Doing Higher Education

Sandra Hummel
Mana-Teresa Donner *Editors*

Student Assessment in Digital and Hybrid Learning Environments

Springer VS

Doing Higher Education

Reihe herausgegeben von

Sandra Hummel · Mana-Teresa Donner
Editors

Student Assessment in Digital and Hybrid Learning Environments

Springer VS

Editors
Sandra Hummel
Center for Scalable Data Analytics
and Artificial Intelligence
TU Dresden
Dresden, Germany

Mana-Teresa Donner
Center for Scalable Data Analytics
and Artificial Intelligence
TU Dresden
Dresden, Germany

ISSN 2524-6380 ISSN 2524-6399 (electronic)
Doing Higher Education
ISBN 978-3-658-42252-3 ISBN 978-3-658-42253-0 (eBook)
https://doi.org/10.1007/978-3-658-42253-0

This Springer VS imprint is published by the registered company Springer Fachmedien Wiesbaden GmbH, part of Springer Nature.
The registered company address is: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Contents

Editors and Contributors

About the Editors

Sandra Hummel is an educational scientist and research group leader at the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) at Dresden Technical University. Her research focuses on human-centered AI-based teaching and learning. She is the coordinator of several EU projects concerned with designing AI tools and technologies that are responsive to the unique needs, goals, and abilities of heterogeneous learners.

Mana-Teresa Donner finished her master studies in *Psychology* and *Adult and Continuing Education* at University of Graz. Her research focuses on learning analytics (project: Learning Analytics—Students in Focus), higher education teaching and on self-regulated learning. She is a PHD candidate (Educational Science) and investigates the learning psychology perspective of AI-based mentoring in the Junior Research Group *Situating AI-based Mentoring* at ScaDS.AI (TU Dresden).

Contributors

Gede Ariadi Management Department, Faculty of Economic and Business, Universitas Kristen Satya Wacana, Salatiga, Indonesia

Johanna Braun Department of Banking and Finance, University of Zurich, Zurich, Switzerland

Nataša Brouwer Faculty of Science, University of Amsterdam, Amsterdam, Netherlands

Corinna Bruckmann Clinical Division of Periodontology, Medical University of Vienna, Vienna, Austria

Julia Dohr Digital Teaching Services, WU Vienna, Vienna, Austria

Mana-Teresa Donner Center for Scalable Data Analytics and Artificial Intelligence, TU Dresden, Dresden, Germany

Alexandra Dorfer Department of Academic Services, University of Graz, Graz, Austria

Andrea Ghoneim Digital Teaching Services, WU Vienna, Vienna, Austria

Hady Haririan Department of Periodontology, Dental Clinic, Faculty of Medicine, Sigmund Freud University, Vienna, Austria

André Heck Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, Netherlands

Julia Hense Society for Media and Competence Research mbH, MMB Institute, Essen, Germany

Nils Hernes Center for Digital Systems (CeDiS), Free University Berlin, Berlin, Germany

Sandra Hummel Center for Scalable Data Analytics and Artificial Intelligence, TU Dresden, Dresden, Germany

Hale Ilgaz The Faculty of Open and Distance Education, Ankara University, Ankara, Turkey

Tomas Meijer College of Life Science, University of Amsterdam, Amsterdam, Netherlands

Florian Mosböck Digital Teaching Services, WU Vienna, Vienna, Austria

Katharina Resch Centre for Teacher Education, University of Vienna, Vienna, Austria

Gudrun Salmhofer Department of Academic Services, University of Graz, Graz, Austria

Lisa Scheer Competence Center for University Teaching, University of Graz, Graz, Austria

Roland Schläfli Department of Banking and Finance, University of Zurich, Zurich, Switzerland

David Schmocker Center for University Teaching and Learning, University of Zurich, Zurich, Switzerland

Alexander Schulz Center for Digital Systems (CeDiS), Freie Universität Berlin, Berlin, Germany

Marthe Schut Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, Netherlands

Michiel van Wijk Institute of Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, Netherlands

Michaela Wagner-Menghin Department of Psychiatry and Psychotherapy, Clinical Division for Social Psychiatry, Medical University of Vienna, Vienna, Austria

Ingrid Wahl Department of Communication, University of Vienna, Vienna, Austria

Dessy Seri Wahyuni Informatics and Education Department, Graduate Program, Ganesha University of Education, Buleleng, Bali, Indonesia

Christa Walenta Ferdinand Porsche FernFH, Wiener Neustadt, Austria

Günther Wenzel Ferdinand Porsche FernFH, Wiener Neustadt, Austria

Benjamin Wilding Department of Banking and Finance, University of Zurich, Zurich, Switzerland

Denizer Yildirim The Faculty of Open and Distance Education, Ankara University, Ankara, Turkey

Introduction

Sandra Hummel and Mana-Teresa Donner

Assessment is a fundamental factor in monitoring the learning process of students and therefore an essential component of effective teaching and learning in the online environment. The COVID-19 pandemic has caused significant disruptions in the higher education sector regarding assessments, because with the need for social distancing and remote learning, online examinations have become essential for evaluating student learning outcomes (Akimov & Malin, 2020). The role of online assessment in higher education during the pandemic has been critical for several reasons: Online assessments have enabled students to continue their studies without interruption, despite the closure of physical campuses and classrooms and have helped to ensure academic integrity by providing secure and reliable platforms for testing and preventing cheating. Furthermore, the possibility to conduct online assessments has provided students with the flexibility to complete assessments at their own pace and in their own time (Hickey & Harris, 2021), which has been particularly important for students who may have had to manage other responsibilities during the pandemic. Digital examinations have allowed also for quick and efficient feedback to students, which has been essential for supporting their learning and addressing any areas of difficulty (Morgenroth & Wieczorek, 2020). However, during the (Corona-induced) wave of digitization,

S. Hummel (✉) · M.-T. Donner
Center for Scalable Data Analytics and Artificial Intelligence, TU Dresden, Dresden, Germany
e-mail: sandra.hummel@tu-dresden.de

M.-T. Donner
e-mail: mana-teresa.donner@tu-dresden.de

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_1

the altered forms of assessment presented us with new challenges, and we have gained new experiences with digital evaluation methods.

