German University of Digital Science

3 SCIENTIFIC REPORT

Assessment Methods for Online Teaching

June 10, 2024

Prof. Dr. Christoph Meinel
German University of Digital Science

Prof. Dr. Mike Friedrichsen
German University of Digital Science

Dr. Thomas Staubitz
German University of Digital Science

Selina Reinhard M.Sc., Daniel Köhler M.Sc.
Hasso Plattner Institute

german-uds.de

Scientific Reports of the
German University of Digital Science | 3

Christoph Meinel | Mike Friedrichsen | Thomas Staubitz
Selina Reinhard | Daniel Köhler

Assessment Methods for Online Teaching

German University of Digital Science

Online teaching does not only require dedicated formats of delivering the learning materials and particular didactic approaches. There are also many specific considerations about assessment forms that have to be taken into account. While in our previous report (SR2)[1] [49] we have focused on specific forms of content delivery and didactic approaches, we now have directed the spotlight on online assessment. We highlight several assessment methods from simple multiple choice tests to automated methods of testing more complex tasks to forms of collaborative assessment, such as project based learning, peer assessment, and team projects. Furthermore, we cover a selection of aspects to ensure academic integrity, such as online proctoring or plagiarism checks.

As described in SR2, we consider the digitalization of education to be a chance rather than a challenge as it improves the availability of education for less privileged learners, learners in remote or rural areas, and life-long and part-time learners.

Operating openHPI, the Hasso Plattner Institute's MOOC platform, for more than 10 years and offering a large number of courses on a variety of topics there, the authors have gained an extensive experience on many aspects of online assessment. Although MOOCs are a form of e-learning in a very special setting, many of the experiences and research results in this context can be transferred and adapted to the more formalized approach of online learning of the German University of Digital Science.

In the report at hand we provide an overview of general considerations about assessments such as assessment levels, assessment principles, and assessment forms. We have a look at grading and cheating, and introduce several forms of online assessment methods. We report on challenges and opportunities in the context of scalability and finally provide an introduction to peer assessment and the assessment of team assignments.

[1] Scientific Report 2 - Innovative Formats for Online Teaching

Contents

1. Introduction

The COVID-19 pandemic has significantly increased the pace of digitalization across diverse spheres and facets of daily existence. In response to the imperative of averting contagion and augmenting public health safety, a pivotal facet of early life — education — has undergone a pronounced transition towards remote modalities across expansive geographical expanses. Analogous to prevailing trends in professional environments, numerous educational establishments encompassing schools, universities, and vocational training centers have embraced digital instruction, at least to some extent. Regrettably, the caliber of education dispensed through online mediums varies markedly contingent upon the degree of preparedness exhibited by instructors. Whereas certain entities, such as select study programs, have adeptly integrated diverse elements into their digital instructional frameworks, others conspicuously lag behind in this regard.

For the last decade, the Hasso Plattner Institute (HPI) has been offering online courses covering a range of subjects through the online education platform known as openHPI[1][50]. Massive Open Online Courses (MOOCs) hosted on openHPI have witnessed a similar upsurge during COVID-19 lockdown periods, as individuals found themselves with limited activities and thus turned to online learning opportunities. Within the framework of online education facilitated by openHPI, we have had the opportunity to explore diverse educational concepts in digital environments. Before moving to the German University of Digital Science (German UDS), the authors have been involved in the development and operation of this platform in leading roles. Building on this experience, with this report, we aim to provide a summary of assessment methods to be applied in online education.

1.1. Problem Statement

A recent article by Ananda Klaar, an ethnology student and author of an op-ed column about universities in the well-known German magazine "Der Spiegel", sported the headline "Wir brauchen neue Prüfungsformen – und zwar schnell!"[2] [102]. The author criticizes that current forms of assessments and exams are causing pressure, fear and sleepless nights. Most of all, the exams do not support learning, and neither prepare for the future tasks in the job nor life in general [102]. The article refers to a report by one of the largest German health insurance companies examining the

[1] `https://open.hpi.de`

[2] We need new forms of assessment - and we need them now!

health of German students[3]. The report describes that 68% of the students are exhausted by stress, often caused by exams. Still many exams rely on the amount of knowledge that a student is able to learn by heart, and generally forgets right after the exam. Grades are not only fostering pressure on the students but also competitive behavior patterns between students, which would be better replaced by collaborative behavior patterns [102]. Furthermore, technological developments, for example, the Internet in every hand via smart phones, and particularly the recent developments in the area of generative AI require fundamentally new approaches towards assessment. While the current situation urgently requires changes in on-campus structures, online study environments are facing similar challenges, while also providing some opportunities for solutions.

1.2. Contribution of this Report

In this report, we provide an overview of assessment forms to be used by educators in various online-education scenarios. To allow for appropriate decisions for, or against, certain assessment forms, we highlight, how well the different forms can be used in combination with the variety of educational methods we have described in a previous report [49].

1.3. Report Structure

The remainder of the report is structured as follows: In Chapter 2 we provide a general definition of assessment and particular forms of assessment in educational environments. Furthermore, we set the scope of the report at hand. In Chapter 3 we introduce a variety of methods which are particularly suitable for assessments in modern online learning contexts. Chapter 4 examines assessment methods particularly focusing on their scalability. In Chapter 5 we highlight the aspect of online proctoring and provide a basic comparison of different forms and providers of online proctoring. Furthermore, we give a broad overview on aspects of plagiarism in the context of online learning. In Chapter 6 we return to a particular form of assessment: peer and self assessment. Chapter 6.2 then shows how peer assessment works particularly well when it is combined with group or team tasks. In Chapter A we provide a (non-exhaustive) list of suitable online assessment tools. Chapter 7 finally summarizes our report.

[3]hhttps://www.tk.de/presse/themen/praevention/gesundheitsstudien/tk-gesundheitsreport-2023-2149758

2. Assessment

Evaluating the performance of learners has always been a difficult endeavor. Educators are tasked with ensuring alignment between assessment and the material covered in instruction, mindful of the inadvertent evaluation of supplementary skills. Many assessment methodologies prioritize rote memorization over the application of acquired principles, consequently fostering what is colloquially termed as "bulimic learning" among students. Moreover, certain assessments may merely demonstrate adeptness in navigating specific examination formats rather than true mastery of the subject matter.

Compounding this complexity are the advent of contemporary generative AI tools, exemplified by platforms like ChatGPT, which challenge traditional examination paradigms. Furthermore, within the realm of online learning, assessing learner performance introduces supplementary obstacles, including heightened demands for identity verification, the implementation of rigorous exam proctoring protocols, and the necessity for a secure academic environment conducive to deterring instances of academic misconduct.

2.1. General Considerations

We will start this report with a set of general considerations on different assessment levels, the general purpose of assessment, and selected attempts to define assessment principles. We will have a look at different forms of assessment, grading, and finally cheating.

2.1.1. Levels of Assessment

Assessment in higher education basically happens on five separate levels [51]. On Level one (L1), the learning[1] of individual students within a course is assessed, on Level two (L2) the learning of students across courses within a program is assessed. Levels three to five are covering the assessment of courses, programs, and institutions. The report at hand is focusing on *L1* and from time to time might scratch at the surface of *L2*. Assessment on *L3* to *L5* is done either by students, e.g. in the context of internal quality assessments or in the context of student rankings of programs and

[1] For simplicity, we only refer to *learning* or *knowledge* for now. We will differentiate between knowledge, skills, and competences in one of the following chapters.

institutions, such as the *CHE*[2], or by public or private accreditation organizations. None of these are part of this report.

2.1.2. The Purpose of Assessment

Different types of assessment serve different purposes. On the one hand, some forms of assessment aim to provide students individual monitoring of their learning process. Examples for this are self-tests that allow students to quickly check their understanding of a certain lesson or context. Self-reflections or summaries allow students to achieve a similar goal with a little more depth. On the other hand, instructors use assessment to monitor the students' performance and decide on measurements to customize their teaching plans for the whole group of students or support individual students who are ahead or behind the bulk of the cohort. Instructors also use assessment to provide individual feedback to students.

To sum it up, assessment is used to help students learn, by helping them determine whether they have comprehended the content, and motivating them to go further or to do better.

In real-world situations, however, assessment is often used to filter students from popular programs and decides on their future careers, nowadays already starting in *Kindergarden* and pre-school, throughout their whole educational career up to graduation from university. Particularly, due to this permanent stress, the topic of assessment goes hand in hand with the topic of cheating. Reducing the weight and significance of single assessments by offering several smaller assessments instead of few large assessments can help to reduce this stress and, therefore, one of the motivations to cheat. We will discuss this in more detail in Section 2.5.

2.1.3. Principles of Assessment

Several frameworks to categorize the principles of assessment exist. We have selected two of them, which although not being exhaustive are covering a sufficiently wide range of aspects. We briefly summarize their key aspects in the following, which are good guidelines when designing assessments.

- Brown [7] has developed a set of principles for the assessment of languages, which can be generalized for any form of assessment.
- *The Taylor Institute for Teaching and Learning* at the University of Calgary [18] has developed a set of guidelines particularly focusing on online assessments.

2.1.3.1. Brown's Principles of Assessment

Although Brown's principles from his book *Language Assessment: Principles and Classroom Practices*, published in 2004 [7], were primarily developed for language assessment, they can be applied to almost any kind of assessment.

[2]Center for Higher Education: https://www.che.de/en/

The main protagonists in assessment are students—those who produce and submit the exams or artefacts to be assessed—and raters—those who assess the submitted work of the students, either by giving them a grade or feedback.

(1) **Practicality** Assessments should not be excessively expensive, should be manageable within appropriate time constraints, and they should be easy to administer. Particularly, they should have a time-efficient grading procedure.

(2) **Reliability** Assessments have to be consistent. The same test given to two comparable students should yield similar results. Reliability comes with a set of different aspects. *Students* have to be reliable, they should not be sick, overly anxious, or burdened with other physical or psychological factors which hamper them to reach their full potential. *Raters* also have to be reliable. Brown speaks of *inter-rater* reliability when two or more raters yield consistent scores for the same test. *Intra-rater* reliability means that one rater yields the same results for the same test regardless of fatigue (having graded the same questions of many students) or any bias caused by previous results or behavior of particular students. Furthermore, it has to be ensured that the way the assessments are *administered* reliably produces the same results. Noise level, distractions, etc. have to be reduced to a minimum. Photocopies or handouts have to be provided with a consistent quality. The same applies for light, desk and chairs, etc. Particularly, in the context of online learning, factors such as online connectivity, Internet speed, or issues with online-proctoring have to be added here. Also, the *Test* itself can be a source of unreliability. Tests that are too long, e.g., do not only test the learners knowledge but also their resilience against fatigue. If they are too strictly timed, they rather measure a student's stress resistance. Ambiguous or poorly designed questions can also contribute a certain amount of unreliability.

(3) **Validity** is another quite complex principle coming with several aspects. A valid assessment tests exactly what is supposed to be tested and not something different, such as vision, previous knowledge, or other skills not exactly relevant for the actual purpose of the test. The assessment should be valid for the *content* of the class: e.g. programming tasks are a more valid assessment for a programming class than multiple choice tests. Other forms of validity next to *content validity* are *concurrent validity*—the results of an assessment should confirm the general performance of a learner as measured, e.g. in class contributions, and *predictive validity*—the assessment captures the likeliness of the learners success, e.g. in a future employment. Furthermore, assessments should be valid against socio-economic advantages, e.g. as some learners might be able to afford additional coaching while others are not or better educated parents can support their offspring in a way that less educated parents cannot. Obviously, this so-called *consequential validity* cannot be designed into an assessment alone, but has to be taken care of in the overall design of classes, e.g. by offering free tutorials led by students from previous semesters, etc. Finally, *face validity* determines the ability of an assignment to measure the knowledge or skill it claims to measure based on the judgement of the learners. A *face valid* assessment

will be provided in an expected and familiar format and doable in the given time frame. It will come with clear directions and a difficulty level that is to be expected by the learners based on previous (ungraded) tests and the way the content was delivered in the lectures.

(4) **Authenticity** simply means that an assessment is related to the real world. It should use natural language, the tasks should be properly set in context, the topics should be relevant to the learner—particular awareness has to be given here to differences in culture, background, and gender.

(5) **Washback** delineates the influence that the assessment has on the teaching. "What is assessed becomes what is valued, which becomes what is taught." [99]

Washback is a phenomenon that occurs particularly with high-stakes assessments. Instructors aim that their students pass e.g. centralized tests with good results and, therefore, adjust their teaching to the design of the assessment.

2.1.3.2. Principles of Assessment - Taylor Institute

The following paragraphs in this section briefly summarize the principles of assessment as defined by the Taylor Institute of Teaching and Learning at the University of Calgary, which have a particular focus on the online learning context [18].

(1) **Focus on learning** Assessments should support the learners to focus on the key elements of a course. Therefore, the learning outcomes of a course should be clearly defined and communicated. The assessments should be closely aligned to these learning outcomes. The strategy to achieve this goal is to employ constructive alignment by iterative revisiting and adjusting the learning material and the assessments to match the defined learning outcomes. Rubrics can be used to communicate to the learners what the instructors expect from them and help the instructors, teaching assistants, or peers to grade consistently and transparent. Particularly in online learning contexts, the students have to use a variety of technologies. If technology is employed in assessments, students should have the opportunity to familiarize themselves with the technology in a non-assessment context, so that during the actual assessment they can focus on the the task and not on the technology.

(2) **Balance structure with flexibility** While online learners need a reliable and predictable structure, they also need a certain level of flexibility. E.g. the learners could be allowed a certain choice in the weight of the individual assessments, they could be allowed to choose if they rather hand in a paper, a recorded presentation or e.g. a software they developed. Furthermore, they could be allowed to decide if they want to work on a certain task alone or in a team, or they could be allowed to select an assessment they want to skip or remove from the grade.

(2) **Provide clear instructions and quality feedback** A clear and well-defined, upfront communication of requirements, technology to be used, purposes, and grading criteria enables learners to focus on the given task, being able to effectively demon-

strate their skills and competencies, without struggling with uncertainties and technological issues.

(3) **Consider alternative forms of assessment** Do not only rely on timed exams. Although these are the easiest and less time consuming approach to create and grade exams, they can only cover the assessment of a subset of the learners skills. Additionally, they require some sort of proctoring to avoid cheating. Alternatives e.g. can be e-portfolios, peer assessments, or custom tools for particular subjects, such as auto-graded coding tasks.

(4) **Use open book exams** In most subjects, learning by heart is no more a skill that meets the needs of neither today nor the future. Knowing where to find information and how to retrieve it, is often a more up-to-date skill. Open book exams, therefore, often are a better way to assess a learner's competencies. These exams can e.g. be combined with competitive time limits, to make sure that learners really have the required competencies in retrieving relevant information. The given time constraints have to be communicated upfront and transparently. Furthermore, learners can be asked to reveal their sources. Well prepared learners, with a good understanding of the subject will always outperform those who lack preparation and understanding. Well implemented open book exams will rather assess the competencies at the higher levels of Bloom's taxonomy, such as analysing, synthesizing, or evaluating information rather than the mere reproduction of knowledge. Open book exams, therefore, also should be equipped with a certain level of support, such as e.g. an FAQ section. Furthermore, the expectations in terms of detail and length should be clearly communicated, e.g. by setting word limits.

(5) **Promote academic integrity** It should always be clearly communicated that academic integrity is expected from the learners and lived by the staff. Assessments should be designed to promote learning, not to avoid cheating. The meaning of academic integrity should be openly addressed in different contexts ways. A very simple form is to ask students to sign or accept an honor code, whenever they start a course or an assessment. On the other end of the spectrum, a dedicated class on academic writing, presentations, etc. could also include aspects on ethics and academic integrity.

2.1.3.3. Others

Other principles of assessment have been published by, e.g. the University of Lancaster[3], or Coursetopia a British provider of vocational training[4]. Although the categories of their principles are not all identical to those that have been described in more detail above, the underlying concepts are basically quite similar. The principles by the University of Lancaster e.g. state that the focus of the assessment should cover

[3] https://www.lancaster.ac.uk/curriculum-and-education-development-academy/resources/assessment-principles/

[4] https://www.coursetopia.co.uk/principles-of-assessment/

what was discussed in the program, they should aim for holistic understanding, facilitate learning and improve standards. Furthermore, assessments and feedback should be timely and embedded in an overall learning journey. Coursetopia stays even closer to Brown's principles but lists eight instead of five: ethics, fairness, sufficiency, currency, authenticity, reliability, validity, and learning domains.

2.1.4. Steps in the Assessment Process

To properly assess the learning of the students, it is important to clearly define and communicate the expected learning outcomes and the expected depth of learning. According to these definitions of expected learning, the appropriate measures to assess the learning outcomes have to be chosen. The collected results need then to be analyzed and, finally, appropriate measures to improve the situation if necessary have to be implemented.

2.2. What to Assess?

On of the most important questions that instructors have to ask themselves when they are creating assessments is what exactly they want to assess. This question is very closely related to the question what exactly they intended to teach. An, unfortunately not yet common, approach is to follow a strategy that is quite similar to *test-driven development* in the area of software engineering. Before the actual software is developed, the code to test if the software actually does what it is supposed to do is developed. Following that approach and creating tests and teaching/learning material side-by-side, it is much more likely to create assessments that actually match what has been learned than pushing the development of assessments to the end of the class, creating it right-in-time when it is delivered to the learners. While it is comparatively simple to meet the course's topics in the assessments, it is far more difficult to meet the level of expectations. Mapped to Bloom's taxonomy, the question might be if the learners are supposed just to remember some facts, understand their meaning, apply this knowledge, analyze or evaluate them or create something on their own.

To give an example: A history lesson dealt with the historic figure *Vlad the 3rd of Wallachia*. A question that just asks for remembering the facts would be "What was his preferred method of killing people?", a question that asks to apply this knowledge would be how exactly that influenced Bram Stoker in writing his most famous novel. A question asking the learner to write a short story would be on the creation level, while writing a critique of Stoker's story comparing it to the historical facts might be the evaluation task.

