Reviewing Assessment in Online and Blended Flipped Classroom

Professor Blaženka Divjak[a], Professor Bart Rienties[b], Barbi Svetec[a], Petra Vondra[a], and Professor Mirza Žižak[c]

[a] University of Zagreb, Faculty of Organization and Informatics, Croatia; [b] The Open University, The United Kingdom; [c] University of Zagreb, School of Medicine, Croatia

Abstract

The aim of this study was to investigate the guiding principles, practises, and challenges of assessment in online and blended flipped classroom approaches in higher education. Based on that, we proposed a generic model of teaching, learning and assessment in online and blended flipped classroom, including examples of teaching and learning activities, with corresponding levels of learning outcomes, and assessment tasks and methods. We synthesised the findings and suggested further research on the following aspects: peer-assessment and self-assessment; benefits and challenges of educational technology; ethical aspects of assessment and plagiarism; balance between flexibility and structure. We recommended that further research should put more focus on the utility, including validity and reliability of assessment and test the proposed generic model in different study fields and educational contexts. Our study was based on a systematic literature review, including 90 articles identified in the Web of Science database, with 23 studied in detail.

Citation

Introduction

Since the onset of the COVID-19 pandemic, a strong emphasis has been placed on innovative pedagogies, digital tools, and methods to ensure the quality and inclusiveness of education. In this context, the flipped classroom has stood out as a particularly relevant and innovative pedagogical approach in higher education (HE) (Divjak et al., 2022).

Assessment is essential for enabling engaging and productive learning in flipped classroom environments. It may help overcome the challenges related to the fact that, in flipped classroom, students need to engage with new material independently. (Talbert, 2015) There are also general issues related to e-assessment, and the European Union Member States pointed out assessment and grading as one of the greatest challenges of education experienced early in the pandemic (Council of the EU, 2020).

Regardless of this, in relation to higher education, not many studies have focused specifically on assessment in online and blended flipped classroom, and this has been identified as an open research question (Divjak et al., 2022). Our primary motivation for conducting this study was to provide more evidence and a more systematic overview of practices and challenges related to assessment in online and blended flipped classroom. Without meaningful assessment practices, the effectiveness of innovative pedagogies, including flipped classroom, cannot be increased, nor their potential fully exploited.

Previous studies related to flipped classroom and technology for assessment called for more elaboration of research approaches, research design, and theoretical foundations in future studies (Divjak et al., 2022), and underpinning practices with sound pedagogy and educational theory (Brady et al., 2019). It has also been noted that, in flipped classroom research, empirically well-grounded and rigorous research is rare (Lundin et al., 2018).

The aim of this study was to analyse the existing research related to assessment in online and blended flipped classroom, considering respective research approaches and theoretical foundations, and identify some guiding principles, practices, and challenges. Based on the findings, our aim was to propose a generic model of flipped teaching, learning and assessment, that could support practitioners, in various educational settings, in the sound planning of online and blended teaching and learning in line with flipped classroom pedagogy. Moreover, our aim was to identify further research gaps and provide recommendations for future research.

Theoretical Background

Flipped Classroom

The flipped or inverted classroom is a student-centred approach designed to increase the quality of in-class learning. This is achieved by moving information transmission out of class time (Abeysekera & Dawson, 2015), allowing for more structured, active learning (Strelan et al., 2020), with possibilities for students to investigate and interact during classes (Divjak et al., 2022). In terms of the revised Bloom's taxonomy (Anderson & Krathwohl, 2001), pre-class activities in flipped classroom are related to lower levels of cognitive work, while class time is related to higher levels of cognitive work (Brame, 2013).

Assessment

Assessment is an essential part of learning. It serves not only to measure, but also to support student learning (Brady et al., 2019). Meaningful assessment can guide learning and positively influence the acquisition of learning outcomes (LOs), while inappropriate assessment may hinder learning. Therefore, it is essential that teachers use assessment methods directly related to intended LOs, in line with the principle of constructive alignment (Biggs, 1999), supporting a deep approach to learning (Entwistle & Ramsden, 2015).

There are several criteria affecting the utility of assessment. According to the conceptual model for defining the utility of assessment (van der Vleuten, 1996), the criteria include reliability, validity, educational impact, acceptability, and cost of assessment. Criteria are weighted based on the importance attached to them by particular users in respective contexts. In relation to this, an approach has been proposed (Divjak et al., 2021) in which weights are assigned to intended LOs, and distributed to assessment tasks aligned with the LOs, using multi-criteria decision-making.

Methods used to assess learning include formative and summative assessment. While the first refers to collecting data in order to improve students' learning, the second refers to using data to assess students' knowledge after completing a learning sequence (AERA et al., 2014; Dixson & Worrell, 2016). It has been pointed out (Ramsden, 2003, p. 187) that there is 'no sharp dividing line between assessment and teaching in the area of giving comments on learning'. This points to the need for alignment between formative and summative assessment. It also implies that the same assignments can be used both for formative assessment and feedback, and a final mark contributing to a course grade (p. 190). Whatever the type of assessment, it should meet the minimum reliability and validity standards (AERA et al., 2014). In flipped classroom approaches, particular focus is usually put on formative assessment, which is particularly relevant with respect to individual student work (Talbert, 2017).

Assessment and Technology

Technology has the potential to enhance assessment processes in HE. It offers numerous possibilities, related to, for example, collaborative learning and prompt feedback (Brady et al., 2019; Daly et al., 2010). However, despite the growing availability of technology, its use in assessment is not consistent among academic staff (Bennett et al., 2017; Brady et al., 2019). It has been pointed out (Timmis et al., 2016) that terminology changes reflect the evolution of understanding of the potential technology brings to assessment, with a shift from computer-based testing, to computer assisted (or aided) assessment, to online assessment, and e-assessment. For the purpose of this study, we consider that e-assessment involves the use of ICT to present and deliver assessment tasks, receive and record responses, and provide feedback to students.

Learning Theories and Research Worldviews

Educational practice and research should be immersed in and underpinned by contemporary learning theories, and contribute to their further development. Here we refer to the learning theories most relevant with respect to this study. *Cognitivism* focuses on knowledge acquisition, as 'a mental activity that entails internal coding and structuring by the learner' (Ertmer & Newby, 2013). *Constructivism* presumes that learners themselves create learning, which takes place in contexts (Schunk, 2012, pp. 230, 491). It assumes that teachers should not deliver instruction,

but rather create situations, so that learners 'become actively involved with content through manipulation of materials and social interaction' (Schunk, 2012, p. 231). *Humanism* and humanistic teaching aim to 'encourage personal growth by providing students with choices and opportunities', with teachers as facilitators, providing resources and encouragement (Schunk, 2012, p. 356). *Connectivism* sees learning as a process of connecting information sources, pointing out that the 'capacity to know is more critical than what is currently known', while decision-making is a learning process (Siemens, 2005).

Moreover, researchers need to consider philosophical worldview assumptions, related research design, and specific methods translating their approach to practice (Creswell, 2018, p. 43). Here we refer to the research worldviews (Creswell, 2018) most relevant with respect to this study. *Postpositivist* worldview presumes deterministic philosophies, and is often related to quantitative research and the scientific method in which researchers collect data to support or refute a theory. *Pragmatic* worldview is problem-centred and oriented towards real-world practices, using all available approaches to solve a problem, and often related to mixed methods. *Constructivist* worldview focuses on understanding and interpreting meanings, looking at participants' views in their complexity, and is usually related to qualitative research.

Methodology

Our research was led by the following research questions (RQ) related to online and blended flipped classroom:

1. What are the predominant research designs and the theoretical foundations of the analysed studies?
2. What are the main findings on the guiding principles and practices in assessment?
3. What is the underlying generic model of flipped teaching, learning, and assessment?
4. What are the key challenges and recommendations for future research?

To respond to the research questions, we conducted a systematic literature review of studies related to assessment in online or blended flipped classroom. The selection of studies included several steps, presented in the further text and Figure 1, in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, as a set of evidence-based items for reporting in systematic literature reviews (Page et al., 2021).

The first step was identification of the initial group of papers related to the research topic. A search was done in the Web of Science (WoS) database, chosen based on its ranking among academic databases and coverage of relevant research. The search was conducted by two researchers (BD, BS) on 14 October 2021, repeated on 13 July 2022, and included all relevant papers, with no pre-defined timeframe. It was based on the following search string: ((("online" OR "distance" OR "distant" OR "ICT" OR "blended")) AND ("flipped classroom" OR "flipped learning" OR "inverted classroom") AND "higher education" AND "assessment". The first search resulted in 77 papers, and the second in 90 papers, from the period between 2013 and the date of the search. All the identified papers were included in the next step.

