<td>65</td>
</tr>
<tr>
<td>The lecturer takes notice of me.</td>
<td>24</td>
<td>57</td>
</tr>
<tr>
<td>I work continuously over the whole semester.</td>
<td>30</td>
<td>62</td>
</tr>
</tbody>
</table>

Table 6 shows, a great majority thinks that the FC enables to work with an individual speed and self-paced. Most students think they can manage the knowledge acquisition [43, 57]. However, less than 50% of the students sense the dedication of the lecturer and the interplay positive. And the permanent learning during the semester is not guaranteed by the concept. Three fields force the lecturer to act. In future the contact time hast to be upgraded in quality and social turns towards the students and the continuously working has to be improved.

The evaluation of the case with the help of the FC evaluation leads to the following insights: The learning levels which can be achieved reach up to “assess”. This learning level is only enabled by the additional contact time during the reading seminar. The online tutorials are regarded as easy to use and usable. Sadly, the students do not emphasize fun while working with the tutorials. The students experience the FC-model positive regarding their individual learning time and the knowledge acquisition. The contact time and the role of the lecturer could be revised as well as cooperative learning facilities [58].

5 Conclusions and Limitations

Recent research shows some deficiencies regarding rigor and validity in measuring the effects of flipped classroom [3, 4, 7]. Our research contributes to this discussion, as this paper presents an evaluation instrument to assess the learning quality of flipped classroom courses. The model was developed with the help of proven measures from
the taxonomy of Krathwohl and TAM. Furthermore, the students’ perceptions towards of the knowledge acquisition can be measured. For the development of this scale we used existing flipped classroom evaluations. We emphasis the FC-evaluation presented as easy to use and to understand. Still, the future quality of the evaluation depends on more feedback from more cases in order to improve the model. Only standardized use will lead to comparable and better results.

Future research and use may enhance the model. As the model presented is appropriate to detect fields for further improvement, a well-defined action plan has to be developed and deduced from more cases. It would be of great value to increase the number of respondents for the development of the structure model. Furthermore, the relation between the learning objectives and the assessment of the knowledge effectively gained has to be clarified in future [25].

References

44. Yoshida, H.: Perceived Usefulness of “Flipped Learning” on Instructional Design for Elementary and Secondary Education: With Focus on Pre-service Teacher