The question can be reformulated to: "Do we want to assess knowledge, behaviors, skills, or competencies?" Assessing the pure remembering of facts has been proven to result in so-called bulimic learning. Students start learning right before an exam stuffing the expected knowledge into their brains, can re-call the knowledge for a few days and as soon as the exam has passed extinguish the knowledge from their

brains as quick as possible. The worst example in that direction was encountered by one of the authors several years ago on an online learning platform where each quiz question started with the sentence "What did the professor say in the previous video about x or y?"

Next to the question on which level of, e.g. Bloom's taxonomy, factual knowledge, skills, and competencies are to be addressed, it has to be decided if further professional or so-called future skills, such as team work, proficiency in communication technology, problem-solving skills, critical reflection skills, conceptual understanding, abstraction or reasoning skills, etc. have to be assessed. Next to those also classical skills, such as writing or language skills might add to the picture.

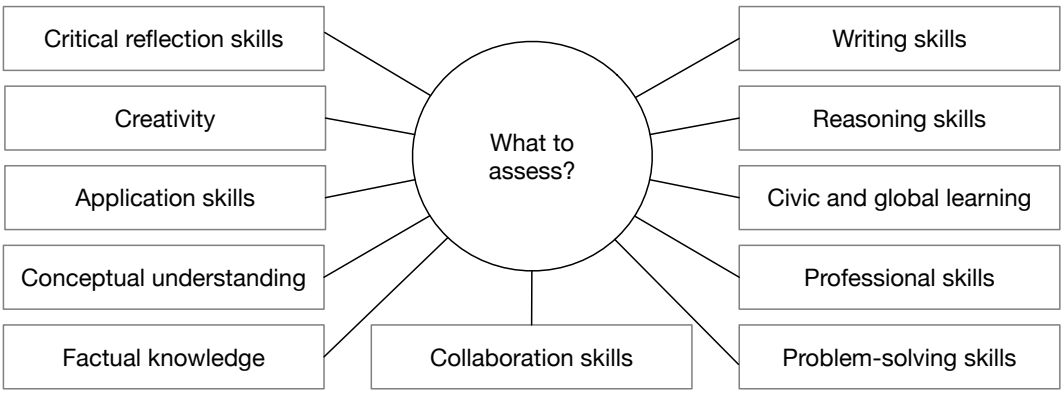

Figure 2.1.: Several aspects that can be covered by different forms of assessments. (Based on [12])

2.3. Formative vs. Summative Assessment

The two essentially different types of assessment are *formative* and *summative* assessment. In brief, summative assessment is measurable and is mostly expressed in grades or similar, while formative assessment helps instructors and learners to adjust the learning process and ideally is expressed in feedback. In other words, formative assessment is assessment **for** learning, while summative assessment is assessment **of** learning [11].

2.3.1. Summative Assessment

Summative assessment (ideally) reflects students' knowledge of a given topic through a grade. Generally, summative assessments are conducted in the form of tests. These tests have to be created in a way that they do not only test the remembering of factual knowledge, but in a way that the students have to cover all levels of Bloom's taxonomy to succeed [49]. Typical examples are final tests at the end of a semester

or quarter, midterm exams, final or capstone projects, papers, oral exams, or presentations. Typically, students are asked to work on such assessments individually and without the help of other people, technological devices, or other resources such as cheat sheets. As these assessments often come with high stakes, cheating and the prevention of cheating is a topic to be addressed in detail. However, summative assessment can also be applied on group work. In this case, a mechanism has to be established that allows the instructor to combine a grade from the group's collaborative work and the individual contributions of the participants.

Information that is extracted from summative assessments can additionally be used in a formative way, particularly by instructors, who always have the chance to improve their teaching in the next iteration of their course. In case that there are mechanisms in place that allow students to learn from their mistakes and potentially even improve their grades in a second attempt or an additional exam, they can also use the results from the summative assessment in a formative way. A very basic example for this is the practice of many school-teachers to ask their students to return a corrected version of their exam once they have received the grade.

On the positive side, summative assessments show if the students have understood a certain topic or concept, they can boost individual confidence of (strong) students, they help to evaluate students' performance and make academic record. They help instructors to identify weak areas and measure the success of their training methods. To some extent, they can also be used by the institution to measure an educators performance.

On the negative side, summative assessments put pressure on the students and lessen the confidence of (weak) students. They often rather assess the ability of students to pass a certain form of tests than their actual knowledge. Furthermore, they are often restricted to the recall of certain information given in a lecture, rather than testing the ability of the students to apply their knowledge. The rise of standardized tests throughout recent years has to be viewed quite critically as they often measure "the average", while excellence in certain areas can easily be missed as they do not take individual needs and capabilities into account. The same exam for all does not necessarily result in a fair[5] selection.

2.3.2. Formative Assessment

Formative assessment provides instructors with the necessary data to adjust the learning process, ideally "on the fly" or at least for the next generation of learners. It enables the instructors to provide on-going, focused, and targeted feedback and, thus, provides learners with the necessary insight to adjust their learning process, ideally also "on the fly" so that they can pass the class or improve their grades. It enables students to take an active role in their learning process by monitoring their

[5]Even the word *fair* itself is defined quite differently. Is it more fair giving everybody the same amount of time and tasks with the same level of difficulty, or is it more fair giving e.g. persons with a disability more time or a different set of tasks that takes the conditions of the disability into account)

progress and using feedback to make adjustments. [11, 20] It empowers students to become independent learners.

Examples for formative assessments are any ungraded class activities starting from ungraded diagnostic quizzes, asking the learners to create concept maps, collecting sentences that identify the main point of a lecture, e.g. in the form of a word cloud, portfolios, journals, presentations, discussion boards to group work on given tasks or worksheets or even longer projects.

On the positive side, formative assessment improves teaching and learning and helps students and instructors to identify strengths and weaknesses and target areas that need work, while adjustments are still possible. Formative assessment is generally low stakes, helps to identify and communicate learning goals, increases rigor and improves academic achievement. Ideally, it helps to enhance student motivation and engagement by personalizing their learning experience, therefore, producing self-regulated learners.

Rubrics can be employed as a guide for scoring and evaluating the quality of the learners' responses. Usually, rubrics include criteria for evaluation, quality definitions for those criteria at particular levels of achievement, and a scoring strategy. Rubrics are often presented in the form of tables and are used by instructors to mark the results and provide feedback, and by learners when planning their work. Rubrics, when used with formative assessment purposes, have shown to have a positive impact on students' learning. Rubrics provide personalized feedback while allowing for students to take ownership of their progress as well as pointing areas of improvement.

	Formative assessment	Summative assessment
	Learn and practice	Assess performance
When	Throughout the course	At the end of the instructional period
Why	Identify gaps and improve learning	Collect evidence of student knowledge or skills
How	Via approaches that support students' needs	Via learning products or a cumulative assessment

Figure 2.2.: Summative vs. formative assessment. (Based on [20])

2.4. Grading

Particularly in the context of summative assessment, grading is an important aspect to be considered. The two main angles to be considered in the context of grading are

- Who is grading?
- Which criteria are employed for grading?

Considering the first question, grading can be done either by machines or by humans. Humans who are grading the work of other (hopefully) humans, can be instructors[6], externals doing this for a fee, or peers. We will dive deeper into these approaches in Section 4. The second question of grading criteria goes hand in hand with the questions introduced in the previous section of what exactly is to be assessed. As this differs significantly from course to course and environment to environment, there is no *one-size-fits-all* solution.

2.5. Cheating

Particularly, but not exclusively, in scalable online learning contexts it has to be clearly defined before an exam is started, what is considered to be cheating. Automated proctoring solutions, e.g., will have to be explicitly configured which behaviors they are supposed to treat as cheating attempts. Once they are configured that way, they will mercilessly punish every participant that is not behaving conforming to these rules. Therefore, it is even more important to clearly communicate the definition of cheating before the participants are starting an exam as it is in a classic classroom setting. Obviously, the rules should be communicated there as well, but the proctors in a classroom have at least some flexibility in minor cases and can work with several layers of warning before ripping the participant's exam sheet to pieces. Automated proctors might allow a certain threshold of misdemeanors but generally they do not provide warnings and in the end it is always a binary decision.

The example already shows that there is actually no difference between online and offline exam settings as human proctors in human based online proctoring tools have the same options as human proctors in a classroom. The differences are rather defined by the number of participants, as for scalable online exams automated proctoring solutions are required. In a similar way, given that modern proctoring tools are applied, there are more differences between a classroom exam with very few students and a classroom exam with hundreds of students, than between similarly sized online and offline exams. With the ubiquitous use of mobile phones, students have very similar opportunities for finding information in each of the settings. Preparing cheat sheets and hiding them in the restrooms appears to be a quite anachronistic view in the light of these developments.

[6] Teachers, tutors, assistants, etc.

2.5.1. Do successful cheaters prove other skills?

In some cases, successful cheaters might actually demonstrate other skills than the ones that are examined. As an example, in some Javascript-based online exams, the participants might be able to find the solution in the exams source code. This might be more or less well hidden. Finding the solution by reading the source code could be considered as cheating, or it could be considered as a demonstration of different skills. The question remains if these skills can be measured, if they should be awarded, e.g. by ignoring them as cheating attempts, or if even in such cases a rigorous punishment is required. These questions are hard to answer in general and will be regarded differently by institutions and individuals.

2.5.2. Reasons for cheating

Participants are inclined to cheat in exams for a variety of reasons. Some of them might just have been lazy and lack a proper preparation. Others are tempted to cheat as the opportunity is given. Particularly in online exams, information is easily accessible on the Internet and there might be a lack of proctoring or supervision. Other common reasons are the pressure that is put on the participants due to the weight of an exam and the resulting fear of failing. Designing courses so that they contain a variety of different, smaller exams appears to be a promising strategy to reduce the motivation for cheating. The effort of preparing for each of the smaller exams is reduced, which can reduce the aversion against starting to prepare for the exam. Furthermore, the weight of each of the exams within the courses overall grading framework is reduced, which can help to lower the pressure of success and the fear of failing.

2.5.3. Why is cheating such a big concern?

While some authors argue that cheating is a particular problem in online settings, we argue that cheating is just as problematic in regular classroom settings. Given that appropriate anti-cheating tools are employed, online exams are even easier to scale than classroom exams with the same probability to detect cheating attempts. Common forms of cheating in online exams, such as sharing answers with fellow learners, accessing illegal resources, using technology, and impersonating other students are also known in classroom settings, particularly if they are large scale.

Also the consequences for the institutions are similar. Confidence and trust of employers in the degrees and certificates are undermined. The reputation of the institution suffers and it will get increasingly hard to attract students.

A particular form of cheating is to employ agencies to write articles, papers, or theses. In this context, another aspect has to be considered. Students who have employed such agencies to cheat are increasingly blackmailed by them when they got away with their cheating and enter high level positions [96].

The National Business Ethics Survey of the U.S. Workforce Report from 2013 revealed that there is a particular high percentage of ethical misconduct in the ranks of

the management. Although the report does not draw the direct connection, there is a certain probability that particularly students who learned that they will get away with cheating during their studies, are likely to continue following this unrighteous path at their future jobs [95].

2.5.4. How to prevent cheating?

Several approaches and technologies exist to detect and prevent cheating particularly in online settings. We introduce selected proctoring solutions in Section 5.1 and selected solutions to detect plagiarism in Section 5.2.1.

3. Online Assessment

The formats of online assessments are varied. Multiple-choice questions (MCQs), in which students must select the correct answer from several options, are often used. Short answer and essay questions, which require longer and more detailed responses, are also common. Technical and IT courses often use programming tasks that require students to write and run code. Project work, which is completed over an extended period of time and often involves presentations or reports, is another common assessment method. Institutions use a variety of platforms and tools to conduct these assessments. Learning Management Systems (LMS) such as Moodle, Blackboard or Canvas offer extensive functionality for conducting online exams. In addition, specialized proctoring software, such as ExamSoft, ProctorU, or Respondus, has been developed specifically for the secure and proctored delivery of online exams [53].

The benefits of online assessments are many and diverse. They provide students with flexibility by allowing them to take exams from multiple locations. Automated scoring of multiple choice and other objective test formats saves time and enables efficient delivery. Large groups of students can be assessed simultaneously, increasing the scalability of these assessments. In addition, online assessments allow the use of multiple assessment methods that combine different question formats and interactive content. Despite these advantages, there are challenges and drawbacks. Technical issues, such as Internet connection failures or software incompatibilities, can affect exam performance. Exam security must be maintained to prevent cheating and ensure student identity. Privacy is also an important issue, as students' personal information and exam content must be protected. Not all students may have equal access to the necessary technical resources, which can affect equity [4].

Proctoring of online assessments can be done in a number of ways. Live proctoring involves real-time monitoring by proctors using a webcam and microphone. Automated proctoring uses software-based monitoring with algorithms to detect suspicious activity. A combination of automated and human proctoring is often used to ensure the security and integrity of exams. Overall, online assessments offer universities a modern and flexible way to evaluate student knowledge and skills. However, their successful implementation requires careful planning and measures to ensure the integrity and fairness of the assessments. Online assessment has become an integral part of higher education, driven by advances in technology and the need for flexible learning environments [28]. We provide more details on online proctoring and a comparison of proctoring solutions and providers in Section 5.1.

The remainder of this chapter examines the typical online assessments used in higher education, discussing their formats, benefits, challenges and best practices for effective implementation.

3.1. Online Assessments in the Higher Education Context

Online quizzes and tests are among the most common forms of online assessment. They can include multiple choice, true/false, short answer and essay questions. They are often used to assess a wide range of knowledge and cognitive skills. One of the main benefits of online quizzes and tests is the immediate feedback they provide to students, allowing them to quickly identify strengths and weaknesses. Automated grading systems make these assessments efficient and scalable, especially for large classes. However, ensuring the security and integrity of online tests can be challenging. Institutions often use secure browsers and online proctoring services to reduce cheating, but these measures can sometimes be intrusive and raise privacy concerns [22].

Written assignments and projects submitted through online platforms such as Moodle or Blackboard are also common in higher education [41]. These assessments allow students to explore topics in depth, conduct research and present their findings in a comprehensive manner. The flexibility of online submissions allows students to manage their time effectively and submit work from anywhere. However, these assessments present challenges in terms of ensuring originality and preventing plagiarism. The use of plagiarism detection software can help to maintain academic integrity. We will discuss plagiarism detection software in more detail in Section 5.2.1. However, grading written assignments can be time-consuming for teachers, especially in large classes.

Discussion boards are another valuable tool for online assessment [98]. These forums encourage students to discuss assigned topics, promote critical thinking and allow for the exchange of different perspectives. Asynchronous participation in discussion forums provides flexibility, allowing students to contribute at their own pace. However, maintaining active and meaningful participation can be challenging, and forums require careful moderation to ensure productive dialogue. Assessing individual contributions in a discussion forum can also be complex and requires clear criteria and guidelines.

Live and recorded presentations are increasingly being used as online assessments to develop and evaluate students' communication skills [47]. Live presentations using video conferencing tools such as Zoom or Microsoft Teams allow for real-time interaction and immediate feedback. Alternatively, students can submit recorded presentations that can be reviewed multiple times by teachers. This format allows students to hone their presentation skills and use multimedia to enhance their delivery. However, technical difficulties such as unstable internet connections or inadequate equipment can reduce the effectiveness of live presentations. Ensuring that students are comfortable with the technology and providing support for technical issues is essential for successful implementation.

Peer assessment is an innovative approach in which students assess each other's work against pre-determined criteria. This method promotes collaborative learning and critical thinking as students gain insight from reviewing the work of their peers [14, 43, 65, 97]. Peer assessment can also relieve teachers of some of the burden of marking. However, ensuring fairness and consistency in peer assessment

can be challenging. Providing clear rubrics and training students on how to give constructive feedback can improve the reliability of peer assessment [48].

The assessment criteria are clearly defined so that students know what to look for. These criteria help ensure that assessments are consistent and fair. A key benefit of peer assessment in the digital classroom is that it encourages self-reflection. Students learn to critically examine and improve their own work by analyzing the work of others. This process strengthens their ability to give and receive constructive feedback. In addition, peer assessment encourages collaboration and exchange between students, leading to a stronger community and a deeper understanding of the subject matter. Through the use of forums, blogs, and dedicated assessment tools, students can share their feedback quickly and efficiently. These tools also make it possible to monitor the assessment process and ensure that all students actively participate and benefit from peer assessment. We provide more insights into peer assessment in Chapter 6. Peer assessment is working particularly well when it is combined with team work. We discuss these aspects in Chapter 6.2.

Interactivity is a key element of modern online assessments, and, particularly in the context of online assessment centers during the application process of universities or companies, AI-powered platforms such as Kira Talent and Sonru are examples for this trend [2, 31, 34]. These platforms use an AI system to present questions and record student responses, which are then analyzed by the AI. This method ensures consistent question delivery and can provide an initial assessment of responses based on criteria such as language clarity and coherence. While AI-based assessments offer scalability and efficiency, they require careful question design and the integration of advanced analytics tools. In addition, students need training to navigate these platforms effectively.

Virtual reality (VR) assessments represent an innovative frontier in online education [54]. VR environments allow students and instructors to interact in simulated environments that mimic real-world scenarios. For example, medical students can practice procedures in a virtual lab, or architecture students can explore virtual buildings. VR assessments provide an immersive experience that can increase engagement and realism. However, implementing VR assessments requires access to specialized hardware and software, which can be cost prohibitive. In addition, the development of realistic and functional VR scenarios requires significant expertise and resources [26].

In summary, online assessment in higher education encompasses a variety of formats, each with its own advantages and challenges. From quizzes and written assignments to discussion forums, presentations, peer reviews, AI-powered platforms and VR environments, these tools offer flexibility, scalability and opportunities for interactive and immersive learning experiences. Successful implementation of online assessments requires careful planning, clear guidelines and ongoing support for both students and educators to ensure they are fair, effective and conducive to learning. As technology continues to evolve, so too will the methods and practices of online assessment, promising ever more innovative ways to measure and improve student learning outcomes. Building on previously established concepts and integrating them with new technological opportunities such as AI, we develop

an innovative catalogue of assessment opportunities serving a vast range of different applications. In the following we shortly describe and explore the assessment categories and solutions.