The second step was screening of the identified papers. Titles and abstracts were examined by two researchers (BD, BS). The inclusion criteria were focused on the relevance of the papers with respect to online teaching and learning, blended teaching and learning, higher education, flipped

classroom, and assessment. Consequently, the total of 32 papers in English were recommended for further analysis. The remaining 58 papers were excluded. Out of that, 57 papers did not correspond to one or more of the said inclusion criteria, and were consecutively excluded due to the following: five papers were not focused on teaching and learning in HE, seven were not focused on flipped classroom, and 45 papers did not substantially contribute to findings on assessment. Additionally, one paper was excluded as it was identified as duplicate. The two researchers cross-checked each other's recommendations, in order to ensure objectivity and reduce the potential bias. In the third step, all of the papers were retrieved for a detailed analysis.

The fourth step was eligibility assessment, with the 32 papers examined in detail by seven researchers (BD, BR, BS, FI, MZ, PV, VK). Each researcher examined between three and seven papers, analysing the following: country, study field, level of HE, research worldview, research approach, methodology and methods, learning theory, participants, main findings on assessment, limitations, and future research recommendations. Finally, the researchers evaluated the assigned papers' relevance for the final analysis, which resulted in 9 articles being excluded due to insubstantial findings related to assessment in flipped classroom, and 23 articles included in the final analysis.

Figure 1

Systematic literature review flow diagram
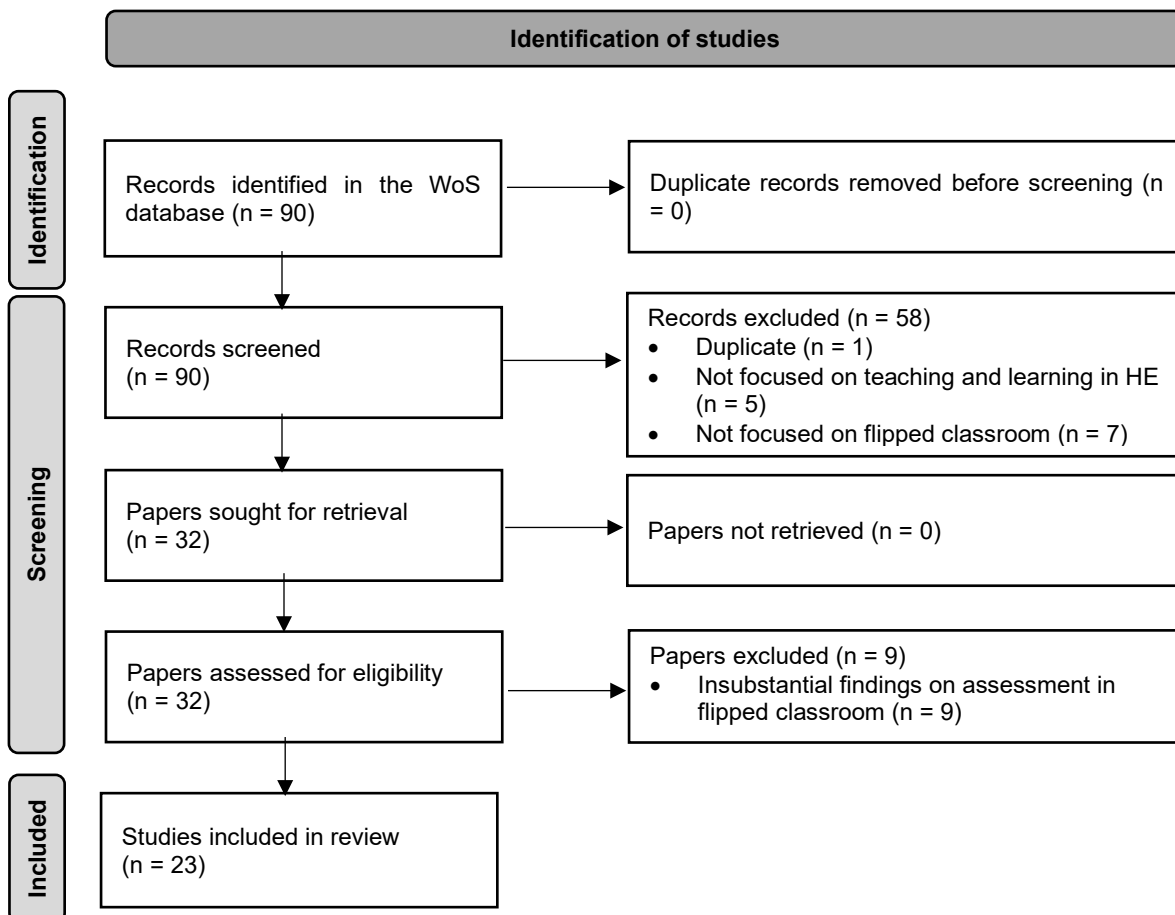
Results and Discussion

The vast majority of the selected 23 papers covered the period between 2017 and 2022, which points to the increase in the relevance of this topic in the past several years. The final list of the selected papers, with the key findings, is presented in the Appendix.

Sample Analysis

Among the selected 23 studies, the majority were performed in the field of Computer Engineering (7) and Computer Science (3), followed by Social Sciences (4), Natural Sciences (3), and Engineering (2). Several studies were also carried out in Education and Educational Research (3), Mathematics (1), and Foreign Languages (1). The vast majority of the studies were conducted at the undergraduate (Bachelor's) level (18), a couple at the graduate (Master's) level (4), and in tertiary vocational education (1).

Concerning the geographical distribution, the studies were performed in Spain (8), Taiwan (4), Australia (2), Norway (2), Brazil (1), Canada (1), France (1), Indonesia (1), Portugal (1), Turkey (1), Israel (1), and the USA (1).

Regarding the research design, most of the studies were quantitative (11), followed by mixed (6) and qualitative (4) research. Designs including surveys were frequently used (10), followed by experimental approaches (5), and case study research (6). Moreover, interviews (2), deep learning (2), focus groups (1), action research (1), and reflection analysis (1) were used.

Some of the studies were categorised in more than one category (e.g., conducted in more than one country or study field). Besides the mentioned original research papers, this study also included systematic literature reviews (2), which were, due to their comprehensive nature, not categorised with respect to the country, and in terms of the study field, were marked as educational research. The overall sample analysis is presented in Figure 2.

Figure 2

Sample Analysis
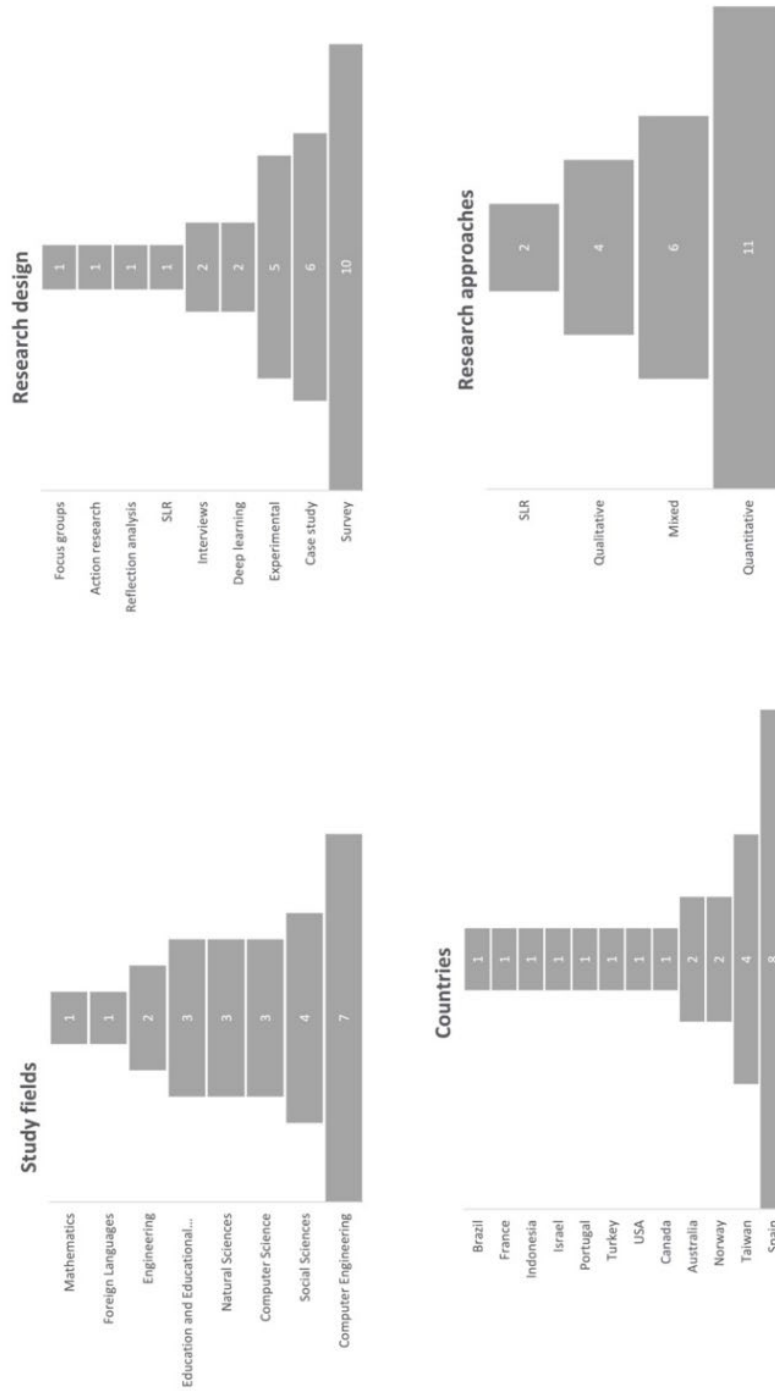
RQ1: Research Designs and Theoretical Foundation of the Studies