The aim of this anthology is to take up the experiences made in the course of the Corona pandemic and to equip us with relevant knowledge for our further teaching. This anthology focuses on new forms of digital assessments and highlight challenges, effective practices and opportunities associated with conducting assessments in digital and hybrid learning environments. Therefore, the articles discuss various aspects to consider when designing scenarios for online exams, including didactic, organizational, technical, and legal factors. Didactic considerations include integrating exams into the overall teaching concept and constructing individual tasks. Organizational factors involve determining the exam process, such as how candidates will log in and how tasks will be created. Technical considerations include selecting suitable exam systems and integrating them into the school's infrastructure. Legal considerations include ensuring legal compliance and accessibility. The roles and responsibilities of each participant in the online exam process need to be defined and incorporated into the overall planning. Various categories of online exams including self-assessment, low cost/feedback, safety, mobile and flexible, massive, gamification/motivation, and adaptive solutions also play a central role offering different possibilities for electronic assessment and learning outcome control, including replacing traditional graded exams. Furthermore, dimensions to consider when implementing online exams, such as the format of tasks, transmission of results, supervision, timeframe, social dimension, and location-specific considerations are presented.

On the structure of this volume

This anthology is composed of two sections: Part A 'General Approaches to Assessments in Digital and Hybrid Environments' includes articles discussing approaches that leverage technology to support student learning and measure learning outcomes effectively. Part B is entitled 'Exam Formats' and deals with assessments measure and provides examples of how digital tools can be used to deliver authentic assessments, such as simulations, case studies, and problem-based learning activities.

The volume opens with a chapter by **Johanna Braun, Roland Schläfli, David Schmocker**, and **Benjamin Wilding**. The authors present their approach towards online-assessment in the domain of Finance. Through Excel-based exercises and examinations (i.e., case studies), students interacted with the topic area in a practice-oriented fashion. An automated grading process and the personalization of exams for individual students allowed for a smooth transition from paper-based exams towards online-assessment and, later, remote open-book examinations. The analyses reveal that online-assessment adds value for lecturers and students in terms of practical relevance and that it can reduce human subjective judgement, improve fairness, and allow lecturers to grade in a more structured and efficient way.

Nils Hernes presents the digital remote assessment format "E-Examinations@Home" deployed at Freie Universität Berlin in 2020. In the case study, the author describes the challenges to digital assessment in higher education during the onset of the covid pandemic as well as the approaches taken to successfully introduce large-scale digital remote assessments. The author focuses on didactic, technological, logistical and legal solutions to provide sound digital remote assessments within a university context.

Alexandra Dorfer, Gudrun Salmhofer, and **Lisa Scheer** make a strong case for student feedback and point to its important contribution in the development of teaching and learning. They discuss the difference between formative student feedback and summative course evaluation alongside guiding questions and considering different teaching scenarios. Theoretical concepts and empirical data are complemented by practical examples of how feedback can be systematically implemented.

Mana-Teresa Donner and **Sandra Hummel** compare digital and face-to-face assessments through a two-stage study design. The first stage involves guided interviews with teachers, and the second stage involved examining examination forms before and after the switch to online examinations due to COVID-19. The study shows that online exams differ especially in exam formats, examination design, strategies of exam realization, and communication with students.

Julia Hense debates how digital assessments can be organised technically and designed according to accepted didactical and administrative standards. According to the author, digital assessments do not only have to follow the requirements of the examination regulations but from a didactic perspective, the question arises, which exam format should be chosen and is compatible with the imparted learning content and according learning objectives and, above all, how the precise exam questions should be designed.

Following the trail of assessment literacy and bearing in mind the specific legal situation in Germany, **Alexander Schulz'** article aims at outlining a basic framework for a literacy of practice for digital remote assessments.

Hale Ilgaz and **Denizer Yildirim** present a design-oriented view within the scope of qualified distance education by considering the e-evaluation design according to the discipline regarding the problems experienced in e-evaluation during the pandemic process. The perspective present in this chapter reflects many years of practical experience in distance education.

Florian Mosböck, Julia Dohr, and **Andrea Ghoneim** present a case study of a successful online-assessment practice in university teaching. It focuses on the support measures that have been implemented in the years 2020/21 by the Vienna University of Economics and Business (WU) to meet the needs of all identified stakeholders, especially students, in written online distance exams.

The case study by **André Heck**, **Marthe Schut, Michiel van Wijk, Tomas Meijer**, and **Nataša Brouwer** discusses the creation and use of cloud-based e-assessment modules in the mathematics and statistics courses for first year students in the bachelor programme Biomedical Sciences at the University of Amsterdam. Two scenarios for e-learning are presented: that of an e-assessment module used alongside a textbook and that of a course completely digitalized in a single cloud-based environment. The authors discuss the use of ICT for learning support and for empowering the learning experience in a design of continuous e-assessment of mathematical and statistical abilities.

Michaela Wagner-Menghin, Corinna Bruckmann, and **Hady Haririan** present an online-learning scenario for acquiring clinical reasoning skills in periodontology. Their study explores the extent to which anonymous aggregated peer-evaluation replicates teachers' evaluation of clinical reasoning performance. The authors conclude that anonymous aggregated peer-evaluation is a feasible formative assessment activity, but should not be used as a summative assessment, as peers hesitate to indicate weaknesses when this translates to a 'failed' decision.