3.2. Online Written Exams

Written examinations are an essential part of academic assessment at German UDS. With the advent of digital technologies and the increasing shift of educational processes to the Internet, there are numerous opportunities to conduct written examinations online. These methods not only improve flexibility and accessibility, but also allow for innovative approaches to student assessment. A widely used form of online written assessment, as briefly mentioned above, is quizzes and tests. These tests include different question types such as multiple choice, true/false, short answer and essay. The automated scoring of these tests allows for immediate feedback to students, which supports the learning process and ensures efficiency in large classes. To ensure the integrity of the exams, secure browsers and online proctoring services must be used to minimize the risk of cheating [3, 22]. Written assignments and essays are another important method of online assessment. This form of assessment allows students to present their knowledge in a detailed and structured way. It encourages critical thinking and the ability to analyze in depth. The flexibility of online submission allows students to submit their work at their own pace and within set deadlines. Learning platforms can support this process by not only facilitating the submission and management of work, but also by providing plagiarism detection tools to ensure the originality of submitted work. Clear assessment guidelines and rubrics contribute to the transparency and fairness of assessment [2, 25]. Project work and portfolios offer students the opportunity to work on a topic over an extended period of time and to document their progress. This form of assessment encourages the practical application of theoretical knowledge and allows for creative approaches to solutions. Continuous assessment through regular rounds of feedback helps students to improve their projects. The final presentation of project results can take place either live in virtual conferences or as a recorded video.

Discussion forums and reflection reports are also important elements of the online assessment process. They allow students to express their thoughts on specific topics in writing and to actively participate in academic discussions. This form of assessment encourages critical thinking and the exchange of ideas. Asynchronous participation offers flexibility, while regular moderation by lecturers ensures the quality of the discussions [44, 98]. Innovative approaches such as automated essay grading and virtual reality (VR) offer new possibilities for online written exams. AI-powered tools such as Gradescope and Turnitin can analyze and grade essays by checking aspects such as structure, grammar and content. This allows for faster and more objective assessment of large volumes of written work. At the same time, the combination of automated assessment with manual checking by teachers can ensure the quality of assessment. Virtual reality (VR) also opens up new perspectives for written examinations. VR environments allow students to solve practical tasks in sim-

ulated scenarios and apply their skills in a realistic way. This immersive experience promotes student engagement and motivation. However, implementing VR exams requires access to appropriate hardware and software, as well as the development of realistic and functional exam scenarios [26, 46, 54].

Combining traditional methods such as quizzes and essays with innovative approaches such as AI-assisted assessment and VR exams creates a versatile and comprehensive assessment environment at German UDS. Through careful planning, clear guidelines and continuous support, these digital examination formats are successfully implemented and developed. They help to meet the ever-changing demands of higher education and ensure the quality of academic assessment at German UDS.

3.3. Online Oral Exams

Oral examinations provide a direct opportunity to assess students' knowledge and communication skills. Particularly during the Covid-19 pandemic, several innovative methods for conducting oral examinations in an online context have been established. These methods improve not only the flexibility and accessibility, but also the efficiency and quality of the examinations. One of the most common forms of online oral examinations is live video conferencing. Using complementary platforms, such as Zoom or Microsoft Teams, enables students and lecturers to interact in real time. The advantage of this method is that examiners can react immediately to students' answers and provide direct feedback. In addition, students' non-verbal communication skills, such as facial expressions and gestures, can be assessed. Implementation requires careful organization and planning of exam dates, as well as technical preparation such as ensuring a stable internet connection and working cameras and microphones. Training the examiners in the use of the videoconferencing software will also help to ensure a smooth implementation. Another option is to use pre-recorded video contributions. With this method, students prepare their answers to predefined questions and record them as videos. The videos are uploaded to a learning platform within a specified time. This form of assessment gives students the flexibility to present their answers at their own pace and without the pressure of a live exam. They can replay their recordings several times to ensure the best possible quality. Clear guidelines on video quality and length, as well as secure platforms for submitting and storing videos, are essential for the successful implementation of this method.

Interactive exams via AI-based platforms (such as Kira Talent or Sonru) are an innovative way of conducting oral exams. In these exams, students answer questions presented by an AI system. The answers are recorded and analyzed by the AI. This method has the advantage of consistent questions and automatic timing, which contributes to fair and efficient assessment. Implementation requires careful development and programming of questions, as well as the integration of advanced AI analysis tools. It also requires the training of students in the use of the platform to ensure that they are comfortable with the system [34, 42]. Virtual reality (VR) examinations are a particularly innovative form of oral examination that will be used in the future. This allows students and examiners to interact in a virtual environment

that simulates a traditional examination situation. This immersive experience encourages student engagement and enables realistic assessment. The implementation of VR exams requires access to specialized VR hardware and software, as well as the development of realistic and functional VR scenarios. Extensive training of examiners and students in the use of VR technology is essential to ensure effective use [8, 63]. In addition to these main methods, synchronous chat-based examinations offer another way of delivering oral examinations in an online context. Here, students and examiners interact in real time via text-based messaging on platforms such as Slack or Microsoft Teams. This method allows for direct communication and is particularly suitable for assessing written communication skills. The integration of audio and video elements can further enrich the assessment process. Setting up dedicated chat rooms, clear instructions and time frames, and technical support are necessary for the successful implementation of these assessments [47, 52]. Each method offers unique advantages and helps to tailor oral examinations to the needs of the students and the requirements of the subject area. Through careful planning, clear guidelines and ongoing support, these digital examination formats are implemented and developed at the German UDS. This ensures that oral examinations are not only fair and effective, but also meet the high standards of higher education and ensure the quality of academic assessment.

3.4. Online Homework Assignments

Homework assignments are a central element of the learning process and performance assessment. In the digital age, there are numerous opportunities to carry out these assignments online, which both increases flexibility and accessibility for students and supports innovative teaching methods. One of the basic forms of homework assignments are written assignments and essays. These assignments allow students to demonstrate their knowledge and understanding of specific topics in detail and develop their analytical skills. Learning platforms provide the technical infrastructure to manage these assignments effectively. Students can submit their work electronically, simplifying the process of submitting and returning assignments. The flexibility of online submission allows students to submit their work at their own pace and within set deadlines. Teachers have the opportunity to add comments directly to the digital documents and provide comprehensive feedback, which supports the students' learning process [4, 6, 28]. In addition to written assignments, project-based tasks are also a common form of homework assignments. These assignments require students to work on a topic over an extended period of time and present their findings in the form of reports, presentations or portfolios. These project-based assignments encourage the application of theoretical knowledge in practical contexts and develop skills such as project management and teamwork. Learning platforms provide suitable tools for creating and managing these projects. The ability to work collaboratively allows students to work in groups and innovate projects together, encouraging collaboration and the exchange of ideas [14, 100]. Innovative approaches to homework assignments include the use of multimedia elements. Students may

be asked to create videos, podcasts or interactive presentations to fulfil their assignments. This form of assignment encourages creativity and the use of digital tools, which are relevant in many professional contexts. Additionally, platforms such as Adobe Spark and Canva are used to enable students to create professional looking content that clearly demonstrates their ideas and insights. These assignment formats help to expand students' digital skills and strengthen their ability to communicate effectively in different media [1, 53]. Another innovative approach is gamified tasks, where game-like elements are integrated into the tasks. Gamification can increase student engagement and motivation by allowing students to earn points, level up or receive awards while completing tasks. This method can be used in particular where complex concepts need to be taught in a fun way [5, 38]. To ensure the originality and integrity of submitted work, the use of plagiarism detection software is of great importance. These tools help to ensure that students' work is their own and adheres to academic standards. In addition, automated assessment tools can be used to facilitate the work of teachers and provide quick feedback to students. These technologies analyze the submitted work for structure, grammar and content and provide an initial assessment that can be supplemented by the teacher. From traditional written assignments to project-based assignments to multimedia and gamified assignments, Homework Assignment offers a variety of ways to promote learning and comprehensively assess students' skills. By implementing these approaches, teachers can provide students at German UDS with a rich and varied learning environment that meets the demands of modern education and prepares students for their future professional challenges.

3.5. Online Presentation Assignments

Online presentation assignments offer students the opportunity to develop their communication skills, present complex topics and hone their argumentation skills. In the digital age, there are many innovative approaches and platforms for delivering these presentations effectively online. One of the most basic forms of online presentation assignments is the live presentation using video conferencing tools such as Zoom or Microsoft Teams (see 3.3). These platforms allow students to deliver their presentations in real time to a virtual audience, with interactivity and instant feedback being key benefits. Live Q&A sessions allow students to demonstrate their ability to answer questions spontaneously and discuss complex topics. Technical preparation and training are essential to ensure presentations run smoothly and technical difficulties are minimized [2, 28].

Pre-recorded presentations are an alternative. Students can record their presentations and then upload them to platforms such as Moodle, Blackboard or YouTube. The advantage of this method is that students can replay their recordings several times and submit the best version, thus improving the quality of the presentations. In addition, recorded presentations can be viewed and assessed by teachers and peers at any time, allowing for flexible time management, which is particularly important in settings with students internationally distributed over several time zones.

Clear guidelines on video quality and length, as well as instructions on how to create professional videos, are essential for successful implementation. An innovative approach to online presentation assignments is the use of interactive presentation tools such as Prezi, Adobe Spark or Canva. These platforms allow students to create dynamic and visually appealing presentations that engage audiences and enhance the comprehension of complex content. Interactive elements such as zoom effects, embedded video and animation can enhance presentations and encourage audience engagement. Teachers can use these tools to show students how to create professional-looking presentations and incorporate interactive content [2].

Virtual reality (VR) is another innovative option for online presentation assignments. With VR platform, students can deliver their presentations in a virtual environment that provides an immersive and interactive experience. This technology allows presenters to navigate through virtual spaces, show 3D models and interact with the audience in a whole new way. However, implementing VR presentations requires access to the appropriate hardware and software, as well as extensive training for students and teachers in the use of the technology [54].

Another modern approach is to create asynchronous presentations in the form of digital portfolios or blogs. Platforms such as WordPress or Mahara allow students to publish their presentations as a series of posts or pages containing text, images, videos and other multimedia elements. This method encourages continuous work on a topic and allows for in-depth exploration. Students can regularly update and expand their presentations, providing a dynamic and long-term record of their learning progress [34].

Online Presentation Assignments offer multiple benefits to students. They not only promote the development of important skills such as critical thinking, research and oral communication, but also technical competence in the use of digital tools. They also allow for flexible scheduling and the participation of students from different geographical regions, which is particularly beneficial in an international study environment. The implementation of these tasks requires careful planning and support from teachers. This includes the provision of clear instructions, technical tools and resources, as well as training in the use of different presentation tools. In addition, teachers should provide regular feedback and opportunities for students to reflect on and improve their presentation skills.

Overall, online presentation assignments offer a modern and flexible way to comprehensively assess and develop students' skills. By integrating new technologies and innovative methods, these assignment formats can be continuously developed and adapted to the changing needs of higher education. This ensures a high quality and contemporary education that prepares students for their future professional challenges.

3.6. Online Discussion Assignments

Online discussion assignments promote interaction between students and lecturers, enable the exchange of ideas and deepen the understanding of complex topics. In the

digital context, there are numerous innovative approaches to making online discussions effective and engaging. One central method for carrying out online discussion assignments is the use of discussion forums on learning platforms. In these forums, students can write posts on predefined topics and respond to the posts of their fellow students. This asynchronous form of discussion allows students to formulate and consider their thoughts thoroughly before publishing them. At the same time, it offers the flexibility to participate at different times, which is particularly beneficial for students with different schedules or in different time zones. Clear instructions and guidelines for participation as well as regular moderation by teachers can ensure the quality of discussions. This includes the provision of grading rubrics to ensure transparency and fairness in assessment [44, 100]. Another approach to online discussion assignments are synchronized discussions via video conferencing tools such as Zoom or Microsoft Teams. These real-time discussions offer the benefit of immediate interaction and feedback. Students can improve their oral communication skills and learn to respond spontaneously to questions and comments. However, conducting synchronous discussions requires careful planning and scheduling to ensure that all participants are available. Technical preparation is also crucial to ensure that all students have access to the necessary equipment and a stable internet connection. The innovative German UDS platform offers a hybrid form of online discussion that integrates elements of both synchronous and asynchronous communication. Students are able to ask questions and receive answers from both lecturers and fellow students. The platform promotes collaboration and the exchange of knowledge and creates a supportive learning environment. The ability to mark and rate posts helps to monitor the quality of discussions and ensure that relevant and useful information is highlighted. A particularly innovative approach is discussions in virtual worlds. These immersive environments provide a unique opportunity for students to interact in a simulated world that replicates real-life scenarios. Virtual worlds encourage student engagement and creativity as they can use their avatars to participate in discussions and explore complex topics in an interactive format. Implementing such discussions requires access to VR hardware and software as well as training to ensure that all participants are familiar with the technology [63]. Another modern approach to online discussion assignments are discursive blogs and wikis. Students can publish their thoughts and analyses in the form of blog entries or wiki pages, which can be commented on and further developed by their fellow students. This form of discussion encourages continuous engagement with a topic and collaboration in the creation of content. Regular feedback and clear instructions on the structure and content of the contributions are crucial to ensure the quality of the discussions [98]. Online Discussion Assignments offer multiple benefits for German UDS students. They not only promote critical thinking and the ability to analyze and discuss complex ideas, but also the development of written and oral communication skills. The flexibility that these assignments provide allows students to actively participate in the learning process regardless of their location or schedule. In addition, they create a supportive community in which students can learn from each other and build knowledge together. The successful implementation of online discussion assignments requires careful planning and support from teachers. This

includes the provision of clear instructions and guidelines, regular moderation of discussions and the use of appropriate platforms and tools. Innovative approaches such as virtual worlds and hybrid discussion platforms can further increase engagement and interactivity and provide students with a rich and diverse learning experience. Overall, Online Discussion Assignments at German UDS help to create a dynamic and interactive learning environment that prepares students for their future professional challenges. By integrating new technologies and innovative methods, these assignment formats can be continuously developed and adapted to the changing demands of higher education to ensure a high-quality and contemporary education.

3.7. Online Simulation Assignments

Online Simulation Assignments are an innovative and effective way for German UDS students to deepen their theoretical knowledge and test their skills in a controlled, realistic environment. These assignment formats allow students to work through complex scenarios that provide them with valuable insight and practical experience without involving physical resources or real-world risks. One of the most basic forms of online simulation assignments is the use of web-based simulation tools. These platforms allow students to experience and control simulation-based scenarios in real time. Examples include simulation environments that simulate realistic challenges. For example, students can manage businesses, plan projects or conduct scientific experiments. These simulations develop students' critical thinking and problem-solving skills by requiring them to make decisions and analyze their impact. The advantage of web-based simulations is that they are accessible and can be integrated into the regular online curriculum. Implementation requires careful selection of simulation software that both supports the learning objectives and is user-friendly.

Another approach are gamified simulations. These simulations integrate game-like elements to increase student engagement and motivation. In gamified scenarios, students can complete tasks in the form of games in which they earn points, level up and receive rewards. This method is particularly suitable for complex or dry topics, as it makes the learning process more entertaining and interactive. The challenge in implementing gamified simulations is to find a balance between fun and educational value to ensure that the learning objectives are not lost [38, 42].

Virtual Reality (VR) and Augmented Reality (AR) offer particularly immersive opportunities for online simulations. Students can be fully immersed in virtual worlds and experience interactive, three-dimensional scenarios. The implementation of VR and AR simulations requires investment in the appropriate hardware and software, as well as extensive training of students and faculty in the use of the technology. The advantage of this method is the immersive learning experience and the opportunity to make mistakes in a safe environment and learn from them [46].

An innovative approach to online simulation assignments is simulation-based business games, where students work in groups to make complex, strategic decisions. These simulation games can present economic, political or social scenarios and encourage teamwork and the development of leadership skills. By interacting with their

peers in a simulation-based environment, students learn to communicate effectively, resolve conflicts and think strategically. Implementation requires detailed planning and facilitation by teachers to ensure that scenarios are realistic and learning objectives are met [13, 62]. AI is also playing an increasing role in the design of online simulation assignments. AI-driven simulations can create dynamic and personalized learning environments that adapt to students' individual needs and progress. This technology makes it possible to analyze complex data and develop realistic scenarios that respond to students' learning progress. However, the implementation of AI requires careful programming and continuous adaptation of algorithms to ensure that simulations are pedagogically valuable and technically sound [33]. Overall, online simulation tasks provide a versatile and effective method for promoting hands-on learning. They allow students to apply theoretical knowledge, make decisions and experience their effects in a controlled environment. The integration of new technologies and innovative approaches enhances the quality of the learning experience and prepares students for the challenges of their future careers. Through careful planning, selection of appropriate platforms and ongoing support from faculty, these simulations can be successfully implemented and developed to meet the high standards of higher education.

3.8. Summary

One of the main reasons why the innovative use of online assessments is so significant is the flexibility and accessibility that these methods provide. Students can access the exams from anywhere and at any time, which is invaluable, especially in a globalized world and for online students. This flexibility allows online universities to serve a wider range of students, including working professionals and other target groups that traditional educational institutions may not be able to reach. In addition, the use of technologies such as artificial intelligence, virtual reality and gamified learning environments encourages student engagement and motivation. These innovative methods render the learning process more effective, more exciting, and more interactive, which is essential for a modern and future-orientated institution that meets the needs of current and future generations [89].

The innovative use of online assessments enables students' learning progress to be analyzed and evaluated in detail. By collecting and evaluating data, teachers can optimize the educational process in a targeted manner and respond to students' individual needs. This data-driven personalization of learning leads to better learning outcomes and higher student satisfaction, which further strengthens a university's reputation.