Research Approaches and Designs

In general, research worldviews were not often elaborated or explicitly mentioned. However, their characteristics were often discernible from the overall research concept. Several studies (Jensen et al., 2018; Llamas-Nistal et al., 2019; Santos, 2021; Wang, 2017, 2019) adopted a predominantly postpositivist worldview, using experimental approaches and collecting quantitative data to support their hypotheses. Others (Arruabarrena et al., 2019; Mojtahedi et al., 2020; Moreno-Ruiz et al., 2019) were more inclined to a pragmatic worldview, often using mixed research methods to find solutions to practical problems. There were also studies led by a constructivist worldview (Wanner & Palmer, 2015), looking to understand meanings constructed by participants, extracted - to some extent - by qualitative methods.

The majority of the analysed studies adopted quantitative approaches to research. This predominantly included two research designs often used in educational research (Creswell, 2012, p. 293). First, survey designs, predominantly cross-sectional, (Lopes et al., 2019; Moreno-Ruiz et al., 2019; Ruiz-Jimenez et al., 2022; Yilmaz et al., 2020), used to identify participants' beliefs and attitudes (Creswell, 2012, p. 376). Second, experimental designs (Jensen et al., 2018; Llamas-Nistal et al., 2019), used to establish possible causes and effects, while controlling several variables that could influence the outcome (Creswell, 2012, p. 295). Generally, in education, and also in our sample, this frequently refers to quasi-experimental designs (Chien & Hsieh, 2018; Jensen et al., 2018; Lin, 2019), as in educational settings it is often difficult to randomly assign participants to different treatment groups (Creswell, 2012, p. 309). Several studies used mixed methods designs (Mojtahedi et al., 2020; Santos, 2021; Zainuddin et al., 2021), using both quantitative and qualitative data to provide a better understanding of the research problem (Creswell, 2012, p. 535). Purely qualitative approaches, most appropriate for addressing research problems when variables are unknown and there is a need to explore (Creswell, 2012, p.16), were the least common, and included primarily case-studies (Bye, 2017; Luth-Hanssen et al., 2018; Mikic-Fonte et al., 2020).

Learning Theories

Flipped classroom approaches offer possibilities for active learning and stimulating students' autonomy. Therefore, it is no wonder that in the majority of the studies the influence of constructivism could be recognised (Estriegana et al., 2019; Jensen et al., 2018; Lopes et al., 2019). Several studies also explicitly referred to constructive alignment (Biggs, 1999) (Bye, 2017; Luth-Hanssen et al., 2018; Mojtahedi et al., 2020; Wanner & Palmer, 2015). Furthermore, being student-centred and attempting to stimulate students' engagement, flipped classroom approaches also relate to humanism. For example, Wanner and Palmer (2015) proposed flexible assessment, offering choices and individualisation to students and considering their potential learning difficulties. Moreover, as flipped classroom is strongly supported by technologies, some aspects of the studies demonstrated links to connectivism. For example, the approach presented by Wang (2019) included students browsing relevant online material. However, teaching depends on what is being taught, and facts, concepts or problem-solving are taught in different ways (Ertmer & Newby, 2013). Therefore, depending on the level of processing needed for a specific task, flipped classroom approaches may rely on the mentioned contemporary learning theories,

but for some tasks, they also rely on cognitivism. In this respect, it is interesting to point out that Bye (2017) referred to the cognitive load theory, as a bottleneck which appears when complex information must be processed in a short time span.

RQ2: Main Findings on Guiding Principles and Practices in Assessment

In general, few studies were strongly and specifically focused on assessment. Assessment was often referred to as part of overall teaching and learning strategies, based on or integrating flipped classroom. However, the guiding principles, practices and challenges presented in the papers present important findings of this systematic literature review research. In the analysed approaches, the principle of constructive alignment (Biggs, 1999) was recognised. The alignment of assessment with teaching and learning activities and LOs was pointed out by several authors (Bye, 2017; Luth-Hanssen et al., 2018).

Continuous, multi-dimensional assessment was emphasised, which usually included various formative and summative tasks (Ruiz-Jimenez et al., 2022). Learning was often described in weekly blocks, starting with studying pre-class materials such as videos (Jensen et al, 2018; Llamas-Nistal et al., 2019; Mikic-Fonte et al., 2020; Mojtahedi et al., 2020; Ruiz-Jimenez et al., 2022; Wanner & Palmer, 2015; Yilmaz et al., 2020), interactive tutorials (Jensen et al., 2018), web links or reading materials (Bye, 2017; Jensen et al., 2018; Mojtahedi et al., 2002; Wang, 2019; Wanner & Palmer, 2015), or other homework exercises (Bye, 2017). These pre-class activities were usually followed by 'simple' assessments covering the introductory content, often in the form of online quizzes (Jensen et al., 2018; Llamas-Nistal, 2019; Mojtahedi et al., 2020; Wang, 2017, 2019; Wanner & Palmer, 2015; Yilmaz et al., 2020), possibly supported by gamification (Lopes et al., 2019; Mikic-Fonte et al. 2020). Quizzes were often taken before (Jensen et al., 2018; Mojtahedi et al., 2020; Wanner & Palmer, 2015), but sometimes also during (Klegeris, 2021; Mikic-Fonte et al., 2020; Wang, 2017, 2019; Zainuddin et al., 2021) or after face-to-face classes (Llamas-Nistal, 2019; Yilmaz et al. 2020). Such assessment activities were primarily envisaged as formative (Mikic-Fonte, 2020; Wang, 2017, 2019; Yilmaz et al., 2020). While in some approaches (Mojtahedi et al., 2020) pre-class online quizzes covering weekly videos did not contribute to the final grade, in others (Mikic-Fonte et al., 2020) in-class exams covering weekly videos contributed to the final grade, whereas gamified assessment did not. In other approaches, however, in-class gamification activities also contributed to the final grade (Estriegana et al., 2018).

Online quizzes seem particularly relevant in terms of simple assessment tasks, related to lower levels of LOs. In one of the studies (Mojtahedi et al., 2020), students perceived pre-class quizzes as the most effective learning activity. Another study (Zainuddin et al., 2021) found that interactive gamified e-quizzes enhanced students' engagement in online classes.

Assessment also included complex tasks, related to higher levels of LOs. The progression in the levels of LOs was clearly demonstrated in a study (Jensen et al., 2018) pointing out that, following each pre-class assignment, students took so-called Explore Assessments, short online quizzes at low levels of Bloom's taxonomy (mostly remember and understand), while every two weeks they took Apply Assessments, at a variety of levels (mostly apply and above).

In relation to higher levels of LOs, active learning sessions included discussions and teamwork (Bye, 2017; Klegeris, 2021; Mojtahedi et al., 2020; Ruiz-Jimenez et al., 2022; Wang, 2017, 2019),

often done in class, with a frequent emphasis on problem-based and project-based learning (Klegeris, 2021; Moreno-Ruiz et al., 2019; Ruiz-Jimenez et al., 2022; Santos et al., 2021; Wang, 2017, 2019). For example, students worked on assignments such as essays and reports (Luth-Hanssen et al., 2018; Wanner & Palmer, 2015), as well as case-based assignments focused on critical analyses (Mojtahedi et al., 2020), discussions of clinical cases (Klegeris, 2021) or coding challenges (Bye, 2017).

Active learning sessions sometimes included or were followed by peer-assessment (Mojtahedi et al., 2020; Wang, 2017), student presentations (Bye, 2017; Moreno-Ruiz et al., 2019), in-class quizzes (Bye, 2017; Mojtahedi et al., 2020; Wang, 2017), gamification and student competitions (Bye, 2017; Ruiz-Jimenez et al., 2022), interactive online lessons enabling assessment of knowledge (Mojtahedi et al., 2020), and/or homework (Wang, 2017, 2019). After (online) lessons, in some cases students also conducted self-assessment or self-reflection (Lopes et al., 2019; Luth-Hanssen et al., 2018; Wang, 2017, 2019), or students' artefacts were compiled in portfolios (Luth-Hanssen et al., 2018; Wanner & Palmer, 2015). Depending on the approach, these activities were either formative or summative in nature.