Dessy Seri Wahyuni and **Ariadi Gede** scrutinize the effect of peer assessment in enhancing students' writing, especially scientific papers, in MOOCs (Massive Open Online Courses) in Indonesia. The results show how peer assessment can improve in the undergraduate students' writing skills for scientific proposals.

Katharina Resch presents a study of assessment practices and related student learning in an e-service learning course using weekly learning diaries, group discussions and online presentations at the University of Vienna in Austria. The findings show the usefulness of structured learning diaries for formative assessment whereas needs-based online group discussions do not exhibit the same impact.

Ingrid Wahl, Christa Walenta, and **Günther Wenzel** evaluate the results of an admission test among applicants and the subsequently enrolled students of a distance learning study program. The authors argue that admission tests which reflect the requirements of the intended study program can provide applicants with study-related information and thus contribute to self-selection.

We hope that the diverse approaches discussed in the contributions to this anthology will find suitable frameworks for further scientific development in the future and that readers encounter not only new ideas but also validation for their teaching methods within this volume. We express our gratitude to the authors for their contributions and to the reviewers for their thorough assessments.

Graz, December 2023
Sandra Hummel and Mana-Teresa Donner

References

Akimov, A., & Malin, M. (2020). When old becomes new: A case study of oral examination as an online assessment tool. *Assessment & Evaluation in Higher Education, 45*(8), 1205–1221. https://doi.org/10.1080/02602938.2020.1730301.

Hickey, D., & Harris, T. (2021). Reimagining online grading, assessment, and testing using situated cognition. *Distance Education, 42*(2), 290–309. https://doi.org/10.1080/01587919.2021.1911627.

Morgenroth, C., & Wieczorek, B. (2020). Zwischenbilanz zum Corona-Hochschulrecht aus Sicht der Hochschulpraxis, Teil I – Online-Lehre. *Ordnung der Wissenschaft*. https://ordnungderwissenschaft.de/wp-content/uploads/2020/12/02_01_2020_Morgenroth_Wieczorek.pdf.

Sandra Hummel is an educational scientist and research group leader at the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) at Dresden Technical University. Her research focuses on human-centered AI-based teaching and learning. She is the coordinator of several EU projects concerned with designing AI tools and technologies that are responsive to the unique needs, goals, and abilities of heterogeneous learners.

Mana-Teresa Donner finished her master studies in *Psychology* and *Adult and Continuing Education* at the University of Graz. Her research focuses on learning analytics (project: Learning Analytics—Students in Focus), higher education teaching and on self-regulated learning. She is a Ph.D. candidate (Educational Science) and investigates the learning psychology perspective of AI-based mentoring in the Research Group *Situating AI-based Mentoring* at ScaDS.AI (TU Dresden).

Part I

General Approaches to Assessments in Digital and Hybrid Environments

Experiences and Critical Reflection on Online-Assessment with Excel Case Studies – Review on a Successful Online-Assessment Practice as Well as the Adaptation to a Remote Setting Due to the COVID-19 Pandemic

Johanna Braun, Roland Schläfli, David Schmocker, and Benjamin Wilding

Abstract

While forms of online-assessment have evolved since the 1980s, their use was fuelled by the COVID-19 pandemic. This paper reflects on an online-assessment practice that has been developed at the Department of Banking and Finance (DBF) of the University of Zurich, as well as the necessary adaptation to a remote setting due to COVID-19. Before 2010, exams were conducted in a paper-based setting and often failed to capture the high complexity of lectures. According to the constructive alignment principle, learning objectives, activities, and assessments must be well-aligned. Changing the assessment setting from paper-based to online on university-managed devices enabled

J. Braun (✉) · R. Schläfli · B. Wilding
Department of Banking and Finance, University of Zurich, Zurich, Switzerland
e-mail: Johanna.braun@uzh.ch

R. Schläfli
e-mail: Roland.schlaefli@uzh.ch

B. Wilding
e-mail: Benjamin.wilding@uzh.ch

D. Schmocker
Center for University Teaching and Learning, University of Zurich, Zurich, Switzerland
e-mail: David.schmocker@uzh.ch

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_2

us to account for such criteria more systematically. Our exams were transformed into case studies in Microsoft Excel to resemble practical use cases and were graded using automated methods to create a fair outcome. Due to the outbreak of COVID-19, exams were held remotely in an open-book format, and our approach had to be adapted. To prevent cheating through peer-based collaboration (e.g., sharing exam solutions and results), we started creating personalised use cases with individual parameters and worksheet layouts. We strongly believe that online-assessments add value for lecturers and students. Exams were redesigned in terms of content, became more application-oriented, and increasingly allowed cross-linked knowledge to be tested. Furthermore, the automated correction and personalisation of exams reduce human subjective judgement, improve fairness, and allow lecturers to grade in a structured and efficient way.

Keywords

Automated grading • Excel case-study • Personalised exam

1 Introduction

At the Department of Banking and Finance (DBF) of the University of Zurich (Switzerland), we have been using online-assessment with Microsoft Excel (referred to as Excel in the remainder of this text) as an examination form since spring 2010. Compared to traditional paper-based examinations, online-assessments capture lecture contents of higher complexity, especially in practice-oriented courses. Furthermore, online-assessments allow for fairer grading procedures, better evaluation objectivity, and more efficient marking. In this paper, we first introduce the guiding didactic aspects and practical thoughts that lead us to use online-assessments based on Excel as an examination form. We then describe the development of our on-site online-assessment scenario and the various adaptations stemming from the COVID-19 pandemic. The paper concludes with a critical reflection on important learnings and possible avenues for the further development of online-assessments using Excel.