4. Scalable Assessment

Finally, depending on the type of the online course, the assessment tools have to allow scalability. This is particularly important as the maximum number of users that can be assessed without some sort of technical support by a single teacher is very low. Once that the technical support is available, however, in most cases scalability is endless and only limited by technical resources, which generally can be scaled much faster and cheaper than human resources. A very simple example is a multiple choice quiz. If a teacher has to grade the results manually, although this will be done much faster than grading any other form of assessment, the amount of students the teacher can grade is limited. Teaching assistants could be added, but they are not always that easy to recruit and also are comparatively expensive. If the multiple choice quiz is offered as an online exam, the grading can be easily automated. Once that is done, it no more matters if the work of a hundred, a thousand, or hundred thousands of learners has to be graded. At max, the available machinery needs a little boost, which in times of cloud computing, virtual machines and containers can be done more or less on the fly.

Next to automation, outsourcing the grading to the learners is a possible alternative. In the following we briefly outline several variants of automating or outsourcing assessments.

4.1. Automating Assessment

Automation of assessment is the most basic requirement to enable scaling of course landscapes and learner engagement. It describes finding tools and solutions to allow a learner's input to be graded by a computer without requiring the interaction from a human reviewer. On the one hand this allows theoretically infinite amounts of students to be graded, on the other hand, it usually provides a student with immediate feedback on their submission.

Multiple Choice, Multiple Answer Quizzes (MCQ) The easiest and, therefore, most common automated assessment form are multiple-choice or multiple-answer exams. Basically, a question is shown together with a set of answers and the students have to select the correct ones. The difference between multiple choice and multiple answer is that in one case only one correct answer is possible, in the other case multiple correct answers are possible. As this difference is marginal for the given context, both will be referred to as MCQs.

These exams are highly scalable and come as an integrated feature with most Learning Management Systems (LMS). Depending on the tools that are integrated

to the LMS, these questions can be parameterized to complicate cheating or added to question banks to ease re-use. To all those who have been students themselves or have dealt with students as teachers, it probably does not need to be mentioned that students also know that teachers like to re-use questions to reduce their workload and therefore spare no effort to get hold of last years questions.

The downside of MCQs is that they only cover a subset, the lower four levels [27], of Bloom and Anderson/Krathwohl's learning taxonomy. Furthermore, creating demanding MCQs that do not only test if students can reproduce the knowledge provided by the lectures (levels 1 and 2) but also test if they can apply their knowledge or analyze given data (levels 3 and 4) is challenging.

Drag & Drop, Connect the Dots, etc. There are a few formats that extend MCQs by providing a more pleasant and interactive user interface for more complex MCQs. Best known are *fill in the gaps*, *drag & drop*, *connect the dots*, *drag the words*, *mark the words*, etc. as provided e.g. by the H5P library[1]. Many LMS also provide such exercises as built in features to some extent. Using these exercises can certainly help to enhance the learning content with some interactivity, however, many of these libraries are JavaScript-based, which might allow creative learners to find the solution in the website's source code.

Coding Exercises Particularly on the beginner's level, coding exercises can be highly formalized and, therefore, allow a comparatively simple automation of the grading. Auto-graders for programming exercises generally can apply two different approaches dynamic and static evaluation (or a combination of both.) Dynamic evaluation compares the output of the students' code to the output that is expected by the teachers. Mostly so-called unit tests are employed to achieve this goal. The advantage of this approach is that it ensures that the students' code actually works as expected and provides proper results. The disadvantages are that writing the tests for these exercises is not really trivial and many aspects, such as particular wording, or uppercase/lowercase issues have to be taken into account. Furthermore, the code has to be executed to run these tests. This can be done either on the university's server or in some cases on the students' client computer. If the code is executed on the server, this comes with scalability and security issues, if it is executed on the client, the students are able to manipulate the results. Finally, dynamic evaluation generally just checks the final result or output of the code and not its quality, e.g. if a certain approach was used to achieve this output. To test for such properties is possible but requires a lot of expertise and effort. Static evaluation addresses some of these issues and several tools to inspect the quality and style of the code exist, however, many auto-graders mostly rely on dynamic evaluation and sometimes use static evaluation as an add-on. The main disadvantages of static code evaluation are that it requires

[1] https://h5p.org/content-types-and-applications

separate tools for each programming language[2] and does not necessarily guarantee that the code will run and produce the correct/expected results. A big advantage of static code analysis is that the code does not have to be executed and, therefore, comes with less scalability and security issues. Several auto-graders for coding exercises exist. CodeOcean, e.g., is such a tool that has been developed at the Hasso Plattner Institute. CodeOcean is completely browser based and the code and the according tests are executed on a server. This comes with several challenges for the security of the environment and the scalability but on the other hand also provides a simple point of entry for the learners, who can learn coding right away without having to install complex software first. Some of the authors of the report at hand, belonged to the research group at the Hasso Plattner Institute that has been developing such tools and examined their use in the context of MOOCs on the openHPI[3] and openSAP[4] platforms. Staubitz et al., experimented with existing web-based coding tools such as JS Fiddle [80] before developing their own solution focusing on coding novices and allowing basically coding exercises in any programming language [72, 73]. They examined the use of this tools in different contexts, such as a Python course targeting school kids [45] or a course on test-driven development in Java [84]. Serth examined how professional tools and approaches, such as pair programming, can be integrated into the programming education at schools [66]. Teusner et al., examined the learners' retention rate in programming MOOCs [93] and aspects of finding the optimal programming exercises [93]. Here they addressed questions, such as the amount of time to be spent on the programming tasks, the optimal difficulty level, the amount of guidance to be offered and the amount of necessary repetition to strengthen the retention. To do so they analyzed the data of over 3 million executions and scoring runs of participants' task submissions in the CodeOcean coding environment [93]. In the following, they also examined the impact of the descriptions of programming exercises [91] and the effects of automated interventions during the participants' attempts to solve these exercises [92]. Furthermore, Serth et al., examined the effects of contextual tips on the participants' success in these exercises [69]. Next to these educational aspects, Serth et al. also examined more technical questions, such as the improvement of security aspects of execution environments for auto-graders [67]. Other auto-graders are e.g., *Praktomat* by the KIT[5], *Jack* by Uni Duisburg-Essen, or more recently also *JupyterHub*. Elhayany et al., have examined the options to integrate JupyterHub into MOOC environments [16] and also ran some experiments with such an implementation at scale [15]. Most of the listed auto-graders are open source projects and their code is available on GitHub. Operating such systems securely, however requires a certain degree of professional skills. Staubitz et al., established CodeHarbor, a tool that allows to exchange

[2]SonarQube is a cloud-based approach to combine such tools for different programming languages. It is mostly used in professional code development and not specifically targeted towards grading. https://www.sonarsource.com/products/sonarqube/

[3]https://open.hpi.de

[4]https://open.sap.com

[5]Karlsruhe Institut of Technology

auto-gradable programming exercises between different auto-graders and to collaboratively develop such exercises [83]. An quick introduction on the possibilities and features of CodeOcean and CodeHarbor can be found in [68].

Math Exercises Similar to coding exercises, there are several approaches to automatically grade math exercises. One of them is the MatLab Grader[6], a commercial tool by MathWorks. Song et al., have established a systematic approach to design and automatically assess MatLab problems and evaluated their approach in a MOOC on communication systems. Their evaluation of the students' learning performance showed that the approach was quite effective [70]. The browser-based version of Matlab supports the LTI interface and can, therefore, easily be integrated with most LMSs. Octave is an open source alternative to MatLab. It also provides a simple web application but does not seem to support the LTI interface. To integrate such tools with an LMS is still possible, but requires more effort and expertise.

AI-based Assessment So far, the automated assessment of complex or even creative tasks is not yet fully and satisfyingly possible. The current state of the discussion rather addresses the opposite direction: Should the students be allowed to use AI tools to solve the assignments and to which extent? Perkins, et al., [56] have created an AI assessment scale that deals with exactly this question. The scale starts on the "No AI" level, which generally forbids the use of AI tools for any purpose and extends to the "Full AI" level, which allows the use of AI tools to support the generation or even completely generate the output for the assignment. With the increasing quality of generative AI, however, there are more and more approaches to use tools, such as ChatGPT to take over parts of the assessment, e.g. to generate assessment rubrics, or to independently assess the work of the student. Cleveland [10] reports about an experiment using ChatGPT to grade the work of 3rd graders. Interestingly, even ChatGPT, when asked to grade an essay warned her that it only "can provide feedback on the writing assignment based on certain criteria", but that "grading assignments typically involves subjective evaluation and requires human judgment" [10]. Even for plain text-based assignments, such as essays, AI grading so far can best be used as a supporting tool, although the results already seem to be quite encouraging. When it comes to assignments such as creating presentations or videos, painting or sculpturing, or playing or composing music, the challenges are even higher. Nevertheless, we recommend to closely follow the developments in this area as things are developing very quickly.

Learning Tools Interoperability (LTI) Interface Most of the listed tools can be connected to a wide variety of LMS via the LTI Interface. LTI is a standard that is supported by basically all relevant LMS and a wide variety of tools (not only auto-graders), which allows to flexibly combine the tools with many different LMS.

[6] https://de.mathworks.com/products/matlab-grader.html

4.2. Outsourcing Assessment

The alternative to automating the assessment in terms of scalability is outsourcing the assessment.

Peer Assessment The most appropriate form for the scalable outsourcing of assessment is the so-called peer assessment. Peer assessment solves the scalability issue by outsourcing the assessment from the instructors to the participants instead of automating this task. Additionally, it inherently pushes all exercises to the *Evaluating* level of Bloom's taxonomy, as the participants have to assess (or evaluate) the work of their peers. Albeit, it possibly can be used for exercises and tasks on all levels of Bloom's taxonomy. It is recommended, however, to focus on the top three levels, as it comes with a certain overhead. Basically, the learners are grading each other. The more learners there are, the more graders are available. It works particularly well for team assignments as individuals can assess the work of other teams, so the number of reviews that have to be written by each learner is low while the number of reviews that are received by a team is high. The higher the number of received reviews, the more accurate is the overall result as outliers can be eliminated. Peer assessment requires a certain level of maturity of the learners and, therefore, works particularly well in the life-long learning context. However, peer assessment has also been employed successfully in the school context and with regular university students. We will dive into this topic more detailed in Chapter 6.

Assessment by External Experts Next to the "internal" outsourcing via peer assessment, the grading can also be outsourced to "external", paid experts via marketplaces such as e.g. Amazon Mechanical Turk[7]. However, this approach requires a large amount of quality assessment to make sure that the assessors have a sufficient qualification. While peer assessment adds an extra quality to the given exercise for the learners as they can experience the work of other learners and the learning experience is lifted to one of the higher levels in Bloom's taxonomy, outsourcing the grading to paid experts does not add any inherent quality to the task. Therefore, peer assessment will still be a valid approach when AI-based evaluation for complex and creative tasks has reached a certain level of maturity. Paid outsourcing of the grading, however, will become obsolete.

[7] https://www.mturk.com/

5. Cheating Detection and Prevention

The detection and prevention of cheating has to be addressed on several levels. First, processes have to be established to detect cheating. These processes have to be communicated to the students, so that they are aware that cheating is a concern an will be prosecuted. On this level, as in classroom exams, online exams have to be proctored. We address this aspect in detail in Chapter 5.1. Open end text assignments, additionally, have to be checked for plagiarism, improper usage of generative AI, and texts that have been authored by professional agencies. We address these aspects in detail in Chapter 5.2.1. Additionally to the implementation of cheating detection processes, on this level the importance of integrity for the learning process has to be communicated to the students constantly. This will increase their self-esteem and help them to actually learning instead of just passing exams. The institution's academic integrity policy should be pointed out and available to the students so that they understand what is expected. A simple consent in form of a checkbox or a short reminder of the most important points that has to be read or checked before an exam already can be an incentive not to cheat.

On the second level, exams and assignments have to be designed in a way that makes it inherently hard to cheat. Simple measures can be, e.g. not to reuse questions and tasks from previous exams. This, however, significantly increases the workload for the teaching teams and prevents putting effort in the continuous improvement of existing assignments. Furthermore, the questions and assignments should be designed in a way that they address understanding and application of knowledge or encourage critical thinking, creativity, and originality rather than asking for a pure replication of the knowledge. Finally, parameterizing questions, randomly ordering the answers for each attempt, or using question pools to provide the students with different but comparable questions will also make it harder for the students to cheat.

On the third level, it is important to eliminate or at least reduce the reasons for cheating, such as the pressure to be successful and fear of failure. Proper preparation of the students, exams that match the the courses' contents, and establishing a level of understanding among the students in which detail certain topics will be addressed in the exam are helpful in that concern. Other factors are the individual weight of each exam and the time pressure the students face during the exam.

5.1. Online Proctoring

Several assessment methodologies necessitate a certain degree of surveillance. The conventional on-campus method entails assembling all students within a sizable lecture hall or equivalent space, wherein multiple teaching assistants monitor the

premises for any illicit behaviors among the examinees. For particularly large exams, an additional verification of identification may be mandated, while for smaller class settings, it is sufficient for the instructor to be familiar with their students.

Online proctoring is an advanced form of examination supervision. Exam candidates take their exam online and are monitored by special proctoring software. This software, or the person that operates the software, detects undesirable behavior and thus ensures the integrity of the examination process. Proctoring functions include verifying the identity of candidates, recognising unauthorized aids, identifying other people in the candidate's environment, and restricting the use of certain browser functionalities. The proctoring solutions are either directly integrated with the exams in existing learning management systems or the the exams are integrated with the proctoring software, or exam and proctoring software are opened in parallel tabs or browser windows. Generally, participants only need a computer or other technical device on which they can take the exam, as well as a camera and microphone to enable monitoring by the proctoring software. Some proctoring solutions require the installation of special browser plugins or standalone software to enable proctoring during the exam. Very few solutions furthermore require particular additional hardware, such as a 360°camera. Depending on the solution, the user is recorded via a webcam, the user's screen is recorded, and an additional camera stream, e.g. from a mobile phone is recorded.

These solutions enable institutions to allow participants to take their exams from home, which promotes flexibility and convenience and is essential for dedicated online forms of learning, such as the German UDS. The security and integrity of the exam is guaranteed by the efficient monitoring and control provided by the proctoring software to an extent that is at least comparable if not even better than the traditional in-person classroom proctoring.

5.1.1. Types of Online Proctoring

Modern proctoring solutions all come with some level of automation and AI support for the proctors. A wide variety of solutions and providers is offering their services world wide. Particularly, in the European context an important aspect is the provider's GDPR (General Data Protection Regulation) compatibility. A distinction is usually made between the following three types of proctoring: live proctoring, record-and-review proctoring and fully automated proctoring.

Live Proctoring Participants are monitored in real time by a human examiner, known as a proctor, during their exam. The proctors are often supported by AI technologies that can point out potential rule violations. This method enables a proctor to monitor several participants at the same time. The examinees usually have to activate their camera and microphone so that the proctor can observe all relevant activities. In some cases, the screen is also shared. The proctor usually only intervenes in the event of misbehavior and allows participants to receive immediate feedback on the validity of their exam attempt. This form offers the opportunity to recognise and correct minor infringements at an early stage. In addition, technical problems

can be clarified directly. However, this option can be more costly, as participants need to be monitored throughout the duration of the exam. It also often requires prior appointment to ensure that sufficient proctors are available. Some providers also offer the option to *bring-your-own-proctors* (BYOP). In this case the proctoring provider only delivers the proctoring software, while the course provider directly employs and provides the proctors.

Record-and-Review Proctoring In this version, the proctoring software records the participants' activities during the exam. At the end of the exam, a proctor reviews the records and decides on the validity of the exam attempt. Here too, the software can support the proctors by using AI methods. Advantages of this variant are that the examinees can take the exams whenever they want and there is no need to schedule a time slot. It is also often more cost-efficient. However, occurring issues cannot be solved during the exam and furthermore participants must accept that feedback on the validity of their exam can take several hours or even days.

Fully Automated Proctoring Automated proctoring works completely without human supervision. In this approach, AI-based technologies take over the monitoring and independently detect undesirable behavior, which can lead to the invalidation of the test. As no human examiners are directly involved in the process, this variant often proves to be the most cost-effective. Exams can also be taken at any time and participants receive immediate feedback. However, this method tends to be less accurate than other methods that combine human examiners and AI methods. It is also more difficult to clarify any problems that arise. Therefore, sometimes hybrid proctoring is offered. Exams are automatically monitored and recorded. In the event of a suspected breach of rules, a human proctor is informed and looks specifically at the relevant part of the recording. This approach potentially offers greater security, as behavior is less likely to be falsely identified as misconduct. Hybrid proctoring thus combines the efficiency of automated monitoring with the precision of human judgement. Generally, hybrid proctoring combines automated proctoring with record-and-review proctoring. A combination of automated and live proctoring can be used to reduce costs as proctors can monitor more examinees at the same time.

5.1.2. Further Functionalities

The functionalities offered by the proctoring solutions include participant verification, where examinees hold an official document such as their ID card in front of the camera to match their name and picture. During the exam, the camera and microphone are usually monitored in order to recognize attempts at cheating. This includes the presence of other people in the room, substitution of the examinee and the use of unauthorized aids. Furthermore, many providers enable room checks in which the room and workstation are shown at the start of the exam or an additional camera is added, for example via a smartphone. Some providers also offer the option of monitoring participants' laptops via screen sharing. In addition, the option of a secure browser is often offered, which restricts participants' access to the Internet

to the exam only, blocking other sources of information. This is often achieved by displaying the exam window in full-screen mode and preventing the learners from exiting full-screen or opening additional windows or tabs. Some proctoring solutions also prevent copying and pasting of text and similar actions.

5.1.3. Proctoring Providers

An initial market desk research delivered a total of 57 proctoring providers. The information about these providers is summarized in the table below. The listed data reflects the information available on the respective websites and comparison portals. It has to be noted that missing information does not necessarily mean that a particular feature is not offered by the provider, but merely indicates that no corresponding information was found. This selection does not claim to be complete but it lists the most popular offers at the time frame end of 2023 to beginning of 2024.