Some approaches integrated several periodical examinations, including theory and problem-solving or project-based learning, as well as independent research or assigned reading (Estriegana et al., 2019; Klegeris, 2021; Wang, 2017). Such examinations were envisaged as either formative (Wang, 2017, 2019) or summative (Jensen et al., 2021; Lopes et al., 2019; Luth-Hanssen et al., 2018; Mikic-Fonte et al., 2020;). Moreover, in some cases, assessment also included virtual laboratory practices (Estriegana et al., 2018). Final summative assessment usually included final exams (Bye, 2017; Jensen et al., 2021; Lopes et al., 2019; Mojtahedi et al., 2020; Wang, 2017, 2019), submission of assignments or projects (Mojtahedi et al., 2020; Wang, 2017, 2019;), as well as final versions of portfolios (Luth-Hanssen et al., 2018).

Final grades were often based on different proportions of formative and summative assessment results (Mikic-Fonte et al., 2020; Mojtahedi et al., 2020; Wang, 2017). However, in some approaches, the final grade was based exclusively on the final exam (Bye, 2017). In some examples, the final grade included peer-marks (Klegeris, 2021), while in other authors did not recommend peer-assessment to affect the final grade (Arruabarrena et al., 2019).

One of the studies (Klegeris, 2021) supported previous findings suggesting that multidimensional evaluation is superior to any uni-dimensional testing, as it can provide students with improved feedback on their strengths and areas for improvement. Moreover, it indicated that student performance in any type of assessment cannot predict their scores in others.

Continuous assessment was often followed by feedback, provided several times during the assessment process, helping students to review and direct their learning. There were examples of automated feedback (Lopes et al., 2018; Luth-Hanssen et al., 2018; Yilmaz et al., 2020), as well as feedback provided by peers and teachers (Moreno-Ruiz et al., 2018; Luth-Hanssen et al., 2018). In one of the approaches, students were allowed to indicate the preferred kind and focus of feedback (Wanner & Palmer, 2015). One of the studies (Ruiz-Jimenez et al., 2022) showed that, as reported by students, continuous work and feedback received in formative assessment enabled them to learn more, and improved their perspectives on the results of final assessment.

Several authors referred to clear criteria as the basis for assessment. Criteria were mentioned as relevant in relation to various types of feedback, whether automated (Luth-Hanssen et al., 2018; Yilmaz et al., 2020), peers' or teachers' (Luth-Hanssen et al., 2018), or self-assessment (Blau & Shamir-Inbal, 2017). In relation to this, the importance of assessment rubrics was stressed by several authors, in particular in relation to peer-assessment (Arruabarrena et al., 2019; Mojtahedi et al., 2020; Moreno-Ruiz, 2018).

Students were often included as active agents in assessment. One way of giving students an active role was peer-assessment (Arruabarrena et al., 2019; Blau & Shamir-Inbal, 2017; Mojtahedi et al., 2020; Wang, 2017). A question was opened whether peer-assessment should be anonymous (Klegeris, 2021) or non-anonymous (Arruabarrena et al., 2019). Moreover, some authors pointed out the importance of peer-assessment scaffolding, by introducing guidelines for supportive feedback and 'feed-forward' (as envisaged by Hattie & Timperley, 2007), with the teacher seen as the role-model for peer-assessment (Luth-Hanssen et al., 2018). Interestingly, one of the studies (Mojtahedi et al., 2020) found that, in terms of enhancing learning, students considered peer-assessment to be the least useful (in comparison to e.g. quizzes or tutorials). However, still more than a half found that providing peer-assessment and receiving feedback enhanced their learning. Furthermore, some approaches included self-assessment or self-reflection (Blau & Shamir-Inbal, 2017; Lopes et al., 2019; Wang, 2017, 2019; Yilmaz et al., 2020). One of the studies (Wang, 2017) found that in-class learning can enhance self-reflection, and self-reflection can foster out-of-class learning and achievement. An important aspect related to peer-assessment and self-assessment refers to ensuring their validity and reliability (Arruabarrena et al., 2020; Divjak et al., 2022; Yilmaz et al., 2020). In this respect, it was pointed out that previous studies had shown that peer-assessment demonstrated high levels of validity and reliability, whereas the validity of self-assessment was questioned (Arruabarrena et al., 2019).

In one case (Luth-Hanssen et al., 2018), students were included in the development of assessment criteria, used by peers and teachers. Moreover, one of the studies (Zainuddin et al., 2021) found that both students and teachers could benefit from student-created quizzes and recommended 'flipping the ownership' to challenge students. In another approach (Arruabarrena et al., 2019), peer-assessment and co-creation were used to develop student-generated content. Finally, it was shown (Luth-Hanssen et al., 2018) that involving students in online study programmes, and especially assessment, fostered their engagement, understanding of their learning paths, and taking responsibility for achieving LOs.

The studies pointed to the dichotomy between the need for flexibility of assessment on the one hand, and clear structure on the other. Some authors (Wanner & Palmer, 2015) argued for flexible assessment, which included a personal assessment plan. Students were given the choice regarding the assignment type, submission deadline, and feedback type. The study showed that students wanted more personalised learning and assessment, but they still wanted their courses and assessment to have structure. On a similar note, it was pointed out (Bye, 2017) that the right balance should be found between too rigidly defined learning activities and problem-based learning activities defined too vaguely.

In the analysed approaches, assessment was strongly supported by technology. Comprehensive online learning management systems (LMS) were frequently stressed. Moodle was given particular prominence (Lopes et al., 2019; Wang, 2017, 2019; Yilmaz et al., 2020), but there were

examples of other LMSs used by particular higher education institutions (HEI) (Bye, 2017; Jensen et al., 2019; Moreno-Ruiz et al., 2019). There were also examples of tools used specifically for assessment, whether commercial, such as Socrative (Lopes et al., 2019) or Zuvio (Lin, 2019), or developed by particular HEIs (Llamas-Nistal et al., 2019). Moreover, gamification was part of several approaches (Lopes et al., 2019; Mikic-Fonte et al., 2020; Zainuddin et al., 2021), supported by tools such as Kahoot (Lopes et al., 2019). In some of the approaches (Luth-Hanssen et al., 2019), learning analytics were used to assess students' effort, progress, self-assessment and experience of the learning environment. Furthermore, one study (Lin, 2019) showed that students learning with online peer-assessment showed no significantly better learning achievement compared to those with conventional peer-assessment. Nevertheless, they showed a significantly higher time involvement and autonomy in flipped learning. Besides the benefits, some authors also mentioned the downsides to the use of technology and online assessment, in particular those related to ethical issues, such as plagiarism and cheating (Divjak et al., 2022; Brady et al., 2019; Llamas-Nistal et al., 2019; Mojtahedi et al., 2020).

Finally, it was noted that, to enhance assessment, the development, design and use of technology should be based on relevant pedagogies and educational theories. However, one study (Brady et al., 2019) found that technology for assessment is 'still at an early stage of adoption with limited pedagogical underpinnings or theoretical frameworks'. From the academic staff perspective, assessment design was not of essential concern when introducing technology for assessment, as there were various pressures and educational, as well as operational, drivers. Clearly, technology should be used in line with effective, valid, and reliable assessment practices. An overview of the identified key elements of assessment, together with their mutual relationships as perceived by the authors of this article, is presented in Figure 3.

Figure 3.

Key elements of assessment in online and blended flipped classroom
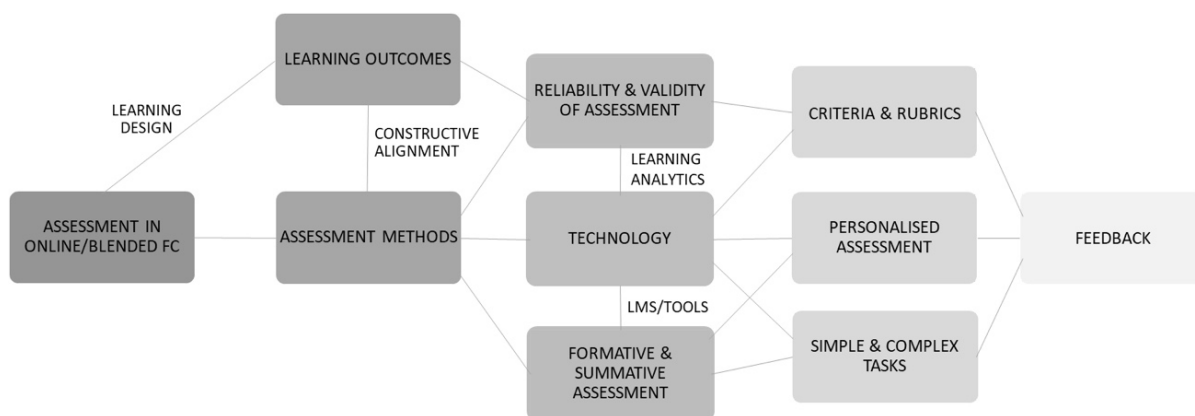
Limitations of the Analysed Studies

Although this is not a limitation of the studies per se, it should be noted that most of the studies included in this systematic literature review were not focused specifically on assessment in online and blended flipped classroom approaches. The studies dealt with online or blended flipped classroom or e-assessment in more general terms. Therefore, relevant findings are in some cases rather limited, but still provide valuable input for further discussion.