2 Didactic Concept and Literature

The didactic concept in our online-assessment courses is based on three key elements: a constructive alignment between learning materials, exercises, and examination; a focus on practical skills that resemble tasks in work life; and a fair examination and grading process through the use of automation. This section will focus on the above-mentioned qualities of our concept.

2.1 Constructive Alignment

According to the constructive alignment principle, it is of great importance that learning objectives, teaching and learning activities, and assessments be closely aligned (Biggs, 2003). All three elements should reach similar levels of complexity and need to be developed in coordination with each other. Alignment in teaching offers several advantages for students and lecturers, the most striking of which are that it helps students better understand the objectives of a course and avoids misunderstandings between teachers and students. Furthermore, the concept of constructive alignment supports a central requirement of the Bologna Reform, namely, a competence-oriented design of teaching as activities are being aligned with learning objectives (e.g., Schaper et al., 2012).

Constructive alignment entails mirroring, in the exam, those competencies and skills that were acquired during the course. As such, an examination “fit for purpose” includes complex tasks that have been previously addressed and uses methods comparable to those taught throughout the course. Because content that is taught and learnt at the university level is often complex and demanding, exam tasks must be designed accordingly. This correlates with efforts aimed at developing examination tasks at higher levels of learning (cf. Bloom, 1973; Anderson & Krathwohl, 2001), by, for example, integrating application, analysis, and synthesis activities. New assessment forms such as online-assessments are better suited to these didactic demands. Tasks in handwritten exams which must be solved using a calculator are time-consuming and, due to these time constraints, often focus on lower-order thinking skills. Examinations in the online-assessment setting incorporate higher-order cognitive tasks, as lower-order activities like simple calculations may be carried out with fitting tools.

In our courses, the examination and mandatory exercises are aligned with this format, providing students with the opportunity to familiarise themselves with the tools over the semester and then apply their skills in the same environment during the exam. In addition to exercises, Excel examples are used throughout the course as part of the e-Learning content to underline the importance of financial use cases and support key learning goals (e.g., specific calculations).

2.2 Focus on Practical Skills

In their work life, students will be required to perform calculations not only correctly, but also efficiently, often using specific software. On top of understanding the theoretical aspect of calculations and performing them, students must acquire the digital skills needed to use such software. University courses often implicitly require students to learn software that can be applied to complex calculation tasks (e.g., Excel or R), but do not necessarily reward students for these learnings, as final exams are usually paper-based.[1]

The use of practical case study exercises in calculation software is one way in which theoretical and practical knowledge can be combined to improve the alignment between theoretical exercises and practical skills. In case study exercises, students are presented with a specific and real-life-related "case" or story describing a problem resembling a practical situation in the industry (Kaiser, 1983). The task is to find a solution, as well as to reflect on the outcome by using both theoretical knowledge and practical skills.

2.3 Fairness in Correction Through Automation

A key requirement when performing any kind of exam is to guarantee a valid and fair grading process (Liu & Zhang, 2020). Studies show that students highly value equal and fair treatment (e.g., Emeka & Zilles, 2020). Thus, procedural fairness—referring to the perception of correctability and consistency (Gordon & Fay, 2010)—is important in the design, conduction, and correction process of any exam. Correction through automatisation can contribute to this in several ways.

Firstly, when designing online exams using automated correction, lecturers should not only invest in the creation of exam questions but also define the passing criteria early in the design process. Though this practice is also recommended for all other exam forms, it can be neglected more easily when exams are, for example, corrected manually, as adaptations to the passing criteria are quick to be made. From a didactic viewpoint, more options to adapt the grading process post-exam may threaten the evaluation's objectivity. The second reason applies to exams with many participants in particular, but can also be true for smaller exam formats. To process numerous exams in a timely manner, the marking workload is often shared between multiple people. Even if the evaluation criteria are defined

[1] There are few courses in which students hand in exercises performed with specific software as a requirement for the final exam and/or contribution towards the final paper-based exam.

in advance and clearly communicated to the examiners, factors like interindividual perceptions, experiences, or working methods may skew the evaluation process. In particular, the rating of individual tasks shared between different people can lead to deviations and unfair rating processes with lower evaluation objectivity. The use of automated correction helps overcome differences in individual correction and renders the marking process more fair. Thirdly, it can be appropriately argued that the personal constitution of the person marking and the associated error rate is unstable, causing potential procedural unfairness. Even if the same person corrects an entire exam or a single task, their attention span and motivation can vary greatly throughout the process.

3 On-Site Online-Assessment Using Managed Devices

Over the past 10 years, the DBF has performed several exams per year in an on-site online-assessment setting. Before 2010, all exams at the DBF were conducted as traditional paper-based exams. In this exam setting, all calculation steps were carried out on paper with a calculator. The questions consisted mainly of open text questions (e.g., "What are synergies in the context of Mergers and Acquisitions?"), manual calculations, and an explanation of the results obtained (e.g., "Calculate the theoretical value per share in both cases and compare this with the current share price of the company. What can you conclude?"). The manual calculations were often very repetitive, meaning that students had to repeat certain calculation steps throughout the exam to receive the end result (e.g., discounting and adding cash-flows).