Of the 57 providers, 22 are headquartered in the USA, 13 in India and 8 in Europe. Of the 57 solutions, 31 offer some form of live proctoring, 24 record-and-review proctoring, 32 automated proctoring and 9 hybrid proctoring. Many providers offer several forms of proctoring to choose from. Furthermore, 28 providers state that they enable simple integration with popular learning management systems, and 25 of them offer an optional or mandatory secure browser.

Provider	HQ EU	Live Proctoring	Record and Review	Automated	Hybrid	Secure Browser
AI Proctor				✓		
AutoProctor				✓		
British Council		✓	✓	✓		
Caduceus						
ConductExam		✓				
Constructor	✓	✓	✓	✓		✓
Datamatics	✓	✓	✓			
DigiExam	✓	✓		✓		
DigiProctor		✓	✓	✓		✓
Dugga	✓	✓		✓		✓
Eklavvya				✓		✓
Examity		✓	✓	✓		
ExamOnline		✓	✓	✓		✓
ExamRoomAI		✓	✓			✓
ExamSoft			✓			
hirePro		✓		✓		
HonorLock					✓	✓
Integrity						
Advocate					✓	✓
Janison		✓	✓	✓		
Learning Spiral		✓	✓	✓		
Learnosity		✓	✓			
MapleLMS		✓	✓			✓
Mercer Metti		✓	✓	✓		✓
Merittrac						
Online						
Exam						
Maker		✓	✓	✓		
OnVUE					✓	
Oxagile		✓				
Pesofts		✓		✓		
Proctor360			✓	✓		
ProctorEdu		✓	✓			
ProctorExam	✓	✓	✓			
ProctorFree				✓	✓	✓
proctorio		✓	✓	✓		✓
ProctorTrack						✓
ProctorU		✓	✓	✓		✓
Pruefster	✓	✓	✓			
PSI		✓	✓			✓
Quilgo				✓		
Respondus				✓		✓
smartServices					✓	
SMOWL	✓		✓	✓		✓
SpeedExam					✓	
Sumadi	✓	✓	✓	✓		✓
Surpass		✓	✓			✓
Synap					✓	✓
Talview		✓	✓	✓		✓
TAO						✓
TestnTrack		✓		✓		
ThinkExam		✓		✓		✓
Top Hat						
ULearn				✓		✓
Unstop				✓		
uxpertise					✓	
Wheebox				✓	✓	✓
Wise Proctor						
Witwiser				✓		✓
Youtestme		✓	✓	✓		

5.1.4. Proctoring solutions suitable for European universities

As proctoring is a strong intrusion into the learners' privacy, we strongly focused on GDPR-compliant providers, which have a European headquarter and are claiming to store the data in the European Union. Of the 57 analysed providers, only eight are headquartered in the EU. While many other providers also claim to be GDPR-compliant, data-storage and operation of the company outside of the EU comes with additional obstacles. Due to the constantly changing regulations and unclear status of e.g. the EU-U.S. Data Privacy Framework [17], as well as the generally unstable global political situation, we only consider companies headquartered in the EU to be safe for the time being. Although several other providers claim to fulfil all GDPR requirements, further legal investigation is required, and as the past has shown political contexts can be shifting quickly in any direction.

We, therefore, classify providers who do not have a European headquarter as non-compliant for the time being and focus on the eight providers that are located within the EU. We contacted those eight providers in order to obtain more detailed information on the offered solutions and the pricing. The pricing models are differing substantially and individual calculations have to be done based on the estimated number of users and exams. Furthermore, the pricing models are often due to changes. Most providers have asked us not to reveal the details. Generally, live proctoring solutions come closest to proctoring in a classroom. Issues can be addressed directly and students can be asked to stop certain behaviors or they can be excluded from continuing their exams. In most cases they are also the most expensive solutions. In some cases, the visible costs can be lowered when the institution that is offering the courses bring their own proctors. Nevertheless, generally, this does not really reduce the total costs but just shifts the expenses. Record-and-Review are the second most expensive solutions and fully automated solutions are the cheapest. Some providers offer flat rates per user or per exam while others base their pricing on the proctored time. Most providers offer discounts based on the increasing number of participants, exams, or proctored time. Generally, it strongly depends on a variety of factors which model is the best match in terms of functionality and pricing.

5.1.4.1. Constructor

Constructor[1] is a proctoring provider based in Switzerland. Constructor offers various forms of proctoring, both AI-supported and with human proctors. For live proctoring, fixed time slots have to be arranged in advance, with several participants taking the exam at the same time. The costs depend on the number of exams and the number of participants per exam. Constructor also offers record-and-review proctoring. In this variant, participants receive feedback within 15 minutes of completing the exam. The solution is easy to use, without the need to install additional software or plugins, but only works in the Chrome browser. Constructor also offers a secure browser, enables a room scan before the start of the exam and recognizes the use of

[1] https://constructor.tech/lp/proctor

illegal aids or resources, such as smartphones. The secure browser can e.g. detect the use of ChatGPT or the switching between tabs. Constructor offers interfaces to several LMS, such as Moodle, Open edX, and others.

5.1.4.2. Datamatics

Datamatics[2] is also based in Switzerland. It offers live proctoring as well as record-and-review proctoring. In both variants proctors are supported by AI technologies. We did not receive further information on request.

5.1.4.3. DigiExam

DigiExam[3] is based in Sweden and offers an automated proctoring solution, The costs are based on the proctored exam time. Live proctoring is possible but only as a BYOP solution. A smartphone can be added as a second camera. DigiExam offers an LTI interface to integrate with LMS.

5.1.4.4. Dugga

Dugga[4] is another proctoring provider based in Sweden and offers live and automatic proctoring. Dugga also offers a secure browser. Our request for further information was not answered.

5.1.4.5. ProctorExam

ProctorExam[5] is based in the Netherlands and offers live proctoring as well as record-and-review proctoring. ProctorExam offers the possibility to monitor the participants' surroundings by connecting an additional camera, such as a smartphone. They also offer an LTI interface that enables integration with Moodle and other LMS. ProctorExam belongs to a company called Turnitin, which is specialized on plagiarism detection and prevention tools. According to the statements of the Turnitin sales representatives, ProctorExam is a deprecated solution and will be replaced with a newer product soon. This product seems to be a full-flegded exam solution including the proctoring functionality as one of the features.

5.1.4.6. Pruefster

Pruefster[6] is based in Germany and offers both live proctoring and record-and-review proctoring. Even in live proctoring settings, participants can start their exams at any time (24/7) and do not have to pre-select a fixed time slot. To enable this, Pruefster lets proctors monitor learners from different institutions in different exams next to each other. Pruefster also offers an BYOP option. In the BYOP option, the pricing is based on exams and the number of participants per exam. In the full service proctoring variant an additional fee per hour of proctoring and participant has to be

[2]https://www.datamatics.com/industries/education-technology/proctoring
[3]https://www.digiexam.com/online-proctoring/proctoring-software
[4]https://dugga.com
[5]https://proctorexam.com
[6]https://pruefster.com/de/solutions

paid. Pruefster offers to connect an additional camera for better monitoring of the room. Furthermore, there is an initial one-time setup fee. It does not offer a secure browser, but permitted aids can be defined and monitoring of the participant's screen is also possible. Pruefster offers an LTI interface to connect to LMS and also can be used standalone next to the LMS or wherever the exam is offered.

5.1.4.7. SMOWL

SMOWL[7] is headquartered in Spain and offers fully automated proctoring. Record-and-review is optionally available. The pricing is based on the number of exams and participants. An attractive flatrate based on the number of participants is alternatively available. SMOWL also offers a secure browser that recognizes which other programs are being used. Optionally, an additional camera, such as a smartphone, can be added to monitor the room or the participant's hands. It provides an interface for easy integration into Moodle and Open edX. SMOWL has been successfully employed on openHPI, the HPI's MOOC platform, from 2016 to 2023. In a comparable study in 2016 [85], SMOWL was the only proctoring provider offering fully automated proctoring in its portfolio, and thereby enabling a cost-effective and convenient procedure for the examinees.

5.1.4.8. Sumadi

Sumadi[8] is a provider originating from Latin America but its headquarter is located in the Netherlands. Sumadi offers all forms of proctoring. Sumadi also offers a secure browser that prevents the use of prohibited applications and visits to unauthorized websites. There is also an option to connect a second camera. The pricing is based on the number of participants and started hours of exams. Sumadi offers an LTI interface to be integrated with LMS.

5.1.5. Further Information

During the Covid-19 pandemic, online exams and with it online proctoring became a topic of special interest. Instructors had to find answers to organizational, technical, and legal questions. Steinbeck, et al., have examined existing literature, explored possible technical solutions that have been used in several case studies with the goal to establish a blue print for such situations [88]. Nigam et al., have systematically reviewed literature on AI-based proctoring solutions, focusing on psychological, cultural, and technological parameters. They found several aspects speaking in favor of online and AI-based proctoring solutions but also found some drawbacks, particularly concerning security and privacy issues. However, they also admit that in-person proctoring also is not always ethically perfect [55]. Hilliger, et al. [30], more generally examined research on pedagogical and technical approaches to improve trustworthiness in remote assessment by summarizing and comparing the

[7] https://smowl.net
[8] https://sumadi.net

submissions to a special issue of the Journal of Computer Assisted Learning that dealt with this question [30].

5.2. Plagiarism

Plagiarism is known back to the days of the Romans, when the poet Martial complained that his words have been plagiarized [57]. The basic definition of plagiarism is "the process or practice of using another person's ideas or work and pretending that it is your own [58]." The definition of the University of Oxford adds that it is still a plagiarism when the work is added "with [...] consent of the original author[...] [59]." Plagiarism is considered to be a violation of academic integrity and social norms, particularly but not only, in the context of education and research in many countries. [94]. In some cultures, however, the rules are different, reiterating someone else's work is sometimes seen as a sign of respect or personal authorship is seen less important. Introna, et al. [35] examined these cultural differences, and found very different interpretations of plagiarism. A Chinese student "did not think it was correct to rewrite an author's words since the author was well known and respected[.]" and, therefore, included them in their text. Other students from India, Greece, Spain and Mauritius, who were accused of plagiarism, were aware of the concept in general but the interpretation of the details were very different in their respective home countries [35]. Furthermore, even in those countries where plagiarism is frowned upon, students have to learn the rules, as schools often tend not to put a strong focus on this issue. According to Foltynek, et al. [19], "plagiarism constitutes a threat to the educational process because students may receive credit for someone else's work or complete courses without actually achieving the desired learning outcomes [19]." So, particularly, in the context of online education, with students coming from diverse cultural backgrounds it is of utmost importance that a common understanding of what is acceptable and what is not acceptable is established. Early on, plagiarism has to be detected and pointed out to the students. Avoiding plagiarism has to be part of the learning process and plagiarism detection should not only be applied when it is too late for the final theses.

Students have to be informed what is acceptable and what is not. Introductory courses to train them in proper academic conduct and the use of tools to support them to detect and avoid potential plagiarism are some of possible approaches to do so. So-called plagiarism detection tools should not only be employed to detect and punish culprits. They should be available for all students to support them in proper writing.

5.2.1. Plagiarism Detection

In 2020, an international team of researchers [19], examined 15 text-based text-matching systems and compared their accuracy in eight languages. The results differed significantly depending on the examined language. They concluded that "although some systems can indeed help identify some plagiarized content, they

clearly do not find all plagiarism and at times also identify nonplagiarized material as problematic [19]."

For the examined English language texts, the offers of Turnitin and PlagScan provided the best results. In the meantime, PlagScan has been bought by Turnitin, rendering this toolset quite unchallenged [19].

It is important to note that these systems do not actually detect plagiarisms, but similarities. The final decision whether these similarities are proper or improper use of previous work, is not provided by the system but has to be done by the teacher. The same piece of similar text can be a plagiarism, but with the same probability can also be a proper quotation of previous work. Therefore, these tools cannot replace but only support the teachers' decision.

5.2.2. Plagiarism Prevention

Supporting students to recognize plagiarisms and preventing them to submit their work in such a state is preferable than waiting for the students' final submission and punishing them. However, as mentioned above the systems only detect similarities and not plagiarisms. Therefore, the students have to be trained to understand the difference, which is a desirable skill anyway. From our perspective, therefore, it would be preferable to allow students to use such tools themselves, once they have received an introductory training and discuss certain exemplary elements of existing reports in class.

6. Peer Assessment

Peer review and peer assessment is a quite established process in the academic context. For example contributions to high quality conferences and journals are often reviewed and assessed by peers. In this context, it can be assumed that the reviewing peers have a similar background and experience as the authors of the papers and articles. More recently, peer review and peer assessment also have been established in the context of teaching. A particularly strong role was played here by research in the context of Massive Open Online Courses (MOOCs). Next to the automation of assessment, peer assessment was seen as the only option to allow large scale assessment of open and creative tasks. Some of this report's authors have previously published a substantial amount of research in that context.

Cornell University's Center for Teaching Innovation also highly recommends integrating peer assessment in the learning and teaching process:

> Peer assessment, or review, can improve overall learning by helping students become better readers, writers, and collaborators. A well-designed peer review program also develops students' evaluation and assessment skills [90].

However, everybody who has ever submitted a paper to a conference or journal for peer review, also is very likely to have experienced the feeling that sometimes the acceptance of the paper is sort of a roulette game and strongly depends on the preferences, mood, experience, stress level, etc. of the assigned reviewer. Furthermore, particularly in an educational setting, the reviewers are often no experts, which is seen as problematic by some learners. Our research has shown, however, that such experiences and prejudices are rather anecdotal and that the majority of participants of peer assessments in MOOCs appreciated the opportunity to learn from the work of their fellow learners and train their evaluation and assessment skills [74]. Furthermore, our overview on existing literature on the topic of peer assessment in MOOCs revealed that this method is quite established and researched in many aspects [21].

Completion rates have been an aspect of strong interest in MOOC research. The low completion rates often have been a deadbeat argument for traditional educators, that MOOCs are flawed by design. It has been shown that MOOCs including a peer assessment have particularly low completion rates [37]. We have confirmed this observation many times from our own experience. Our research has shown, however, that the majority of the participants did not have an issue with peer assessment as a grading methodology, but with the significantly higher workload and level of commitment that has to be shown to succeed in open and creative tasks. Time has shown that the completion rate discussion in itself was flawed as it compared apples and oranges when comparing MOOC participants to regular students. The same

applies for this particular form of tasks and assessment. While creative and open end tasks are too time consuming for many participants who are following these courses as a form of edutainment, and only attract a minority of advanced and particularly motivated learners[71], they are an essential component in full-fledged study programs no matter if offline or online.

6.1. Peer Assessment Tools

In small offline classes, peer assessment , basically, could be done simply with a piece of paper and (if blind or double blind) a hat. If the classes get larger this becomes increasingly uncomfortable. The same applies if blindness or double-blindness are strict requirements. In peer assessment, blindness delineates that the reviewer does not know who's work s/he is reviewing. Double-blindness means that also the reviewee does not know by whom s/he is reviewed. Generally, double-blindness is the preferred operational mode for summative peer assessments to minimize bias. However, given that the population of reviewers and reviewees is sufficiently familiar with each other, blindness is becoming increasingly difficult to maintain. In such cases, it can be more convenient to use dedicated tools for peer assessment. In pure online settings, such as MOOCs or the study programs of the German UDS, peer assessments are not possible without tool support. Fortunately, a variety of such tools exist and they are often embedded as central elements of a learning management system. In the context of the openHPI platform, we have developed such a tool ourselves. It was based on the core principle that reviewing and grading the work of the peers is as essential as submitting a solution to the task [79]. Therefore, it featured a built-in mechanism to ensure that only participants who have reviewed their peers are receiving any credits for their own work. This mechanism was also employed quite successfully to detect lurkers in peer assessed team tasks [86]. Open edX' built-in peer assessment tool, similar to openHPI's peer assessment tool, allows both summative and formative assessment. It also allows several combinations of peer, self, and expert assessment. Instructors define rubrics for the summative part, which are used by the participants to grant points to themselves and/or their peers. In the formative part, the participants are asked to provide their peers with constructive criticism. Open edX' implementation of peer assessment does so far not seem to automatically disqualify participants who have not reviewed the work of their peers, but rather follows a voluntary approach. An evaluation how this work in our context. In most peer assessment tools, whether built-in to an LMS or standalone, a peer assessment consists of four to five steps:

1. Step 1: Submission—The students submit a digital artifact, which is defined by the designers of the task, for review.
2. Step 2: Training—The instructors provide exemplary solutions and grades. The learners can grade these examples and compare their grading to the grading of the reviewers. This step often is optional.

3. Step 3: Review and grading—The learners grade the work of their peers and write helpful reviews. Instructors can revise the results and adjust if necessary. Ideally, only students who have properly reviewed the work of their peers will receive points for their own work. Reviewing is mandatory.
4. Step 4: Self-evaluation—Learners can evaluate their own work in comparison to the work of their peers. This step often is optional.
5. Step 5: Ideally, there is a simple feedback mechanism to allow the graded to rate the graders. The openHPI tool e.g. allowed the rated to rate the received reviews before showing the points received by the individual raters. Other tools even allow anonymous discussions between grader and graded.

If peer assessments should be blind or double-blind or not is disputable. From our perspective, in summative assessments, particularly if the participants already know each other, double-blindness is strongly recommended to avoid biases. However, particularly in small groups or communities that exist for a longer time, even with completely anonymized submissions, double-blindness cannot always be fully guaranteed as the members at some point know the work or style of their peers.

In the past we have examined many aspects of peer assessment and published our results in conferences and journals. We delineated the functionalities of our peer assessment tool [79] and a set of related tools [75] to allow team tasks to be peer assessed. We analyzed the general influence of collaborative tasks on the course outcomes [76]. We collected a first round of feedback and made some observations [78, 87]. Later on we examined the participants opinions on peer assessment in MOOCs more systematically. The results were quite positive. The participants particularly enjoyed being able to see and examine the work of their fellow students in detail to obtain a better feeling for the quality of their own work and as a source of inspiration for future tasks [74]. We examined peer assessment in the secondary education context [82], and differences between male and female participants [36]. In [21], we reviewed the existing literature on peer assessment to obtain a broader overview on existing approaches [21]. Finally, we had a more detailed look on the effects of team composition on the peer assessment outcomes [77].