As for the inherent limitations, the majority of the studies were focused on a specific case or a small-scale experiment, with a short time-frame, relatively small sample, limited student populations, specific study field, particular level of HE, specific educational institution, and context.

Another limitation is that, in discussing flipped classroom and assessment, authors often did not refer to LOs, but rather focused on students' experiences or grades as single proxies of achievement (Chien & Hsieh, 2018; Moreno-Ruiz et al., 2019; Wanner & Palmer, 2015). There was also not much focus on the utility of assessment (as envisaged by van der Vleuten, 1996). Moreover, as also found in the analysed systematic literature review (Brady et al., 2019), not many studies focussed on online assessment programmes from the perspectives of academic staff.

Finally, not all of the analysed papers were soundly theoretically founded, whether in terms of research approaches or learning theories, so the theoretical background was often not easily discernible or not clearly linked to the concept and the findings of a study. Therefore, it may be challenging to use such findings in the further development of a particular learning theory. A similar observation was made in the analysed systematic literature review (Brady et al., 2019), which found there was sometimes no demonstrable connection between theoretical frameworks and practices, or between study designs and findings.

RQ3: Generic Model of Flipped Teaching, Learning and Assessment

To summarise, the systematic literature review findings point to several common characteristics and issues we find important to consider when designing assessment in the online flipped classroom. To start with, assessment needs to be constructively aligned with the intended LOs, as well as teaching and learning activities. Pre-class activities and assessment are often linked to lower levels of LOs, whereas active and independent learning and accompanying assessment are related to higher levels of LOs. Moreover, assessment is usually continuous and multi-dimensional, including both formative and summative tasks, which can be done individually or in teams/groups. It contains feedback given several times in the learning process, whether automatic or provided by peers or teachers. When it comes to assessment of complex tasks, it is important to establish clear assessment criteria, possibly in the form of an analytic rubric, to enhance the reliability and validity of assessment.

Based on that, we are proposing a generic model of flipped teaching, learning, and assessment in online and blended flipped classroom, presented in Figure 4. The model represents the most common flipped classroom learning block, which includes teaching, learning, and assessment activities related to one comprehensive learning topic or one chapter. It lies on four pillars: teaching and learning methods, assessment tasks, assessment methods, and related levels of LOs according to the revised Bloom's taxonomy (Anderson & Krathwohl, 2001). A flipped classroom learning block represents a constituent unit of a course, and such units line up one after another during the teaching and learning process (e.g., throughout a semester). However, the sequence of activities is not universal: different learning paths are possible. For example, some problem-solving task can trigger additional work on introductory, low-level LOs (presented in Figure 4. in form of a reversed arc). LOs of all building blocks contribute to the LOs of the entire course. Assessment is continuous, including both formative and summative tasks.

Figure 4

Generic model of flipped teaching, learning, and assessment
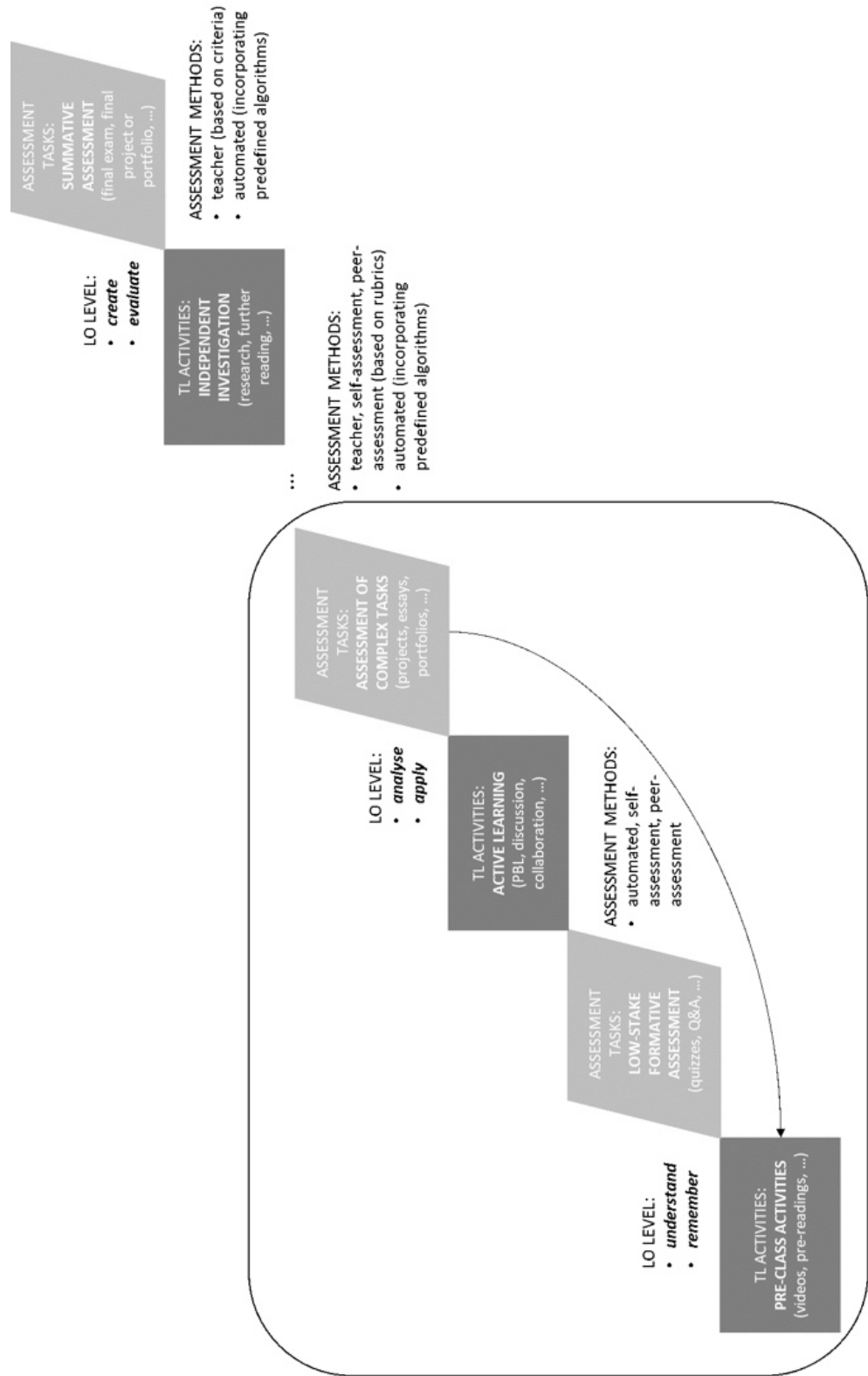
The learning process starts with introductory pre-class learning activities, such as watching videos or reading relevant materials. These activities are usually related to lower levels of LOs, mostly remember and understand. Therefore, they are followed by simple, low-stake formative assessment tasks, such as pre-class and in-class quizzes and/or in-class Q&A sessions, assessing students' understanding of the basic notions and concepts. Assessment methods can include automated feedback provided by online tools, self-assessment or peer-assessment. Instant feedback can help students to better understand their own comprehension of the introductory content and direct their own learning. Moreover, formative assessment results can present an important input for teachers when designing active learning in the classroom, helping them to put more focus on the aspects which are the least clear to their students, but also to create complex, relevant, and authentic learning tasks.

Active learning sessions are based on in-class and/or collaborative work on complex tasks, which might include project-based learning, problem-solving in teams or discussions. These activities are related to higher levels of LOs, mainly apply and analyse. As the basis for assessment of the acquisition of LOs, students provide artefacts, such as essays and reports, projects or portfolios. These artefacts can be assessed using various methods: they can be subject to teachers' assessment, self-assessment or peer-assessment. Providing clear assessment criteria is of utmost importance, possibly in the form of a rubric, which can help structure assessment and feedback. It is also possible to provide automated feedback via online tools, in which case, to support criteria-based assessment, it is possible to incorporate predefined algorithms (as suggested by Divjak & Maretić, 2015). Providing clear criteria is essential in enhancing the reliability and validity of assessment, which can be especially relevant in relation to peer-assessment (Divjak & Maretić, 2017).