These calculation steps were time-consuming and the exams, due to time constraints (maximum exam time of two hours), often focused on questions with lower-order thinking skills. Additionally, practitioners working in the firm valuation field regularly analyse balance sheets and income statements, create financial statements and forecasts, and perform various valuation methods, predominantly in Microsoft Excel. Thus, there was a misalignment between the exam contents (including the teaching materials) and the real-life applications in a work environment. In 2010, a first course (i.e., "Valuation of the Firm") was transformed into a digital case study exam in Microsoft Excel. Restructuring the exam improved the alignment between examination and practice and enabled us to pose complex analysis tasks. To ensure that students were well-prepared for this new type of examination in Excel and familiar with the question types and technology used, Excel case studies were introduced as learning activities throughout the course. These learning activities, in the form of formative and summative assessments, assured that exercise and exam formats corresponded to the constructive alignment principle. This alignment of the course content with the exam also highly influences students' preferences for a specific assessment format (i.e., online-assessments), as is shown in Chap. 5.

After the initial success of the online-assessment practice and several small technical and organisational adjustments (e.g., adjusting Excel sheets to different laptop screen sizes, standardising the training of proctors on how to proceed with technical problems, providing more information on keyboard layout and screenshots of the online-assessment environment), a second course, called "Corporate Finance 2", was adapted to the case study setting in 2017. Similar to the first course, the calculations were carried out in Microsoft Excel or another calculation software used in a professional environment. Additionally, we were able to add calculations that were not at all possible to perform in the previous handwritten exam setting due to their requiring numerical optimisation. The previous exam was severely constrained in content and feasibility. These two courses were the only ones tested in this online-assessment setting until 2020.

3.1 Online-Assessment Environments at the University of Zurich

Before the pandemic, during the exam period at the University of Zurich, three lecture halls were transformed into dedicated online-assessment rooms. This allowed all faculties at the university to hold their online-assessment exams in a controlled environment with a steady technical setup, which was centrally supported by the university's informatics service. Each online-assessment was performed on-site in rooms with a capacity of between 32 and 233 students. Students were supervised by lecturers and assistants, similar to traditional paper-based exams. Each student was provided with a university-owned laptop in which was installed the necessary software for the exam (e.g., Microsoft Excel), but which was disconnected from any networks besides the Learning Management System (LMS).[2] The exams were closed-book in general, as cheating and collaboration could be reliably prevented thanks to the supervision and controlled technical equipment. See Fig. 1 for an exemplary student workspace in this examination setup.

Though the controlled and centrally organised workplaces came with advantages, our lack of influence over the technical setup (as provided by the

[2] OLAT at UZH, https://www.olat.uzh.ch.

Fig. 1 Examination setup in the lecture hall

university's informatics service) meant that the applications and devices quickly became outdated in comparison to the student's own devices and working environments. These differences meant that students were left to work with systems different from those they were used to on their private devices, making it especially challenging for users of operating systems other than Windows (e.g., macOS). This critique is reflected in the student survey discussed in Chap. 5. Especially in open-ended questions, students complained about other operating systems, outdated Excel versions, different keyboards, a lack of shortcuts, and other issues. Therefore, this pre-defined setup was not always perceived as fair by the students, even when communicated clearly in advance. Some students had to adjust to the presented settings during the examination, whereas other students did not have to adjust at all. This gave the latter group a perceived, albeit small, advantage.

3.2 Case Studies in Excel and Question Types

To create an exam centred around practical analysis tasks in the Banking and Finance industry, we started creating longer case studies consisting of multiple exercises, each of which is broken down into several smaller tasks. An exam usually consists of two to seven exercises. We guide the students through the exercises and provide a predefined structure aid in solving the tasks. Smaller tasks are most often linked to each other, meaning that students must continue based on the calculated results of previous tasks (within the same exercise) and link later calculations back to previous results, as is the case in regular work life when one is using Excel and templates. Students must use appropriate Excel formulae (e.g., = F3 + G3 or = sum(F3:G3)) to solve tasks correctly.

A key advantage of having students provide formulae is that we can see each step of their calculation, rather than just a solution that is either right or wrong. In Excel, students can perform more complex calculations in one step (e.g., calculate the standard deviation of given values by entering = STDEV()) as opposed to multiple steps on a calculator. Furthermore, Excel allows us to include tasks that are too complex for manual calculation (e.g., computing matrices). When one is working in Excel, repetitive calculation steps that provide no additional value for assessment purposes can simply be copied or autocompleted, e.g., performing a calculation five times for five consecutive years. Also, in Excel, students can easily copy the formula or drag to fill.

Exercises usually start with a description of the initial situation (e.g., the balance sheet and income statement of a fictional company) and often introduce a small story that is worked with throughout said exercise (e.g., “Company A produces X and wants to derive the valuation of one of its segments.”). This starting situation is analysed through tasks of increasing depth and, often, difficulty (e.g., first some simple formulae, then more complex combinations with a final output or judgement). Figure 2 showcases an exemplary exam exercise from one of our online-assessment courses. To examine different thinking patterns, question types are varied throughout the case studies. Closed-form questions consist of numerical results computed in Excel (see N in Fig. 2) or manual selections from a set of options (see D in Fig. 2). They usually have either one or multiple correct solutions. Contrarily, open-form questions consist of a free-form input (see F in Fig. 2) in which students can respond in text form. Open-form questions are often used in combination with closed-form questions to ask for theoretically founded explanations of a performed calculation or to have students estimate or judge a value. We also use more advanced question types that require the application of numerical optimisation (i.e., using the Excel solver), primarily when the concept of the optimisation is crucial for overall understanding (e.g., in portfolio optimisation). Though these advanced question types require more sophisticated tools to compute, their outputs are usually seen as some of the more primitive question types when grading. To prevent a scenario in which students are overtaxed with the online-assessment setting and the transfer of theoretical concepts to Excel, our online-assessment courses are graded in part based on preparatory exercises (what we call "Involving Activities"), which students solve during the semester preceding the exam. Excel formulae and working with the tool itself must be practised in an environment resembling that of the real exam. Furthermore, the Involving Activities are graded to provide feedback and are awarded exam points upon successful completion.