Two important topics to deal with in peer assessments are the fairness and the quality of the reviews. Both topics can be easily addressed in most cases by providing the reviewers with explicit information what is expected and a set of questions to be addressed as well as the expected length of the answers. We have rarely encountered cases of so-called rogue reviews, which in the beginning have been one of our biggest concerns. Using the median value within the grading rubrics to determine the final grade already eliminates extreme outliers. In our experience, these outliers have very rarely been cases of intended misbehavior. Mostly some reviewers just had a different opinion. The probability of a fair and appropriate total grade increases with the number of reviews. Allowing the reviewed to respond to the reviewer by rating the received review, commenting it, or reporting it in severe cases, helps to further reduce cases of misconduct. Generally, the bigger issue than malicious misbehavior, is misbehavior out of laziness or lack of time. If the rate that is received for a review can be added as a component of the final grade of the reviewer, this can be used as

an additional incentive for reviewers to write meaningful reviews. The peer review system that we used to work with only showed the grade given by a single reviewer to the reviewed when s/he had rated the written review. This worked quite well as an incentive for the reviewed to actually rate their reviewers. In several cases, however, this resulted in situations where the reviewed complained that they gave a reviewer a 5-star rating as their review said: "Well done. Good job!" Then they were disappointed that those reviewers actually did not give them the number of points they expected based on the positive review. We received several complaints that the reviewed would have preferred to see the grade they received from the reviewer before they rated the review. However, the whole idea of this feature was exactly to avoid such tit-for-tat situations. While "Well done. Good job!" might imply a good grade, it is a lousy review deserving a bad rating. Similar to the the expected questions to be addressed and the expected length of the review, this needs to be trained, or at least be communicated to the learners explicitly.

In conclusion, we consider peer assessment to be an extremely suitable and reliable tool for the grading of creative and open tasks in scalable learning contexts. To guarantee the anonymity of of the submissions, a certain group size is required, which is proportional to the number of peer assessed tasks. In the following chapter, we will discuss more specific issues with peer assessment in the context of team assignments.

6.2. Group and Team Assignments

Teachers' and students' perception of team assignments often differ significantly. Many teachers are strong supporters of team work, as it teaches many skills and competencies that are required in future life and it also reduces their workload. Particularly highly motivated learners, however, often hate team work. When they are teamed up with less motivated learners, they fear that they either will have to cover all of the work themselves, or that they will receive (at best) mediocre grades if they rely on the contributions of their teammates. Furthermore, it adds a significant communicative overhead to the given tasks. This overhead was reduced to some extent by the introduction and ubiquitous availability of online communication and collaboration tools during the recent years.

In 2016, Riebe, Girardi, and Whitsed [64], conducted a systematic literature review on teamwork pedagogy in higher education. According to their sources, it is no more just "desirable" being able to work in teams, it is "essential." They refer to reports from the major English speaking countries[1], as well as from Eastern Europe and China, which express the view that the ability to work in teams and the "related interpersonal skills are equally or more important than graduates' technical skills" [64] . Hughes and Jones [32], state that, nowadays, teamwork is employed in nigh on all organizations. Furthermore, they report about a poll by the Association

[1] US, UK, Canada, and Australia

of American Colleges and Universities (AACU) in 2009, which revealed that 71% of the employers expect that colleges place a greater emphasis on teamwork skills [32].

Most modern approaches to learning and teaching, such as active learning or project-based learning, work best if several students are collaborating on the same task [29]. Each student brings in a different background, different skills and problem-solving approaches. Depending on the task, these teams are ideally multi-disciplinary. Wampfler et al. [39], have examined the general learning culture, video based lectures, and also assessment during the Covid-19 pandemic when education in many contexts instantly had to switch from on-site to on-line delivery. Their particular context was education in schools, but their results can be applied to university settings as well [39]. Basically they postulate that the goal of contemporary learning environments should be a "practiced we-culture", not only in lecturing but also in assignments and assessments. Figure 6.1 shows the four dimensions of their model.

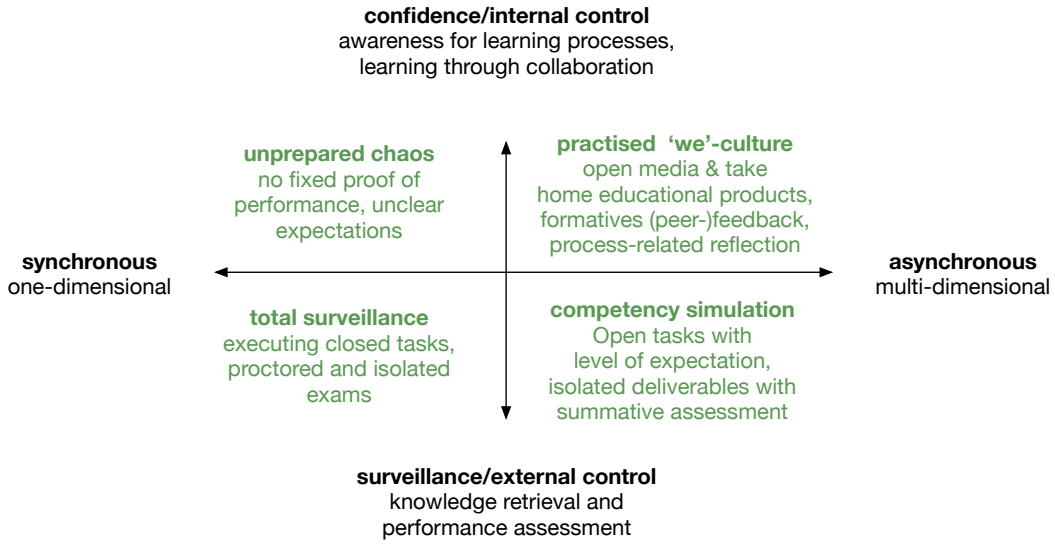

Figure 6.1.: The goal of contemporary learning environments should be a practised "we-culture". (Based on [40], CC-BY-4.0)

For a modern digital learning environment, it is crucial to support the instructors and facilitators in offering a wide range of collaborative options. Starting with a course-wide discussion forum and ending with the possibility to form small groups or teams within the larger course community. Restricting a platform and its courses to a plain video-and-multiple-choice experience, prohibits to leverage all the advantages that come with the massiveness of MOOCs, but also in smaller contexts, such as the study programs of the German UDS, disables participants and instructors to go the extra-mile. We, therefore, consider enriching the courses with hands-on tasks and exercises and enabling collaborative learning among the course participants to be crucial.

Collaborative learning, project-based learning, team-based learning, and team assignments are terms that are often used without a clear distinction. All of them delineate a method of learning that is based on the idea of working with others to solve a common challenge. In this report, we distinguish these terms by the following criteria: *Collaborative learning* delineates the general process of learning by communicating and interacting with others [23]. This process does not necessarily involve teamwork or team assignments. Enabling the learners to communicate about a given topic, or a task that each of them has to solve and submit individually, already fulfills this definition. It is also differentiated between forms of *Large Group Collaboration* (LGC) and *Small Group Collaboration* (SGC). All forms of collaboration in which participants are interacting with—potentially—all other participants of the course in question are considered as LGC. Whenever subgroups of participants are formed, whose members are supposed to interact mainly with other members of this subgroup, this is considered to be SGC. In SGC, the report additionally differentiates between loosely-coupled *groups* and tightly-coupled *teams*.

In loosely-coupled groups, participants are interacting on a certain topic out of their own motivation and interest. In tightly-coupled teams, participants are working on a common task, which is part of a graded assignment. *Project-based learning* delineates a process where learners are working on a sufficiently complex task; and—by doing this—learn the skills and competencies that the instructors intend to convey. The term *team-based learning* is used to delineate the subset of project-based learning, where learners are working in teams on a given project. If this task is graded, the term *team-based assignment* is used. In the report at hand, we focus on the assessment of the team-based assignments, using peer assessment. As we have shown in Chapter 6, peer assessment is a very promising approach, particularly, if the assessment of free-form assignments has to be scalable. If done properly, the combination of team assignments with peer assessment comes with several inherent benefits:

- As the task has to be completed by a team and the assessment of the other teams' work is done individually by each team member, the total number of reviews is much higher than in peer reviews for individual assignments, without increasing the workload of the reviewers.
- As writing reviews is mandatory, the peer assessment serves an additional purpose of filtering out free-riders in the teams [86].

In peer assessments, it is always desirable if a submission receives a large number of reviews, as the accuracy of the grading is becoming increasingly reliable with the amount of reviews. Team assignments, therefore, go particularly well hand in hand with peer assessment if the team members have to grade the other teams individually.

While many learners demand for more hands-on exercises [101], collaborative learning is a top priority only for a small subset of participants [24, 81]. So far several attempts have been made to classify the learner types in online courses. Based on the users' engagement and activity within the course, no-shows, dropouts, free-riders, and active participants are typical classifications. Poellhuber, Roy, and Bouchoucha [61] add a social component to this list by differentiating between active-

socials and active-independents. In the course that they have examined, the active-socials form the smallest group (5.6%). This group, however, also sports the highest survival rate in the course(93%) [61]. Encouraging the learners to become active-socials will, therefore, help to increase the retention rates of both the learners in the study programs as well as the knowledge, skills, and competencies within the learners.

To increase the learners' motivation for collaborative work and team tasks, we consider a proper matching of the teams to be crucial. In an international context, the participants' time zones are the most important matching criterion. This is directly followed by the students' time commitment. Matching students with a similar time commitment for a certain task reduces friction and it is more likely that teams will produce outcomes that are satisfying for all team members.

Another approach is to match students based on their skill levels and bring together those who are only slightly differing, so that the more skilled students can pull up the less skilled students. To make this work, the difference between the levels has to be sufficiently small to not frustrate both advanced and less advanced students.

Whichever method is chosen, it has to be ensured that each student's grade will be composed of a team component which is identical for all team members and an individual component, which can differ from student to student.

One way to do this is to integrate the reviews that were written by the students into the final grade. A further option is to ask the students to grade their individual contribution within the teams. openHPI's peer assessment system supported both options [75, 77, 78]. We are still evaluating how far the new Open edX based system of the German UDS supports these requirements.

Given that reviewing the work of others is a mandatory element of the peer review system and not reviewing the work of the peers results in zero points for their own work, peer assessment is the ideal complement to team assignments as it helps to eliminate lurkers [86] and by design allows a large number of reviews without increasing the workload of the individual learners. In the context of our MOOCs, we recommended a team size of five to eight members. In part this was due to the high dropout rates in MOOCs to make sure that at least three to four members are left in the end. In settings where dropouts are a lesser issue, we would rather recommend smaller team sizes. Larger teams require more organizational overhead and it is easier for lurkers to hide behind the back of more active teammates. The tasks should always be designed in a way that they also could be solved by a single person and the size of the team should be reflected in the final grade for the task.

In conclusion, we strongly advertise to encourage the students to employ all forms of collaborative learning and particularly encourage the instructors to offer project-based tasks to be solved by groups or teams. The matching of the teams should be handled by the instructors rather than by the students, particularly in the early tasks, when the students do not know each other yet. Particularly, in pure online settings such as MOOCs or the German UDS study programs, finding a well-fitting team can be quite challenging for the students [87]. To match proper teams, the instructors have to be supported by technology that allows them to collect the required data and then (semi-)automatically match the students based on homogeneous or het-

erogeneous criteria. While, e.g., time zone and time commitment should be matched homogeneously; age, gender, etc. should rather be mixed heterogeneously [77]. Next to the actual task, the evaluation process for the teamwork should always include a report about the team structure and the team processes that have been employed to solve the task [9]. We see a particular advantage in the openness, mixability, and stackability of our modules here. As learners from around the globe, Master students, MBA students, and even life-long learners pursuing micro-degrees will attend the same courses together, a variety of backgrounds and experiences is inherent in the student population of our courses. This setting is ideal for the students to learn from each other.

7. Summary and Discussion

In the report at hand we have outlined different forms of academic assessment and have introduced an assortment of assessment methods with a particular focus on scalable online learning and study formats. Furthermore, we gave an overview on related topics such as online proctoring and plagiarism prevention and detection.

As in our previous report on online teaching methods (SR2) [49], it is close to impossible to cover every detail of every aspect in this dynamic and quickly changing field. The report at hand can, therefore — particularly in those areas where it becomes tangible and introduces actual products and solutions rather than theories, concepts, or frameworks — only be a snapshot that presents the situation at the time of writing.

Nevertheless, we are convinced that this tangibility is necessary to move beyond pure academic theory and introduce hands-on solutions and are willing to pay the price for this by abandoning the timelessness of the report.

The given overview on online assessment perfectly complements SR2 [49] and will be a very helpful resource for educators on all educational levels, who want — or have to — switch from in person training to online or hybrid formats. Online learning enables learners who are living in remote areas or cannot attend synchronous on-campus classes for other reasons, such as a job, or children, or elderly relatives that need to be attended to complete their studies. Furthermore, it can contribute to reduce traffic and it allows students to live in cities that are more affordable than the typical university towns. Recent global challenges, such as pandemics, wars, or climate change are becoming increasingly likely and educators around the world should prepare themselves for situations where face-to-face classes are simply no more possible.

Particularly in scalable online learning environments, open ended tasks and assignments are challenging for teachers as they require both formative and summative assessment in different phases and both types of assessment are very time consuming. Our previous research has shown that a combination of teamwork and peer assessment is very fruitful in this setting. Next to increasing the scalability of such tasks, it enables and encourages students to learn from each other and in the same time provides them with many opportunities to gain so-called future skills, such as the ability to work in teams, the ability to constructively criticize the work of others and many more. However, we are also aware that this kind of assessment is not the golden grail for every context and type of learner. To reduce resentments and guarantee a delightful experience to all, it is important to properly match the teams and design the tasks in a way that allows multiple ways to succeed. We see a particular advantage in our open and stackable course design here, which allows us to bring students from the widest variety together and encourage them to learn with and from each other.

We already referred to Plutarch's famous metaphor of learning to the ignition of a fire rather than the filling of a bottle [60]. If the Greek philosopher's words are taken seriously in the way teaching is delivered, it also needs to be reflected in the way that learning is assessed.

Bibliography

[1] W. Admiraal, B. Huisman, and O. Pilli. "Assessment in Massive Open Online Courses". In: *Electronic Journal of e-Learning* 13 (Apr. 2015), pages 207–216.

[2] A. Ahmed and A. Pollitt. "The Support Model for interactive assessment". In: *Assessment in Education: Principles, Policy & Practice* 17.2 (2010), pages 133–167. DOI: 10.1080/09695941003694425. eprint: https://doi.org/10.1080/09695941003694425.

[3] R. Babo, L. Babo, J. Suhonen, and M. Tukiainen. "E- Assessment with Multiple-Choice Questions: A 5 Year Study of Students' Opinions and Experience". In: 19 (Jan. 2020). DOI: 10.28945/4491.

[4] Z. Baleni. "Online formative assessment in higher education: Its pros and cons". In: *Electronic Journal of e-Learning* 13 (Jan. 2015), pages 228–236.

[5] A. B. Barreto, F. Bellotti, B. Kapralos, K. Lee, P. Moreno-Ger, and R. Berta. "Assessment in and of Serious Games: An Overview". In: *Advances in Human-Computer Interaction* 2013 (Feb. 28, 2013), page 136864. DOI: 10.1155/2013/136864.

[6] C. Bonk, M. Lee, T. Reeves, and T. Reynolds, editors. *MOOCs and Open Education Around the World (1st ed.)* 2015. ISBN: 9781315751108. DOI: 10.4324/9781315751108.

[7] H. D. Brown. *Language Assessment: Principles and Classroom Practices*. Pearson Education, 2004. ISBN: 0-13-098834-0.

[8] B. Chavez and S. Bayona. "Virtual Reality in the Learning Process". In: Mar. 2018, pages 1345–1356. ISBN: 978-3-319-77711-5. DOI: 10.1007/978-3-319-77712-2_129.

[9] S. Chujfi, H. Traifeh, T. Staubitz, R. Refaie, and C. Meinel. "Exploring Collaboration and Assessment of Digital Remote Teams in Online Training Environments". In: *Workgroups eAssessment: Planning, Implementing and Analysing Frameworks*. Edited by R. Babo, N. Dey, and A. S. Ashour. Intelligent Systems Reference Library. Singapore: Springer, 2021, pages 27–53. ISBN: 9789811599088. DOI: 10.1007/978-981-15-9908-8_2.

[10] D. Cleveland. *Here's how I used ChatGPT to assess student writing*. https://ditchthattextbook.com/chatgpt-ai-assess-student-writing/. [Online; accessed 24-04-2024].

[11] E. Csibi. *Assessment in Higher Education*. https://www.pearson.eu/cee/ fileadmin/CEE/ELT/TT/Assessment_in_Higher_Education_pdf.pdf. [Online; accessed 21-03-2024].

[12] *Deciding What to Assess*. https://onq.queensu.ca/shared/TLHEM/assessments/05_s1_02_deciding_what_to_assess.html. [Online; accessed 21-03-2024].

[13] P. D. Dmitry Abbakumov and W. V. den Noortgate. "Rasch Model Extensions for Enhanced Formative Assessments in MOOCs". In: *Applied Measurement in Education* 33.2 (2020), pages 113–123. DOI: 10.1080/08957347.2020.1732382. eprint: https://doi.org/10.1080/08957347.2020.1732382.

[14] F. Dochy, M. R.Segers, and D. Sluijsmans. "The Use of Self-, Peer and Co-assessment in Higher Education: A review". In: *http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=016507/(100)* 24 (Oct. 1999). DOI: 10.1080/0307507991233137993 5.