The learning process continues with independent investigation, which can include tasks such as further reading, independent research or improvement of artefacts based on the received feedback. In terms of LOs in HE, this is linked to the highest levels, including evaluate and create. It also includes synthesising, as students need to combine previously acquired knowledge to create something new. These activities lead to final summative assessment, which can include, for example, a final exam, a series of monthly exams, or submission of the final version of a project or a portfolio. This final assessment will usually be done by a teacher, according to previously established criteria.

The essence of this model could be applied in various flipped classroom settings, including online, blended, and face-to-face, with some adaptations. However, being based on the findings related to online and blended flipped classroom, it is not entirely invariant to the setting. Most importantly, it incorporates technologies used in online and blended flipped classroom settings, which can provide important added value in comparison to face-to-face environments.

One example of this are videos, as a common element of pre-class activities in a number of analysed flipped classroom approaches. Videos are often provided through LMSs, as an effective way of introducing new content, but can also contain pauses with questions used for formative assessment. Moreover, technology enables quick and simple assessment of students' comprehension via online quizzes taken before or during classes. It also provides valuable possibilities for automated feedback, which can help students reflect on and direct their learning, at the same time reducing teachers' workload. The latter can also be supported by developing

databases of assessment tasks and questions exchanged by teachers (Lopes et al., 2019). Furthermore, online tools support collaborative work in groups or teams, for example, by splitting students in an LMS or using Wikis, blogs or forums, which may sometimes be more convenient in comparison to face-to-face environments, due to various organisational restrictions. Another example is easier organisation, sharing, and management of artefacts by using e-portfolios. Online environments also enable learning analytics, which can support enhancing the reliability and validity of assessment. Finally, online environments can make learning, including assessment, more interesting and stimulating for students, in particular by using gamification.

Although online environments can support teaching, learning and, assessment in flipped classroom in various ways, this also comes with certain limitations (Divjak et al. (2022), for example, related to the user-friendliness of technology, teachers' workload, students' preferences regarding modes of communication etc.

RQ4: Challenges and Recommendations for Future Research

Our analysis has revealed that the majority of the relevant research has been published in recent years, and assessment in online and blended flipped classroom was rarely in focus. To make more general conclusions, further research focused specifically on this issue is needed.

It is important to analyse assessment programmes in online and blended flipped classroom from the perspective of their utility, especially validity and reliability, which has been recognized as critical for assessment in general. This is often implicitly considered as fulfilled, including in peer-assessment, where it may be particularly relevant (Alqassab et al., 2023). Peer-assessment is generally perceived as less reliable because of students' bias, but is essential with respect to all assessment methods. Further research should also consider that, if assessment validity is not ensured, grades cannot reflect the actual acquisition of LOs. In that context, further studies may take into account that not all LOs are equally important and their prioritisation should be reflected in the assessment programme.

When it comes to peer-assessment, the perceived usefulness of peer-assessment in terms of enhancing learning in the context of flipped classroom should be further explored. The analysed articles do not provide conclusive answers to whether peer-assessment should be anonymous and whether it should affect final grades, when used in different flipped classroom learning blocks, as conceptualised in the proposed model.

Further research could also question how to ensure the right balance between personalisation and independent learning on the one hand, and ensuring the clarity and structure of assessment processes on the other, while at the same time maintaining manageable cognitive load. Here, it is important to consider the constructivist perspective, relying on students' individual learning processes, active involvement and self-regulation (Schunk, 2012), and further investigate the power of formative assessment in enhancing students' success measured by summative assessment.

The use of technologies for assessment in flipped classroom approaches could be further explored. There is generally a need (Brady et al., 2019) for more longitudinal studies on the value and effectiveness of using technology for assessment from the perspectives of all relevant stakeholders. Special attention should be focused on teachers, to better understand the costs

and benefits of implementing e-assessment, respective workload implications and alignment with educational theory. Other stakeholder groups may include students, governments, experts (Gerritsen et al., 2017), and employers (van der Vleuten, 1996). Moreover, the possibilities of gamification would be worth exploring in more depth, from the angle of the proposed model.

As the analysed research was often small-scale, it would be valuable to investigate the use of associated assessment in new contexts, to develop more generalised insights. It is also important to examine the proposed generic model from the perspectives of different study fields and educational traditions, minding cultural differences in learning (Entwistle, 2009, p. 39).

Ethical aspects and avoiding plagiarism in e-assessment should also be explored more deeply, especially with respect to the rapid new developments related to AI. Moreover, more research could address social loafing, related to students putting in less effort when they work on group assignments, compared to when they work individually.

Finally, future research should pay greater and more explicit attention to links with relevant theory. Learning theories present the basis for development of well-grounded teaching and learning practices.

Limitations of This Study

The limitations of this systematic literature review are primarily related to the limitations of the analysed studies. A limitation also stems from the fact that the list of relevant papers was extracted from one database (WoS). Moreover, our analysis included only papers published in English, as the only common language of the members of the international research team.

Conclusions

This study focuses on assessment, as the essential part of learning, in the context of online and blended flipped classroom approaches in higher education. We report on the results of a systematic literature review which identified the existing assessment practices and related challenges. The systematic literature review also revealed important guiding principles for assessment, such as constructive alignment, ensuring continuous, both formative and summative assessment, supporting their validity and reliability with clear assessment criteria and rubrics, and directing learning with appropriate feedback. Based on that, we proposed a generic model of flipped teaching, learning, and assessment in online and blended flipped classroom, which could be used and modified depending on the educational context. The model envisages assessment in three phases of flipped classroom, and it includes formative assessment related to basic tasks and lower levels of learning outcomes, followed by either formative or summative assessment of complex task based on criteria, and (final) summative assessment, related to higher levels of learning outcomes. The fact that most of the papers included in the systematic literature review were published in the past several years points to the fact that assessment in online and blended flipped classroom is a propulsive area, which requires further research. Therefore, we provided recommendations for future research to focus on, for example, the utility of assessment, challenges of using technology, avoiding plagiarism, and establishing the balance between flexibility and structure of assessment.

Conflict of Interest

Acknowledgements

References

References marked with an asterisk indicate studies included in the systematic literature review.

Alqassab, M., Strijbos, J.-W., Panadero, E., Ruiz, J. F., Warrens, M., & To, J. (2023). A Systematic Review of Peer Assessment Design Elements. *Educational Psychology Review, 35*(1), 18. https://doi.org/10.1007/s10648-023-09723-7

American Educational Research Association, American Psychological Association, & the N. C. on M. in E. (2014). *Standards for Educational & Psychological Testing.* https://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition

Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). Longman.

*Arruabarrena, R., Sánchez, A., Blanco, J. M., Vadillo, J. A., & Usandizaga, I. (2019). Integration of good practices of active methodologies with the reuse of student-generated content. *International Journal of Educational Technology in Higher Education, 16*(1), 10. https://doi.org/10.1186/s41239-019-0140-7

Bennett, S., Dawson, P., Bearman, M., Molloy, E., & Boud, D. (2017). How technology shapes assessment design: Findings from a study of university teachers. *British Journal of Educational Technology, 48*(2), 672–682. https://doi.org/10.1111/BJET.12439

Biggs, J. (1999). What the Student Does: teaching for enhanced learning. *Higher Education Research & Development, 18*(1), 57–75. https://doi.org/10.1080/0729436990180105

*Blau, I., & Shamir-Inbal, T. (2017). Re-designed flipped learning model in an academic course: The role of co-creation and co-regulation. *Computers & Education, 115*, 69–81. https://doi.org/10.1016/j.compedu.2017.07.014

Brame, C. (2013). Flipping the classroom. *Vanderbilt University Center for Teaching*. Retrieved December 2022 from: http://cft.vanderbilt.edu/guides-sub-pages/flipping-the-classroom/

*Brady, M., Devitt, A., & Kiersey, R. A. (2019). Academic staff perspectives on technology for assessment (TfA) in higher education: A systematic literature review. *British Journal of Educational Technology, 50*(6), 3080–3098. https://doi.org/10.1111/bjet.12742

*Bye, R. T. (2017). The Teacher as a Facilitator for Learning - Flipped Classroom in a Master's Course on Artificial Intelligence. *Proceedings of the 9th International Conference on Computer Supported Education,* 184–195. https://doi.org/10.5220/0006378601840195

*Chien, C.-F., & Hsieh, L.-H. C. (2018). Exploring University Students' Achievement, Motivation, and Receptivity of Flipped Learning in an Engineering Mathematics Course. *International Journal of Online Pedagogy and Course Design, 8*(4), 22–37. https://doi.org/10.4018/IJOPCD.2018100102

Council of the European Union. (2020). Council conclusions on countering the COVID-19 crisis in education and training (2020/C 212 I/03). *Official Journal of the European Union* (C 212 I/9), 9–14.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: qualitative, quantitative, and mixed methods approaches* (5th ed.) SAGE.