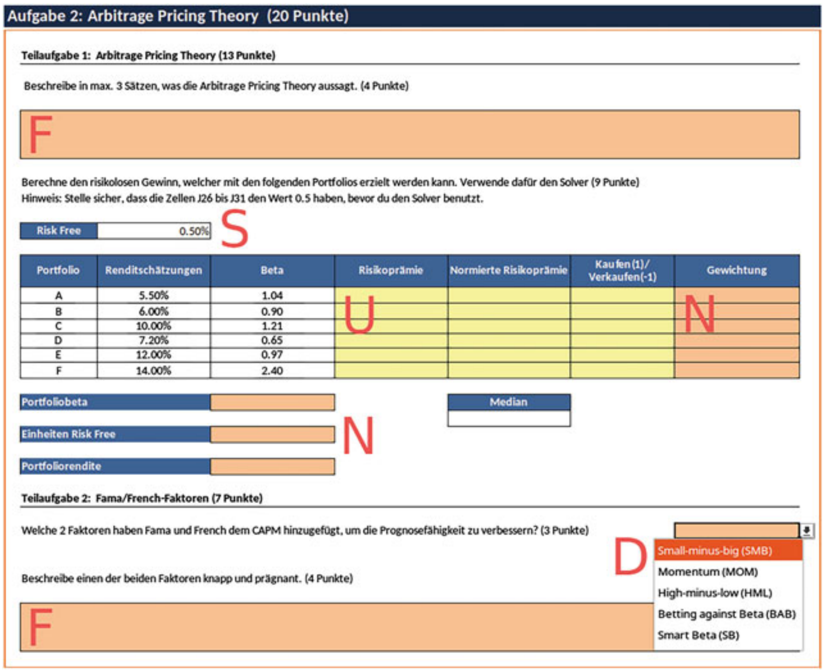


Fig. 2 Case-study example. (*Note.* Exercise of an investment course with different question types and multiple subtasks. **S:** given starting value for computations; **U:** ungraded parts of the exercise (intermediary calculations); **N:** numerical computations using formulae; **D:** drop-down selection from a set of options; F: free-text responses (e.g., explanations))

3.3 Consequential Errors

Long-form exercises in a case study are often based on incremental tasks, meaning that a value is computed and that the next task is a derivation of the result of the preceding one. This leads to challenges in automated grading based on numerical equivalence due to consequential errors (i.e., if A is wrong but B is derived correctly from A, then B is graded wrong as well). Such consequential errors must be accounted for in grading, or their occurrence must be prevented—the latter of which is very hard to achieve in practice. Figure 3 visualises an example of what a consequential error may look like.

Our approach to grading while accounting for consequential errors consists of checking whether an exercise is wrong based on exact numerical results. If so, we use the results of the student for all exercises preceding the one we are grading (by backfilling the sample solution with the student's previous results) and apply the sample solution formulae of the exercise being graded to these results (of which we know that they will compute the correct result). If the resulting numbers match the student's submission, points are awarded based on consequential errors. This approach comes with the downside of being computationally expensive and potentially more error-prone, as it involves many Excel interactions when results

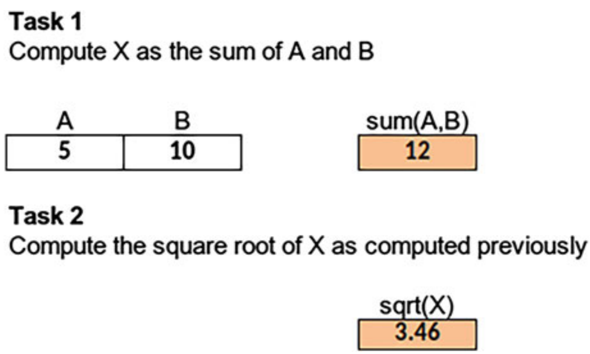


Fig. 3 Exemplary scenario of a consequential error. (*Note.* This illustration showcases a scenario of consequential errors. Task 1 has been computed incorrectly (answer 12, correct 15). Task 2 has been calculated correctly (i.e., one square root of 12 is 3.46). However, when grading naively by comparing numbers, one would expect Task 2 to result in 3.87, which is the square root of 15. The mistake made in Task 1 would thus inappropriately affect the grading of Task 2)

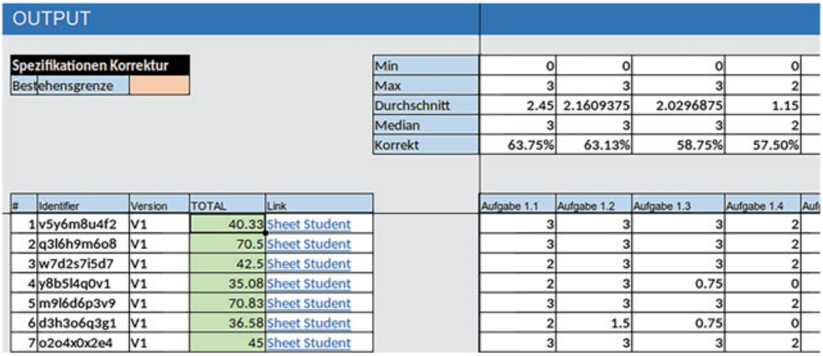
are transferred between different workbooks. Grading may therefore take several hours depending on the number of students.