[15] M. Elhayany and C. Meinel. "Towards Automated Code Assessment with OpenJupyter in MOOCs". In: *Proceedings of the Tenth ACM Conference on Learning @ Scale*. L@S '23. Copenhagen, Denmark: Association for Computing Machinery, 2023, pages 321–325. DOI: 10.1145/3573051.3596180.

[16] M. Elhayany, R.-R. Nair, T. Staubitz, and C. Meinel. "A Study about Future Prospects of JupyterHub in MOOCs". In: *Proceedings of the Ninth ACM Conference on Learning @ Scale*. L@S '22. New York City, NY, USA: Association for Computing Machinery, 2022, pages 275–279. ISBN: 9781450391580. DOI: 10.1145/3491140.3529537.

[17] *EU–US Data Privacy Framework*. https://en.wikipedia.org/wiki/EU%E2%80%93US_Data_Privacy_Framework. [Online; accessed 07-06-2024].

[18] *Five Principles for Meaningful Online Assessment*. https://taylorinstitute.ucalgary.ca/resources/five-principles-for-meaningful-online-assessment. [Online; accessed 21-03-2024].

[19] T. Foltýnek, D. Dlabolová, A. Anohina-Naumeca, S. Razı, J. Kravjar, L. Kamzola, J. Guerrero-Dib, Ö. Çelik, and D. Weber-Wulff. "Testing of support tools for plagiarism detection". In: *International Journal of Educational Technology in Higher Education* 17 (2020), pages 1–31.

[20] *Formative and Summative Assessment*. https://educationaltechnology.net/ formative-and-summative-assessment/. [Online; accessed 21-03-2024].

[21] D. Gamage, T. Staubitz, and M. Whiting. "Peer assessment in MOOCs: Systematic literature review". In: *Distance Education* 42.2 (Apr. 2021), pages 268–289. ISSN: 0158-7919, 1475-0198. DOI: 10.1080/01587919.2021.1911626.

[22] S. H. P. W. Gamage, J. R. Ayres, M. B. Behrend, and E. J. Smith. "Optimising Moodle quizzes for online assessments". In: *International Journal of STEM Education* 6.1 (Aug. 13, 2019), page 27. DOI: 10.1186/s40594-019-0181-4.

[23] J. M. Gerlach. "Is this collaboration?" In: *New Directions for Teaching and Learning* 1994.59 (1994), pages 5–14.

[24] F. Grünewald, C. Meinel, M. Totschnig, and C. Willems. "Designing MOOCs for the Support of Multiple Learning Styles". In: *Scaling up Learning for Sustained Impact*. Edited by D. Hernandez-Leo, T. Ley, R. Klamma, and A. Harrer. Volume 8095. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, pages 371–382. ISBN: 978-3-642-40813-7.

[25] A.-E. Guerrero-Roldán and I. Noguera. "A model for aligning assessment with competences and learning activities in online courses". In: *The Internet and Higher Education* 38 (2018), pages 36–46. ISSN: 1096-7516. DOI: https://doi.org/10.1016/j.iheduc.2018.04.005.

[26] D. Hamilton, J. McKechnie, E. Edgerton, and C. Wilson. "Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design". In: *Journal of Computers in Education* 8.1 (Mar. 1, 2021), pages 1–32. DOI: 10.1007/s40692-020-00169-2.

[27] G. R. Hancock. "Cognitive Complexity and the Comparability of Multiple-Choice and Constructed-Response Test Formats". In: *The Journal of Experimental Education* 62.2 (1994), pages 143–157. DOI: 10.1080/00220973.1994.9943836. eprint: https://doi.org/10.1080/00220973.1994.9943836.

[28] J. Heil and D. Ifenthaler. "Online Assessment in Higher Education: A Systematic Review". In: *Online Learning* 27 (Mar. 2023). DOI: 10.24059/olj.v27i1.3398.

[29] L. Helle, P. Tynjälä, and E. Olkinuora. "Project-Based Learning in Post-Secondary Education – Theory, Practice and Rubber Sling Shots". In: *Higher Education* 51.2 (Mar. 2006), pages 287–314. ISSN: 1573-174X. DOI: 10.1007/s10734-004-6386-5.

[30] I. Hilliger, J. A. Ruipérez-Valiente, G. Alexandron, and D. Gašević. "Trustworthy remote assessments: A typology of pedagogical and technological strategies". In: *Journal of Computer Assisted Learning* 38.6 (2022), pages 1507–1520. DOI: https://doi.org/10.1111/jcal.12755. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12755.

[31] N. Holmes. "Engaging with assessment: Increasing student engagement through continuous assessment". In: *Active Learning in Higher Education* 19 (Aug. 2017). DOI: 10.1177/1469787417723230.

[32] R. L. Hughes and S. K. Jones. "Developing and assessing college student teamwork skills". In: *New Directions for Institutional Research* 2011.149 (2011), pages 53–64. DOI: 10.1002/ir.380. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ir.380.

[33] D. Ifenthaler and S. Greiff. "Leveraging learning analytics for assessment and feedback." In: *Online learning analytics*. Edited by J. Liebowitz. https://doi.org/10.1201/9781003194620: Auerbach Publications, 2021, pages 1–18.

[34] D. Ifenthaler, C. Schumacher, and J. Kuzilek. "Investigating students' use of self-assessments in higher education using learning analytics". In: *Journal of Computer Assisted Learning* 39.1 (2023), pages 255–268. DOI: https://doi.org/10.1111/jcal.12744. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12744.

[35] L. Introna, N. Hayes, L. Blair, and E. Wood. *Cultural Attitudes towards Plagiarism: Developing a better Understanding of the Needs of Students from Diverse Cultural Backgrounds Relating to Issues of Plagiarism*. Technical report. [Online; accessed 24-04-2024]. University of Lancaster, 2003.

[36] C. John, T. Staubitz, and C. Meinel. "Performance of Men and Women in Graded Team Assignments in MOOCs". In: *2019 IEEE Learning With MOOCS (LWMOOCS)*. 2019, pages 30–35. DOI: 10.1109/LWMOOCS47620.2019.8939654.

[37] K. Jordan. *"MOOC Completion Rates: The Data," The Katy Jordan MOOC Project*. http://www.irrodl.org/index.php/irrodl/article/view/2112/3340. [Online; accessed 26-04-2019]. 2015.

[38] Y. J. Kim and D. Ifenthaler. "Game-Based Assessment: The Past Ten Years and Moving Forward". In: *Game-Based Assessment Revisited*. Edited by D. Ifenthaler and Y. J. Kim. Cham: Springer International Publishing, 2019, pages 3–11. ISBN: 978-3-030-15569-8. DOI: 10.1007/978-3-030-15569-8_1.

[39] W. Klee, P. Wampfler, and A. Krommer (Hrsg.) *Hybrides Lernen*. [Online; accessed 24-04-2024]. Beltz Verlag Weinheim, 2021. ISBN: 978-3-407-63223-4.

[40] A. Langela-Bickenbach and P. Wampfler. *Lernkultur, Videokonferenzen und Prüfungen*. https://hackmd.io/ojjzyKPfTWq1oflsoLj1Og. [Online; accessed 18-05-2024].

[41] S. Law. "Using digital tools to assess and improve college student writing". In: *Higher Education Studies* 9 (Feb. 2019), pages 117–123. DOI: 10.5539/hes.v9n2p117.

[42] H.-S. Lee, G.-H. Gweon, T. Lord, N. Paessel, A. Pallant, and S. Pryputniewicz. "Machine Learning-Enabled Automated Feedback: Supporting Students' Revision of Scientific Arguments Based on Data Drawn from Simulation". In: *Journal of Science Education and Technology* 30 (Jan. 2021), pages 1–25. DOI: 10.1007/s10956-020-09889-7.

[43] L. Li and F. Gao. "The effect of peer assessment on project performance of students at different learning levels". In: *Assessment & Evaluation in Higher Education* 41.6 (2016), pages 885–900. DOI: 10.1080/02602938.2015.1048185. eprint: https://doi.org/10.1080/02602938.2015.1048185.

[44] X. Liu, L. Li, and Z. Zhang. "Small group discussion as a key component in online assessment training for enhanced student learning in web-based peer assessment". In: *Assessment & Evaluation in Higher Education* 43 (May 2017), pages 1–16. DOI: 10.1080/02602938.2017.1324018.

[45] M. v. Löwis, T. Staubitz, R. Teusner, J. Renz, C. Meinel, and S. Tannert. "Scaling youth development training in IT using an xMOOC platform". In: *Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE*. Oct. 2015, pages 1–9. DOI: 10.1109/FIE.2015.7344145.

[46] G. Makransky and G. B. Petersen. "The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality". In: *Educational Psychology Review* 33.3 (Sept. 1, 2021), pages 937–958. DOI: 10.1007/s10648-020-09586-2.

[47] J. Mao and K. L. Peck. "Assessment Strategies, Self-Regulated Learning Skills, and Perceptions of Assessment in Online Learning." In: *The Quarterly Review of Distance Education* 14 (2013), pages 75–95.

[48] S. Meek, L. Blakemore, and L. Marks. "Is peer review an appropriate form of assessment in a MOOC? Student participation and performance in formative peer review". In: *Assessment & Evaluation in Higher Education* 42 (Aug. 2016), pages 1–14. DOI: 10.1080/02602938.2016.1221052.

[49] C. Meinel, D. Köhler, and T. Staubitz. *Innovative Formats for Online Teaching*. Technical report. German University of Digital Science, Mar. 2024.

[50] C. Meinel, C. Willems, T. Staubitz, D. Sauer, and C. Hagedorn. *openHPI: 10 Years of MOOCs at the Hasso Plattner Institute*. Technische Berichte des Hasso-Plattner-Instituts für Digital Engineering an der Universität Potsdam 148. Google-Books-ID: 7rilEAAAQBAJ. Potsdam, Germany: Universitätsverlag Potsdam, Oct. 2022. ISBN: 978-3-86956-544-6.

[51] R. Miller and A. Leskes. *Levels of Assessment: From the Student to the Institution*. AAC&U's Greater Expectations Series. Association of American Colleges and Universities, 2005.

[52] M. Montenegro-Rueda, A. Luque-de la Rosa, J. L. Sarasola Sánchez-Serrano, and J. Fernández-Cerero. "Assessment in Higher Education during the COVID-19 Pandemic: A Systematic Review". In: *Sustainability* 13.19 (2021). ISSN: 2071-1050. DOI: 10.3390/su131910509.

[53] A. Moreira and A. Balula. *Evaluation of Online Higher Education - Learning, Interaction and Technology*. Jan. 2014. ISBN: 978-3-319-05424-7. DOI: 10.1007/2F978-3-319-05425-4.

[54] S. Müser and C. D. Fehling. "AR/VR.nrw –Augmented und Virtual Reality in der Hochschullehre". In: *HMD Praxis der Wirtschaftsinformatik* 59.1 (Feb. 1, 2022), pages 122–141. DOI: 10.1365/s40702-021-00815-y.

[55] A. Nigam, R. Pasricha, T. Singh, and P. Churi. "A Systematic Review on AI-based Proctoring Systems: Past, Present and Future". In: *Education and Information Technologies* 26.5 (Sept. 1, 2021), pages 6421–6445. DOI: https://doi.org/10.1007/s10639-021-10597-x.

[56] M. Perkins, L. Furze, J. Roe, and J. MacVaugh. "The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment". In: *Journal of University Teaching and Learning Practice* 21.06 (Apr. 2024). ISSN: 1449-9789. DOI: 10.53761/q3azde36.

[57] *Plagiarism*. https://en.wikipedia.org/wiki/Plagiarism. [Online; accessed 24-04-2024].

[58] *Plagiarism*. https://dictionary.cambridge.org/dictionary/english/plagiarism. [Online; accessed 24-04-2024].

[59] *Plagiarism Information about what plagiarism is, and how you can avoid it.* https://www.ox.ac.uk/students/academic/guidance/skills/plagiarism. [Online; accessed 24-04-2024].

[60] Plutarch. *Moralia. On Listening to Lectures*. LCL. Cambridge, MA, US: Harvard University Press, 1927. DOI: 10.4159/DLCL.plutarch-moralia_listening_lectures.1927. eprint: https://penelope.uchicago.edu/Thayer/E/Roman/Texts/Plutarch/Moralia/De_auditu*.html.

[61] B. Poellhuber, N. Roy, and I. Bouchoucha. "Understanding Participant's Behaviour in Massively Open Online Courses". In: *The International Review of Research in Open and Distributed Learning* 20.1 (Feb. 2019).

[62] G. Polito and M. Temperini. "A gamified web based system for computer programming learning". In: *Computers and Education: Artificial Intelligence* 2 (2021), page 100029. ISSN: 2666-920X. DOI: https://doi.org/10.1016/j.caeai.2021.100029.

[63] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt. "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda". In: *Computers & Education* 147 (2020), page 103778. ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2019.103778.

[64] L. Riebe, A. Girardi, and C. Whitsed. "A Systematic Literature Review of Teamwork Pedagogy in Higher Education". In: *Small Group Research* 47.6 (2016), pages 619–664. DOI: 10.1177/1046496416665221. eprint: https://doi.org/10.1177/1046496416665221.

[65] B. Sekendiz. "Utilisation of formative peer-assessment in distance online education: a case study of a multi-model sport management unit". In: *Interactive Learning Environments* 26.5 (2018), pages 682–694. DOI: 10.1080/10494820.2017.1396229. eprint: https://doi.org/10.1080/10494820.2017.1396229.

[66] S. Serth. "Integrating Professional Tools in Programming Education with MOOCs". In: *2019 IEEE Frontiers in Education Conference (FIE)*. Cincinnati, OH, USA, Oct. 2019, page 2. ISBN: 978-1-72811-746-1. DOI: 10.1109/FIE43999.2019.9028643.

[67] S. Serth, D. Köhler, L. Marschke, F. Auringer, K. Hanff, J.-E. Hellenberg, T. Kantusch, M. Paß, and C. Meinel. "Improving the Scalability and Security of Execution Environments for Auto-Graders in the Context of MOOCs". In: *Fifth Workshop "Automatische Bewertung von Programmieraufgaben" (ABP 2021)*. Edited by A. Greubel, S. Strickroth, and M. Striewe. Volume 5. Virtual Event, Germany: Gesellschaft für Informatik e.V. (GI), Oct. 2021, pages 3–10. DOI: 10.18420/abp2021-1.

[68] S. Serth, T. Staubitz, R. Teusner, and C. Meinel. "CodeOcean and CodeHarbor: Auto-Grader and Code Repository". In: *SPLICE Workshop CS Education Infrastructure for All III: From Ideas to Practice*. Virtual Event, Mar. 2021, page 5.

[69] S. Serth, R. Teusner, and C. Meinel. "Impact of Contextual Tips for Auto-Gradable Programming Exercises in MOOCs". In: *Eighth ACM Conference on Learning @ Scale*. Virtual Event, Germany: ACM, June 2021, pages 307–310. ISBN: 978-1-4503-8215-1. DOI: 10.1145/3430895.3460166.

[70] S. H. Song, M. Antonelli, T. W. K. Fung, B. D. Armstrong, A. Chong, A. Lo, and B. E. Shi. "Developing and Assessing MATLAB Exercises for Active Concept Learning". In: *IEEE Transactions on Education* 62.1 (2019), pages 2–10. DOI: 10.1109/TE.2018.2811406.

[71] T. Staubitz. "Gradable team assignments in large scale learning environments". doctoralthesis. Universität Potsdam, 2020, page 122. DOI: 10.25932/publishup-47183.

[72] T. Staubitz, H. Klement, J. Renz, R. Teusner, and C. Meinel. "Towards practical programming exercises and automated assessment in Massive Open Online Courses". In: *2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. Dec. 2015, pages 23–30. DOI: 10.1109/TALE.2015.7386010.

[73] T. Staubitz, H. Klement, R. Teusner, J. Renz, and C. Meinel. "CodeOcean - A versatile platform for practical programming exercises in online environments". In: *2016 IEEE Global Engineering Education Conference (EDUCON)*. Apr. 2016, pages 314–323. DOI: 10.1109/EDUCON.2016.7474573.

[74] T. Staubitz and C. Meinel. "A Systematic Quantitative and Qualitative Analysis of Participants' Opinions on Peer Assessment in Surveys and Course Forum Discussions of MOOCs". In: *2020 IEEE Global Engineering Education Conference (EDUCON)*. Porto, Portugal: IEEE, Apr. 2020, pages 962–971. ISBN: 978-1-72810-930-5. DOI: 10.1109/EDUCON45650.2020.9125089.

[75] T. Staubitz and C. Meinel. "Collaboration and Teamwork on a MOOC Platform: A Toolset". In: *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*. L@S '17. Cambridge, Massachusetts, USA: ACM, 2017, pages 165–168. ISBN: 978-1-4503-4450-0.

[76] T. Staubitz and C. Meinel. "Collaborative Learning in MOOCs Approaches and Experiments". In: *2018 IEEE Frontiers in Education Conference (FIE)*. Oct. 2018, pages 1–9. DOI: 10.1109/FIE.2018.8659340.

[77] T. Staubitz and C. Meinel. "Graded Team Assignments in MOOCs: Effects of Team Composition and Further Factors on Team Dropout Rates and Performance". In: *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*. L@S '19. Chicago, IL, USA: ACM, 2019, 5:1–5:10. ISBN: 978-1-4503-6804-9. DOI: 10.1145/3330430.3333619.

[78] T. Staubitz and C. Meinel. "Team Based Assignments in MOOCs: Results and Observations". In: *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. L@S '18. London, United Kingdom: ACM, 2018, 47:1–47:4. ISBN: 978-1-4503-5886-6. DOI: 10.1145/3231644.3231705.

[79] T. Staubitz, D. Petrick, M. Bauer, J. Renz, and C. Meinel. "Improving the Peer Assessment Experience on MOOC Platforms". In: *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*. L@S '16. Edinburgh, Scotland, UK: ACM, 2016, pages 389–398.