Divjak, B., Kadoić, N., & Žugec, B. (2021). The Use of Decision-Making Methods to Ensure Assessment Validity. *2021 IEEE Technology & Engineering Management Conference - Europe* (TEMSCON-EUR), 1–6. https://doi.org/10.1109/TEMSCON-EUR52034.2021.9488580

Divjak, B., & Maretić, M. (2015). Geometry for Learning Analytics. *Scientific and Professional Information Journal of Croatian Society for Constructive Geometry and Computer Graphics* (KoG), 48-56.

Divjak, B., & Maretić, M. (2017). Learning Analytics for Peer-assessment. *Journal of Information and Organizational Sciences*, *41*(1), 21–34. https://doi.org/10.31341/jios.41.1.2

*Divjak, B., Rienties, B., Iniesto, F., Vondra, P. & Žižak, M. (2022). Flipped Classrooms in Higher Education during the COVID-19 Pandemic: Findings and Future Research Recommendations. *International Journal of Educational Technology in Higher Education 19*, 9. https://doi.org/10.1186/s41239-021-00316-4

Dixson, D. D., & Worrell, F. C. (2016). Formative and Summative Assessment in the Classroom. *Theory Into Practice, 55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

Entwistle, N. (2009). *Teaching for Understanding at University: Deep Approaches and Distinctive Ways of Thinking* (1st ed.). Palgrave Macmillan.

Entwistle, N., & Ramsden, P. (2015). *Understanding Student Learning* (Routledge Revivals). Routledge. https://doi.org/10.4324/9781315718637

Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 26(2), 43-71. https://doi.org/10.1002/piq.21143

*Estriegana, R., Medina-Merodio, J.-A., & Barchino, R. (2019). Analysis of competence acquisition in a flipped classroom approach. *Computer Applications in Engineering Education, 27*(1), 49–64. https://doi.org/10.1002/cae.22056

Gerritsen-van Leeuwenkamp, K. J.; Joosten-ten Brinke, D.; Kester, L. (2017). Assessment quality in tertiary education: An integrative literature review. *Studies in Educational Evaluation, 55*, 94-116. https://doi.org/10.1016/j.stueduc.2017.08.001

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

*Jensen, J. L., Holt, E. A., Sowards, J. B., Heath Ogden, T., & West, R. E. (2018). Investigating Strategies for Pre-Class Content Learning in a Flipped Classroom. *Journal of Science Education and Technology, 27*(6), 523–535. https://doi.org/10.1007/s10956-018-9740-6

*Klegeris, A. (2021). Mixed-mode instruction using active learning in small teams improves generic problem-solving skills of university students. *Journal of Further and Higher Education, 45*(7), 871–885. https://doi.org/10.1080/0309877X.2020.1826036

*Lin, C.-J. (2019). An online peer assessment approach to supporting mind-mapping flipped learning activities for college English writing courses. *Journal of Computers in Education, 6*(3), 385–415. https://doi.org/10.1007/s40692-019-00144-6

*Llamas-Nistal, M., Mikic-Fonte, F. A., Caeiro-Rodriguez, M., & Liz-Dominguez, M. (2019). Supporting Intensive Continuous Assessment With BeA in a Flipped Classroom Experience. *IEEE Access, 7*, 150022–150036. https://doi.org/10.1109/ACCESS.2019.2946908

*Lopes, A. P., Soler, M., Caña, R., Cortés, L., Bentabol, M., Bentabol, A., Muñoz, M. D. M., Esteban, A., & Luna, M. (2019). *Gamification In Education and Active Methodologies at Higher Education.* 1633–1640. https://doi.org/10.21125/edulearn.2019.0480

Lundin, M., Bergviken Rensfeldt, A., Hillman, T., Lantz-Andersson, A., & Peterson, L. (2018). Higher education dominance and siloed knowledge: a systematic review of flipped classroom research. *International Journal of Educational Technology in Higher Education, 15*, 20. https://doi.org/10.1186/s41239-018-0101-6

*Luth-Hanssen, V. M., Raaheim, A. R., Sorensen, E. K., & Olstad, K. O. (2018). Digital Assessment in Higher Vocational Education in Norway. Students as Partners. *10th International Conference on Education and New Learning Technologies*, 5903-5909. https://doi.org/10.21125/edulearn.2018.1417

*Mikic-Fonte, F., Llamas-Nistal, M., Caeiro-Rodriguez, M., & Liz-Dominguez, M. (2020). A Gamification Module for BeA Platform. *2020 IEEE Frontiers in Education Conference (FIE),* 1–5. https://doi.org/10.1109/FIE44824.2020.9274180

*Mojtahedi, M., Kamardeen, I., Rahmat, H., & Ryan, C. (2020). Flipped Classroom Model for Enhancing Student Learning in Construction Education. *Journal of Civil Engineering Education, 146*(2), 05019001. https://doi.org/10.1061/(ASCE)EI.2643-9115.0000004

*Moreno-Ruiz, L., Castellanos-Nieves, D., Braileanu, B.P., González-González, E.J., Sánchez-De La Rosa, J.L., Groenwald, C.L.O. & González-González, C.S. (2019). Combining Flipped Classroom, Project-Based Learning, and Formative Assessment Strategies in Engineering Studies. *International Journal of Engineering Education, 35*(6(A)), 1673-1683.

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ,* 372:n71. http://doi.org/10.1136/bmj.n71

Ramsden, P. (2003). *Learning to Teach in Higher Education* (2nd ed.). Routledge. https://doi.org/10.4324/9780203507711

*Ruiz-Jiménez, M. C., Martínez-Jiménez, R., Licerán-Gutiérrez, A., García-Martí, E. (2022). Students' attitude: Key to understanding the improvement of their academic RESULTS in a flipped classroom environment. *The International Journal of Management Education, 20*(2). https://doi.org/10.1016/j.ijme.2022.100635.

*Santos, H. (2021). COVID-19 Lockdown Effects on Student Grades of a University Engineering Course: A Psychometric Study. *IEEE Transactions on Education*. https://doi.org/10.1109/TE.2021.3131745.

Schunk, D. H. (2012). *Learning Theories, an Educational Perspective* (6th ed.). Pearson Education Inc.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1). Retrieved from http://www.itdl.org/

Strelan, P., Osborn, A., & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review (Vol. 30).* https://doi.org/10.1016/j.edurev.2020.100314

Talbert, R. (August 10, 2015). Four Assessment Strategies for the Flipped Learning Environment. *Faculty Focus.* https://www.facultyfocus.com/articles/blended-flipped-learning/four-assessment-strategies-for-the-flipped-learning-environment/

Talbert, R. (2017). *Flipped Learning - A Guide for Higher Education Faculty*. Routledge.

Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: opportunities, challenges and risks. *British Educational Research Journal, 42*(3), 454–476. https://doi.org/10.1002/berj.3215

van der Vleuten, C. P. M. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education, 1*(1), 41–67. https://doi.org/10.1007/BF00596229

*Wang, F. H. (2017). An exploration of online behaviour engagement and achievement in flipped classroom supported by learning management system. *Computers & Education, 114*, 79–91. https://doi.org/10.1016/j.compedu.2017.06.012

*Wang, F. H. (2019). On the relationships between behaviors and achievement in technology-mediated flipped classrooms: A two-phase online behavioral PLS-SEM model. *Computers & Education, 142*, 103653. https://doi.org/10.1016/j.compedu.2019.103653

*Wanner, T., & Palmer, E. (2015). Personalising learning: Exploring student and teacher perceptions about flexible learning and assessment in a flipped university course. *Computers & Education, 88*, 354–369. https://doi.org/10.1016/j.compedu.2015.07.008

*Yilmaz, R., Öztürk, T., & Karaoğlan Yilmaz, F. G. (2020). Examining University Students Acceptance of Web-based Formative Assessment System. *Hacettepe University Journal of Education*, 1–12. https://doi.org/10.16986/HUJE.2020063671

*Zainuddin, Z., Farida, R., Keumala, C. M., Kurniawan, R., & Iskandar, H. (2021). Synchronous online flip learning with formative gamification quiz: instruction during COVID-19. *Interactive Technology and Smart Education, 19*(2) 236-259. https://doi.org/10.1108/ITSE-01-2021-0002