3.4 Automated Grading

While Excel-based case studies could, theoretically, be graded manually, this would make a fair and consistent grading process very inefficient. Therefore, we evaluated automated approaches to help us grade our case study exams (Fig. 4). Early approaches to such automated gradings were based on Visual Basic for Applications (VBA) programming. The tight integration with Excel allowed us to perform all programming and run the grading procedure directly within Excel. However, the use of VBA had several drawbacks: Consequential errors had to be defined manually as Excel formulae, which required a lot of preparatory effort, and it was not feasible to incorporate the creation of Excel files for distribution within the program; student submissions were to be preprocessed with separate tools to allow the Excel program to ingest them.

To alleviate these issues, a new version of our grading program was developed using the Python programming language. The new program interacts with Excel through Python libraries and is not directly coupled to Excel, allowing us to incorporate further steps of the examination process. For example, the generation of hundreds of exams based on different versions and the automated grading of consequential errors can be directly integrated into the program. Working with Python proved especially helpful during the pandemic, as we could quickly extend the software based on new requirements by using our existing software libraries.

A general challenge across previously described grading programs is that different types of questions must be treated with different approaches. While closed-form questions have a clear, easily comparable, and potentially rounded numerical result, open-form questions must still be graded manually. To simplify this process, our programs generate an Excel worksheet for each open-ended question, showing all responses in an anonymised format. Lecturers grade all responses per question in one go, reducing the risk of the bias that may have otherwise occurred—thus providing a fairer grading process.



OUTPUT

Spezifikationen Korrektur	
Bestehensgrenze	

Min	0	0	0	0
Max	3	3	3	2
Durchschnitt	2.45	2.1609375	2.0296875	1.15
Median	3	3	3	2
Korrekt	63.75%	63.13%	58.75%	57.50%

#	Identifier	Version	TOTAL	Link	Aufgabe 1.1	Aufgabe 1.2	Aufgabe 1.3	Aufgabe 1.4	Auf
1	v5y6m8u4f2	V1	40.33	Sheet Student	3	3	3	2	
2	q3l6h9m6o8	V1	70.5	Sheet Student	3	3	3	2	
3	w7d2s7i5d7	V1	42.5	Sheet Student	2	3	3	2	
4	y8b5l4q0v1	V1	35.08	Sheet Student	2	3	0.75	0	
5	m9l6d6p3v9	V1	70.83	Sheet Student	3	3	3	2	
6	d3h3o6q3g1	V1	36.58	Sheet Student	2	1.5	0.75	0	
7	o2o4x0x2e4	V1	45	Sheet Student	3	3	3	2	

Fig. 4 Output sheet of a corrected exam (anonymised)

4 Personalised Remote Online-Assessment (with BYOD)

In early 2020, due to the pandemic, all exams were held in an online and open-book setting, as the UZH no longer permitted examinations in lecture halls. Thus, the previous online-assessment approach was no longer viable and had to be adapted to this new setting. Exams had to be distributed to students via the LMS and solved at home on students' own devices. In our specific case, each student was provided with an Excel file that they could download, solve on their device and submit within the examination time frame. This approach introduced new challenges compared to the on-site setting. Firstly, as students worked on their private devices, issues in their environment could arise. For example, their version of Excel could be outdated or their browser could be set up in a manner that did not show the examination environment correctly. However, the critique regarding different operating systems, language settings, etc. was resolved by allowing students to choose the operating system and language settings themselves, as described in Chap. 5. Secondly, access to the learning materials as well as the internet could not be prevented during the exam. Therefore, the exams had to be constructed as open-book exams, focusing less on knowing the material by heart and more on cross-linked knowledge and practical cases that could not simply be looked up online. Questions were therefore adjusted even further towards higher levels according to Bloom's taxonomy (cf. Bloom, 1973), increasingly allowing for cross-linked knowledge to be tested. Lastly, it was not possible to effectively supervise students[3] and we were unable to control the sharing of exam solutions or communication between students during the exam, even if it was explicitly forbidden according to the UZH honour code. This made it difficult to measure individual performance and may threaten a reliable assessment.

To mitigate the challenges introduced by the remote exam setting, we adapted our exam procedures and added new technical countermeasures. To ensure that student environments are compatible with the exam, we now provide detailed guidelines on how the exam works and on the responsibilities of students regarding the necessary technical preparations. Before each exam, we perform a "test run" (often a mock exam) that allows students to familiarise themselves with the infrastructure and the process of downloading and uploading exam files. Furthermore, we rebuilt our exams in the form of open-book exams designed to be as "unshareable" as possible and added measures to prevent and detect cheating. The remainder of this section provides a concise overview of these measures.

4.1 Prevention of Cheating

To hinder the sharing of exam solutions, we create and distribute personalised Excel workbooks to all students, meaning that each student works on their own individual version of the exam. To create these personalised exams, we automatically adjust certain pre-defined input parameters (i.e., cells in Excel as seen in Fig. 2, cell S), after which the remainder of the exam is recomputed using the Excel formulae in the sample solution. For example, we might generate the input parameter "cost of equity" based on a random function that generates either a number inside a pre-defined range (e.g., 10.00–15.00%, rounded to two digits) or randomly selects one from various choices (e.g., choose one number out of the following options: 10.0, 10.5, 11.0, 11.5%). This way, every student has different initial values and will therefore have different numerical results, lowering the incentive to simply share values (i.e., students could share the values, but they might not be correct). This approach works especially well when multiple values are randomised and the number of possible permutations increases. Personalisation is not limited to numbers, but can also be applied to words (e.g., company names), allowing for a clear distinction even in open-ended questions. Students

[3] The laws governing our university allow proctoring with Zoom. However, the student can be observed only frontally. Filming the surrounding room, or filming the student from behind or the side, as well as capturing the student's computer screen, is prohibited. Additionally, recordings of such proctoring are not permitted. Therefore, supervision is very limited and not very efficient.

must argue based on different firms with different numbers, which means that their results are not compatible. However, to promote fairness among the students and prevent invalid outcomes (e.g., a negative stock price), it is important to randomly choose from a pool of didactically sound and practically meaningful values and ranges. These personalised initial values must make sense so that all students can interprete in the same way. An additional variation that reduces the possibility of cheating in an exam is requiring students to estimate certain input values themselves (e.g., "How high do you think the market risk premium will be in the next 10 years?"). A value that is not clearly defined is suitable as an estimated value. A realistic range in the example question is between 4 and 7%. Ideally, the estimate must be justified by a text so that we can check whether the estimate was made by chance or by knowledge. On the one hand, this ensures that all students have different initial values; on the other hand, it enables us to check whether the students are capable of making realistic estimates. Through individual and automated correction, even such exercises can be corrected quickly and efficiently.