[80] T. Staubitz, J. Renz, C. Willems, J. Jasper, and C. Meinel. "Lightweight Ad Hoc Assessment of Practical Programming Skills at Scale". In: *Global Engineering Education Conference (EDUCON), 2014 IEEE*. IEEE. 2014, pages 475–483.

[81] T. Staubitz, J. Renz, C. Willems, and C. Meinel. "Supporting Social Interaction and Collaboration on an xMOOC Platform". In: *EDULEARN14 Proceedings*. 6th International Conference on Education and New Learning Technologies. Barcelona, Spain: IATED, July 2014, pages 6667–6677.

[82] T. Staubitz, R. Teusner, and C. Meinel. "MOOCs in Secondary Education - Experiments and Observations from German Classrooms". In: *2019 IEEE Global Engineering Education Conference (EDUCON)*. 2019.

[83] T. Staubitz, R. Teusner, and C. Meinel. "Towards a repository for open autogradable programming exercises". In: *2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. Dec. 2017, pages 66–73. DOI: 10.1109/TALE.2017.8252306.

[84] T. Staubitz, R. Teusner, C. Meinel, and N. Prakash. "Cellular Automata as an Example for Advanced Beginners Level Coding Exercises in a MOOC on Test Driven Development". In: *International Journal of Engineering Pedagogy (iJEP)* 7.2 (2017), pages 125–141.

[85] T. Staubitz, R. Teusner, J. Renz, and C. Meinel. "An Experiment in Automated Proctoring". In: *European Stakeholders Summit on experiences and best practices in and around MOOCs(EMOOCs 2016)*. 2016. ISBN: 9783739237107.

[86] T. Staubitz, H. Traifeh, S. Chujfi, and C. Meinel. "Have Your Tickets Ready! Impede Free Riding in Large Scale Team Assignments". In: *Proceedings of the Seventh ACM Conference on Learning at Scale*. L@S '20. Virtual Event, USA: Association for Computing Machinery, 2020, pages 349–352. ISBN: 9781450379519. DOI: 10.1145/3386527.3406744.

[87] T. Staubitz, H. Traifeh, and C. Meinel. "Team-Based Assignments in MOOCs - User Feedback". In: *2018 Learning With MOOCS (LWMOOCS)*. Sept. 2018, pages 39–42. DOI: 10.1109/LWMOOCS.2018.8534607.

[88] H. Steinbeck, T. Staubitz, and C. Meinel. "Proctoring und digitale Prüfungen - Durchführungsbeispiele und Gestaltungselemente für die digitale Lehre". In: *DELFI 2021*. Bonn: Gesellschaft für Informatik e.V., 2021, pages 253–264. ISBN: 978-3-88579-710-4.

[89] "Students' Perspectives on the Use of Innovative and Interactive Teaching Methods at the University of Nouakchott Al Aasriya, Mauritania: English Department as a Case Study". In: *International Journal of Technology, Innovation and Management (IJTIM)* 1.2 (Dec. 2021), pages 90–104. DOI: 10.54489/ijtim.v1i2.21.

[90] C. for Teaching Innovation. *Teaching students to evaluate each other*. https://teaching.cornell.edu/resource/teaching-students-evaluate-each-other. [Online; accessed 17-05-2024].

[91] R. Teusner and T. Hille. "On the Impact of Programming Exercise Descriptions". In: *2018 Learning With MOOCS (LWMOOCS)*. Sept. 2018, pages 51–54. DOI: 10.1109/LWMOOCS.2018.8534676.

[92] R. Teusner, T. Hille, and T. Staubitz. "Effects of Automated Interventions in Programming Assignments: Evidence from a Field Experiment". In: *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. L@S '18. London, United Kingdom: ACM, 2018, 60:1–60:10. ISBN: 978-1-4503-5886-6. DOI: 10.1145/3231644.3231650.

[93] R. Teusner, C. Matthies, and T. Staubitz. "What Stays in Mind? - Retention Rates in Programming MOOCs". In: *2018 IEEE Frontiers in Education Conference (FIE)*. Oct. 2018, pages 1–9. DOI: 10.1109/FIE.2018.8658890.

[94] *The Fundamental Values of Academic Integrity, Third Edition*. [Online; accessed 24-04-2024]. International Center for Academic Integrity, 2021. ISBN: 978-0-9914906-7-7.

[95] *The National Business Ethics Survey of the U.S. Workforce*. Technical report. Ethics Resource Center, 2013.

[96] *These students cheated on a test and got away with it. Then the blackmail started*. https://www.smh.com.au/education/these-students-cheated-on-a-test-and-got-away-with-it-then-the-blackmail-started-20240320-p5fe1x.html. [Online; accessed 09-04-2024].

[97] R. Wadmany and O. Melamed. ""New Media in Education" MOOC: Improving Peer Assessments of Students' Plans and their Innovativeness". In: *Journal of Education and e-Learning Research* 5 (Jan. 2018), pages 122–130. DOI: 10.20448/journal.509.2018.52.122.130.

[98] Y.-M. Wang. "Enhancing the Quality of Online Discussion—Assessment Matters". In: *Journal of Educational Technology Systems* 48.1 (2019), pages 112–129. DOI: 10.1177/0047239519861416. eprint: https://doi.org/10.1177/0047239519861416.

[99] *Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning.* https://eric.ed.gov/?id=ED442280. [Online; accessed 21-03-2024].

[100] X. Wei, N. Saab, and W. Admiraal. "Assessment of cognitive, behavioral, and affective learning outcomes in massive open online courses: A systematic literature review". In: *Computers & Education* 163 (2021), page 104097. ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2020.104097.

[101] C. Willems, J. Jasper, and C. Meinel. "Introducing Hands-On Experience to a Massive Open Online Course on openHPI". In: *Proceedings of IEEE International Conference on Teaching, Assessment and Learning for Engineering* (*TALE 2013*). Kuta, Bali, Indonesia: IEEE Press, 2013.

[102] *Wir brauchen neue Prüfungsformen – und zwar schnell!* https://www.spiegel.de/start/studium-hochschulen-brauchen-neue-pruefungsformen-und-zwar-schnell-a-78c4d111-f3b5-4587-be86-b076026e45d0?sara_ref=re-so-app-sh. [Online; accessed 21-03-2024].

A. Assessment Tools

The following list of online exercise and assessment tools is not comprehensive and also not static. Often, tools are no more continued to be maintained after a certain amount of time.

Table A.1.: Selected online assessment tools

Provider	Description
Artemis	An open source tool created and maintained by TU Munich, supporting many different exercise types, including auto-gradable programming exercises, quizzes, modelling exercises, and many more. Code: https://github.com/ls1intum/Artemis Project: https://artemis.cit.tum.de/
CodeHarbor	A repository for auto-gradable programming exercises developed and maintained by the Hasso Plattner Institute's openHPI team. Enables collaborative editing and sharing of exercises. Exercises can be listed on all platforms supporting the IEEE LOM model and shared with all auto-graders supporting the ProFormA format. Code: https://github.com/openHPI/codeharbor Project: https://codeharbor.openhpi.de/
CodeOcean	An auto-grader for coding exercises in a variety of programming languages. Supports Java, Python, Ruby, Javascript, and R out of the box. Additional adapters can be added easily. CodeOcean is developed and maintained by the Hasso Plattner Institute's openHPI team. Code: https://github.com/openHPI/codeocean Project: https://github.com/openHPI/codeocean/blob/master/README.md
CodingBat	A free site of live coding problems to build coding skill in Java and Python. It is a project by a computer science lecturer at Stanford University. Project: https://codingbat.com/java
Jack	An auto-grader with multiple exercise types developed by the University of Duisburg Essen. The code does not seem to be open source. According to their website, the service can be offered on request for other universities as well. Project: https://www.uni-due.de/zim/services/jack.php
JupyterHub	Open source project providing a multi-user version of the Jupyter Notebook. Its main purpose are rather exercises and executable examples than exams. Currently, several parties are working on the integration of auto-grader functionalities. Code: https://github.com/jupyter/ Project: https://jupyter.org/hub
LabXchange	Global science classroom created and maintained by Harvard University. Can be used for free by learners world wide. Its main purpose is to provide labs and exercises and not exams. Supported subjects are Chemistry, Biology and Physics. Project: https://about.labxchange.org/

Table A.2.: Selected online assessment tools

Provider	Description
Mathworks	Commercial provider for math and engineering exercises and exams. Project: https://www.mathworks.com/products/matlab-grader.html
Peergrade	Commercial peer assessment tool from Denmark. Project: https://www.peergrade.io
PeerStudio	Formative peer assessment tool developed by UC San Diego and Carnegie Mellon University. Project: https://www.peerstudio.org/
Praktomat	An auto-grader for programming exercises developed by Karlsruhe Institute of Technology. The original code repository has not been updated since 2 years, but there is a more recent fork maintained by the Hochschule Offenburg. Code: https://github.com/KITPraktomatTeam/Praktomat Fork: https://github.com/hso-praktomat/praktomat Project: https://pp.ipd.kit.edu/projects/praktomat/praktomat.php
SonarQube	Cloud-based tool for static code analysis. Combines several static code analysis tools for a variety of programming languages. SonarQube is not an auto-grader but can support in evaluating the quality of code. Project: https://www.sonarsource.com/products/sonarqube/ Code: https://www.sonarsource.com/products/sonarqube/downloads/
TAO Testing	An open source online testing tool developed by the University of Luxemburg. It is now maintained by the Open Assessment Technologies community. Code: https://github.com/oat-sa Project: https://www.taotesting.com/higher-education/
Learning Management Systems (LMS)	Basically all learning management systems, such as Canvas, Moodle, Open edX, Sakai, Blackboard, Ilias, LonCapa, etc. include at least a simple assessment tool to create multiple choice quizzes or similar. Some of them even provide more sophisticated tools for instructor, peer, and self-assessment. Moodle: https://moodle.org/ Open edX: https://openedx.org// Sakai: https://www.sakailms.org// Blackboard: https://www.blackboard.com/ Ilias: https://www.ilias.de/ LonCapa: https://www.lon-capa.org/
Learning Tools Interoperability (LTI)	The LTI interface allows to connect additional assessment tools, such as the ones listed above, to the LMS. The LTI interface is supported by most LMS. Libraries in many programming languages exist for a more or less simple integration of LTI on both ends (LMS (LTI consumer) and assessment tool (LTI provider).) Project: https://www.imsglobal.org/lti-fundamentals-faq

B. Glossary

A

Anderson, Lorin Willard. American educator and psychologist. Revised Bloom's taxonomy with Krathwol.

AI-based Assessment

Auto-grader Tool that automatically grades exercises in a formalized context. Auto-graders can be simple evaluators of Multiple Choice Exams or complex tools to automatically grade Math or Programming exercises. The term is mostly used in the context of automatically grading Programming exercises. In this context, auto-graders can follow either dynamic or static evaluation approaches or combinations of both.

Automated proctoring see Fully automated proctoring, see also Proctoring.

B

Bloom's Taxonomy A hierarchical classification system for learning objectives, named after Benjamin Bloom. First published in 1956 and in its revised version (by Anderson and Krathwol) still widely used.

Brown's Principles of Assessment H. Douglas Brown, the former president of TESOL International Association (Teachers of English to Speakers of Other Languages), the largest professional organization for teachers of English as a second or foreign language, wrote several books on language learning and assessment of language learning. His five principles of assessment—Practicality, Reliability, Validity, Authenticity, and Washback—can be applied to assessments in many subjects beyond language learning.

Bulimic Learning Undesirable form of "learning" that consists of "gulping down" knowledge very quickly before an exam and "vomitting it out" again right after.

C

Carnegie-Mellon private research university in Pittsburgh, Pennsylvania.

ChatGPT A generative ChatBot by OpenAI.

Collaborative Learning Learning is a social activity. Collaborative learning contains all forms of learning where several individuals are learning together and support each other. In some situations the line between collaborative learning and cheating can become quite thin. We recommend to define exactly where desirable behavior ends. We also recommend to generally encourage collaboration in learning.

COVID-19 pandemic world wide virus epidemic causing lock-downs and many other disruptions in everyday life from December 2019 until about 2023.

D - E

edX MOOC platform founded as a joint operation between MIT and Harvard University.

Expert Assessment most common form of assessment in academic institutions. The work of the students is assessed by experts such as professors or teaching assistants.

Essays Short textual assignments.

F

Formative Assessment This form of assessment attempts to improve future results, e.g. of a future of a teacher. Generally, rather a verbose statement or feedback than a measurable grade.

Fully Automated Proctoring Form of (online) proctoring that comes without human intervention. AI-based tools make sure that the correct person is taking the exam and records or prevents a limited set of undesired behaviors. Generally, fully automated proctoring is the most scalable and also the cheapest form of proctoring. However, it comes with certain limitations in what can be detected.

G

German University of Digital Science (German UDS). New fully digital private university in Germany.

H

HPI Hasso Plattner Institute. Research institute and faculty of digital engineering at the University of Potsdam.

I

Interactivity In the context of e-learning interactivity is often defined as learners interacting with the content. In such cases the content is often presented in forms that allow digital interaction such as clicking on buttons or drag & drop. The Oxford dictionary provides a more general and in our opinion more suitable definition: "The process of people or things working together and influencing each other." Inter-human interaction can be differentiated in learner-teacher (e.g. a learner asks a question in the forum and a teacher answers), learner-learner (e.g. peer assessment), and teacher-teacher (e.g. two teachers discuss a certain topic in a video) interaction.

K - L

Krathwol American psychologist and educational researcher. Revised Bloom's taxonomy with Anderson.

Learning Management Systems (LMS) Online tool to manage learning, courses, and classes. A variety of tools and providers with very different business models from open-source to commercial exist. Most LMS include at least a basic multiple choice quiz system for tests and exams. Some also include additional exam forms. Most LMS support the LTI interface to allow connecting further learning and exam tools.

Learning Tools Interoperability (LTI) Interface standardized interface allowing to connect online learning tools to LMS.

Live Proctoring Form of online proctoring where a human proctor monitors one or more students in real time while they are working on an exam.

M

MOOCs Massive Open Online Courses. Started in Canada around 2008 by Siemens and Downes to proof their connectivist learning theory. Hyped around 2012 by the NY Times and other media. Courses that are offered online by renowned universities (originally) free of charge and open for anyone to join.

Multiple Choice Tests Students are given the questions and a set of answers. Generally, they have to select the correct answers. Depending on the amount of possible correct answers, it is distinguished between "multiple choice" and "multiple answers" questions. Multiple choice questions have exactly one possible answer, multiple answers questions can have several possible answers (0-all). We strongly recommend to never use zero correct answers. We, furthermore, recommend to avoid questions where all answers are correct. Multiple choice questions are generally easier to answer than multiple choice questions. Depending on the used quiz system the grading of these questions can differ significantly. Multiple choice questions can allow more than one correct answer. The students can then select from several possible options and will receive full points whichever of the correct answers they have chosen. Multiple answer questions might provide zero points if a wrong answer is chosen or deduct points for each wrong answer but mostly do not provide less than zero points. The grading of these tests can be automated very simple, therefore, they are employed very often in online settings.

O

openHPI The MOOC platform of the Hasso Plattner Institute.

P

Peer Assessment Form of assessment where the students are assessing each others' work. Currently, the only option to properly assess open ended tasks in a scalable way.

Pedagogy , the method and practice of teaching children.

Problem-Based Learning (PBL) student-centered approach to learning. Instead of "being instructed" students learn by actively solving an open-ended problem. Skills such as team collaboration, communication, creativity are at least as important as the actual expertise in a certain topic area to successfully deliver a solution.

Plagiarism the inadequate re-use of previous ideas without proper quoting and citing. In academic environments building on previous ideas is generally desirable. However, it has to be clearly shown where own ideas start and previous ideas end. Also who has had the previous idea and where to find the original text. There are cultural differences in how this is done and students have to learn how to cite and quote properly and compliant with academic integrity.

Problem-Oriented Learning see Problem-based learning.

Proctoring The monitoring or surveillance of students during exams. In offline situations, this is often done by teaching assistants. In online exam situations several tools and providers exists to support the teachers. These tools can be either fully automated (see automated proctoring), allow live proctoring (see live proctoring) or employ record-and-review proctoring (see record-and-review proctoring).

Project-Based Learning see Problem-based learning.

Prompting The art of writing instructions for generative AI tools.

R

Record-and-Review Proctoring A video of the person taking the exam is recorded. The proctor watches the recording later on. In most cases a proctor will watch several recordings in parallel. Record-and-Review proctoring simplifies the scheduling as learner and proctor do not have to be available in the same time. Generally, it is a little cheaper than live proctoring as well.

S

Scalability In the context of e-learning, scalability is the ability to provide the same, or at least a similar, learning experience to 10 learners as to 10.000 learners.

Self-Assessment Form of assessment where students are assessing their own work. Generally, only used as a complementary option in combination with peer assessment.

Self-Regulated Learning Learners set their own goals and also monitor and regulate the progress in achieving their objectives.

Simulations Imitation of a process that exists, or could exist, in the real world.

Summative Assessment Assessment to grade the work of students. Generally, rather a number than detailed feedback.

T

Team Assignments Assignments that are solved collaboratively by a team. The grading of such assignments should always include a general grade for the performance of the team as well as individual elements for each student.

Current Scientific Reports of the German University of Digital Science

Vol.	DOI	Title	Authors/Editors
1	10.60906/germanuds-h1gg-kd83	**Vision of a University of the Digital Tomorrow - the German University of Digital Science**	M. Friedrichsen and C. Meinel
2	10.60906/germanuds-18vc-6t81	**Innovative Formats for Online Teaching**	C. Meinel, T. Staubitz, and D. Köhler

German University of Digital Science

IMPRESSUM

Series
Scientific reports of the German University of Digital Science

German University of Digital Science
Prof. Dr. Mike Friedrichsen
Prof. Dr. Christoph Meinel

Marlene-Dietrich-Allee 14
14482 Potsdam
office@german-uds.de
www.german-uds.de

ISSN 2943-2030 (Online)
ISSN 2943-2057 (Print)

DOI 10.60906/germanuds-xxdq-e717