Appendix

Author(s) & year	Context	Field & level of HE	Research design & sample (if applicable)	Findings related to assessment practices in online and blended flipped classroom approaches
Arruabarrena et al., 2019	Spain, France	Computer Engineering Undergraduate, Graduate	Quantitative Case study, survey n/a	Peer-review was used to assess student-generated content (e.g., videos, questions). A non-anonymous form of peer-review was used. Authors stressed that peer-review should be based on well-defined criteria or rubrics. They did not recommend peer-assessment to affect the final grade.
Blau & Shamir-Inbal, 2017	Israel	Education Graduate	Qualitative Reflection analysis 36 students	A re-designed flipped classroom model put an emphasis on "technology-enhanced embedded assessment", combining individual reflection and peer feedback. Each course assignment had evaluation criteria for every section, used in self-assessment. The approach also focused on collaboration and students' co-creation of course content and LOs to develop regulation strategies for individual learning and teamwork.
Brady et al., 2019	Global	Educational Research	Systematic literature review 65 papers	The systematic literature review explored academic staff experiences and perceptions of adopting technology for assessment OF (summative) /FOR (formative) /AS (formative) learning. It was found that not many papers talked about online assessment taking into consideration staff perspectives. It was pointed out that technology use in assessment is not consistent among academic staff, although its potential is growing; technology is most often adopted for formative, low-stakes assessment rather than summative assessment. More structural work is needed to better understand the costs/benefits of implementing online/blended assessment and the respective workload implications, as well as better alignment with educational theory. It was found that a number of papers discussed a need for institutional leadership and support on the issues of plagiarism and cheating.
Bye, 2017	Norway	Computer Science Graduate	Qualitative Case study 20 students	Learning activities and assessment were aligned with intended LOs. Assessment included mandatory assignments, consisting of theoretical short-answer questions and project-like programming exercises. However, the final grade was determined solely based on a final individual oral exam. Digital self-tests allowed for immediate feedback and could contain ambiguously phrased questions, triggering deep learning. Gamification was used to motivate students. Focus on problem-solving was found to be the major key to the success of the course.
Chien & Hsieh, 2018	Taiwan	Engineering Mathematics Undergraduate	Mixed Experimental, survey, interview 60 students	The final exam of the prerequisite course was considered as a pre-test. One formative and two summative assessments during the semester were taken as post-tests. Students' learning achievement (formative and summative assessment) and motivation (intrinsic value, self-efficacy, and self-regulation) were notably better in flipped classroom than in the traditional setting.

Divjak et al., 2022	Global	Educational Research	Systematic literature review 205 papers, 18 in detail	The systematic literature review explored the implementation of online flipped classroom in HE during the COVID-19 pandemic and provides future research recommendations. It points out that assessment in online flipped classroom is an open research question, especially when combined with reliability concerns about peer-assessment. Further research is essential, especially on ensuring the reliability and validity of online assessment in flipped classroom approaches, and forms of assessment and institutional action on academic integrity that can positively impact preventing cheating.
Estriegana et al., 2018	Spain	Computer Engineering Undergraduate	Quantitative Survey 285 students	Assessment consisted of periodical continuous evaluation tests, which included theory and problem-solving, and laboratory practices conducted throughout the term. Game-based learning techniques and activities with mobile applications, such as Socrative or Kahoot, were applied to increase motivation and improve student participation.
Jensen et al., 2018	USA	Biology Undergraduate	Quantitative Experimental 657 students	Following each pre-class activity (interactive tutorials, video lectures, book readings), students took *Explore Assessment* – short online quizzes at low levels of Bloom's taxonomy (mostly *Remember* and *Understand*). Every two weeks, they took *Apply Assessments*, at a variety of Bloom's levels (mostly *Apply* and above). Comprehensive final (summative) assessment was taken in class or a proctored testing center.
Klegeris, 2021	Canada	Biochemistry Undergraduate	Quantitative Action research 89 students	In the course with a significant PBL (clinical case) and teamwork component, two types of assessment significantly contributed to the final grade. End-of-block examinations were taken three times per course, in the form of short-answer question quizzes. Peer marks were given anonymously, based on peers' participation, preparation and professionalism, at the end of each block, using an online tool (iPeer).
Lin, 2019	Taiwan	Foreign Languages Undergraduate	Mixed Experimental 57 students	Students learning with online peer-assessment showed no significantly better learning achievement in comparison to those with conventional peer-assessment. Nevertheless, they showed a significantly higher time involvement in flipped learning, as well as higher autonomy related to previewing the content before class.
Llamas-Nistal et al., 2019	Spain	Computer Engineering Undergraduate	Quantitative Experimental 250 students	A program based on the combination of flipped classroom and intensive continuous assessment. An online tool, Blended e-Assessment (BeA), was successfully used to facilitate assessment. Students took exams on paper, but assessment done by teachers, reporting on results, and review of assessment by students were done online (BeA), using scanned copies of the papers.
Lopes et al., 2019	Portugal, Spain	Business and administration, Chemistry Undergraduate	Quantitative Survey 3000 students	The blended learning approach included online tools and gamification (Socrative, Kahoot) in conjunction with flipped classroom. Gamification (performed through Kahoot) was used for formative assessment (part of continuous assessment) and summative assessment. It increased students' motivation for active learning.

Luth-Hanssen et al., 2018	Norway	Electro-engineering Vocational college	Qualitative Case study n/a	Assessment was aligned with teaching and learning activities and LOs. Students were included in establishing assessment criteria, which were used in peer-assessment, as well as by teachers. Portfolios were used for reports and assignments which were formatively assessed by both peers and teachers. Portfolios were subject to final summative assessment.
Mikic-Fonte et al., 2020	Spain	Computer Engineering Undergraduate	Qualitative Case study n/a	The BeA platform was used to combine flipped classroom and gamification. Intense continuous assessment was performed in the classroom once every two weeks, covering the corresponding weeks' contents. Gamification was used to motivate students to watch videos in the weeks when students were not assessed.
Mojtahedi et al., 2020	Australia	Construction Engineering Undergraduate	Mixed Case study, survey 60 students	The assessment program consisted of pre-class online quizzes, interactive tutorial discussion sessions (teamwork), in-class online quizzes covering the pre-class activities, peer-assessment based on a rubric, final assessment and a case-based assignment. Pre-class online quizzes were perceived by students as the most effective learning activity, whereas peer-assessment was the least appreciated.
Moreno-Ruiz et al., 2019	Spain, Brazil	Computer Engineering Undergraduate, Graduate	Quantitative Survey 72 students	An online tool (SIENA) was used, allowing for self-directed learning and self-assessment, and computerized adaptive testing. Continuous assessment included online tests, analysis of messages exchanged in an online forum and assessment of final projects. Students received feedback on each of the tests. Rubrics were used for continuous assessment.
Ruiz-Jimenez et al., 2022	Spain	Business and administration Undergraduate	Quantitative Survey 107 students	The study was focused the effect that students' attitudes and formative assessment have on their perception of learning outcomes in flipped classroom. Students' perception of their final assessment results depended positively on their attitude towards learning and on formative assessment. It was shown that continuous assessment and feedback from formative assessment enabled students to learn more.
Santos, 2021	Spain	Computer Engineering Undergraduate	Mixed Case study, experimental 74 students	The study found that the reliability of online student assessment in flipped classroom is high. The different level of item difficulty in the f2f block may be a cause for low test reliability. Student surveys and student performance suggested that online flipped classroom can be suitable for this specific course.
Wang, 2017	Taiwan	Computer Science Undergraduate	Quantitative Deep learning 488 students	A behavior model was developed that depicted the effects of online engagement on achievement. It was found that engagement in problem-solving had direct, positive effects on formative assessment and the final achievement. Introducing self-reflection and self-assessment activities into flipped classroom helps entice behavioral engagement in online study and social interaction, which in turn promote engagement in problem-solving activities.

Wang, 2019	Taiwan	Computer Science Undergraduate	Quantitative Deep learning 431 students	Formative assessment comprised classroom performance, homework assignments, quizzes, and midterm exams, whereas summative assessment included final exams and final projects. The final grade was based on a certain proportion of the outcomes of formative summative assessments. It was found that in-class behavioral engagement had a direct positive effect on engagement in self-reflection and self-assessment.
Wanner & Palmer, 2015	Australia	Social Sciences Undergraduate	Mixed Survey, focus groups 96 students, 47 teachers	A flexible assessment program was used, in which students were involved in deciding on the method and timing of assessment. Flexible assessment is changing assessment practices to be more learner-centered and directed and improving student engagement. Even though personalized assessment was appreciated, students still needed guidance and clear structure in their learning. The scope of flexibility and choices for assessment still needs to be aligned with LOs.
Yilmaz et al., 2020	Turkey	Computer Engineering Undergraduate	Quantitative Survey 381 students	A web-based formative assessment system was used to support formative assessment, providing students with instant feedback. Students' acceptance of the system was explored using a scale of acceptance containing the following elements: computer self-efficacy, perceived ease of use, social influence, perceived content, state of enjoyment, state of interest, perceived usefulness and usage intention.
Zainuddin et al., 2021	Indonesia	Public Administration Undergraduate	Mixed Experimental, survey, interviews 73 students	The approach included gamified flip learning, with interactive gamification (game-based e-quizzes, with students creating questions) integrated into the assessment process. It was found that it stimulated the engagement of students in flipped classroom, self-paced learning and critical thinking.