Overall, the exam must be personalised with caution, since personalisation increases the number of possible permutations and, thus, the workload for lecturers and assistants. To minimise the risk of invalid outcomes, it is advised to personalise only those few key parameters that change as many consecutive results as possible, instead of personalizing all numbers in the initial task description.

Despite the personalisation of values as introduced above, the tasks are consistent in structure and students still work with the same formulae as their colleagues (e.g., F3 + F4). Formulae can be shared effortlessly among students, making it tempting to simply copy and paste formulae from others to achieve a correct result, even in a completely personalized setting. To mitigate this, we add empty hidden rows and columns to each Excel sheet, varying by exercise and student. This leads to different formulae: e.g., F3 + F4 in the exercise of one student would correspond to D2 + D3 in the exercise of another. Overall, this greatly increases the efforts needed to share formulae, as students would have to adjust the formula to match their personal sheet structure, costing them time that they could otherwise use to solve the exam. The individualisation of formulae can be performed without having to consider didactic aspects, as an empty hidden row or column does not change the exam content if applied correctly. However, it would result in unfair exams if the empty rows or columns were located within the exercise so that only certain, rather than all, students can use Excel's autocomplete function. These two steps of personalisation (varying initial values and adding hidden rows and columns) are performed in the large majority of the exercises and worksheets in our exam workbooks.

When creating personalised exam versions for each student, we first generate a personalised solution workbook with the randomised initial values and insert rows and cells as previously described. This personalised solution workbook is then transformed into the exam file by removing all the values that should be derived by the student. Preparing these two exam files for each student allows us to account for consequential errors even though we are working with personalized workbooks, as we can extract formulae from the matching sample solution when deriving consequential errors.

These two measures of preventing cheating along with the focus on open-book examination questions make collaboration between students more time-consuming, thus decreasing the likelihood of achieving a good grade through cheating. These measures are fortified by the students' intense time pressure during the exam. The outcome of exams incorporating these actions is, therefore, more reliable and fair to all students.

4.2 Detection of Cheating

The design of an exam can reduce cheating attempts, as it makes collaboration more time-consuming and cheating less attractive overall. However, cheating cannot be completely ruled out, as students could always try to work around the systems in place. In addition to our cheating prevention measures, we have implemented means of detecting cheating, with a special focus on the sharing of entire Excel workbooks or worksheets. We include personalised watermarks in visible and hidden places throughout all Excel worksheets. This allows us to track whether students hand in "their" assigned workbook or whether they (partially) copied and uploaded the Excel file of a colleague. This measure is easily implemented and can be verified automatically while grading the exam, as we can simply read the watermarks for each exam and check for any identical values across students. Furthermore, attempting to copy a workbook from a colleague would be of no use to students, as their exams are still graded against their own sample solution with a personalized structure and numbers. If the structure and numbers do not match their sample solution, no points are awarded.

5 Reflection on the Student's Perspective

Several stakeholders are involved in the process of designing, conducting, and correcting an exam. These include lecturers, assistants, people responsible for the technical implementation in the online-assessment environment, and, ultimately, the students taking the exam. Students are key stakeholders, making it crucial to evaluate their perspective on the exam design and process. Exams should be well-designed and challenging but solvable so that performance may be measured objectively. The marking process should be fair and without biases, and feedback on the exam collected and reviewed as quickly as possible after the exam. Additionally, students prefer to compare themselves to their peers (peer distribution). This part of the paper will focus on the survey results we have collected from the students' perspective, specifically regarding online-assessment exams.

From 2011 to 2021, we conducted nine anonymised surveys with the Banking and Finance Bachelor students who participated in the on-site (up to 2020) or remote (after 2020) online-assessment examinations. The questions on the survey changed slightly from the on-site to remote settings, however, the questions in Table 1 were asked in both settings. In total, 354 students answered the surveys, corresponding to an average response rate of 25%. The surveys were distributed after the exam via our LMS to all students enrolled in the courses before they received their grades. The surveys consisted of single-choice questions as well as open text questions. Table 1 gives an overview of the questions asked.

As we see a clear shift in the students' perspectives due to the forced change to remote online-assessments during 2020 and the spring semester of 2021, we split this section into two parts (student feedback on on-site online assessment using managed devices and student feedback on personalised remote online-assessments (with BYOD)) and highlight the changes.

5.1 Student Feedback on On-Site Online-Assessment Using Managed Devices

5.1.1 Self-perceived Level of Computer Skill

The online-assessment exams are digital case studies in Microsoft Excel. Therefore, computer skills, especially in Office tools such as Excel, are relevant in the exam context and potentially influence student perception of online-assessment exams. It was important for us to know whether students felt comfortable handling computer-based tasks in Excel. The average self-perceived level of computer skill and computer knowledge (Q1 in the questionnaire) was high, with an

Table 1 Survey questions 2011–2021

No	Questions	Answer options
Q1	How advanced are your computer skills (in your opinion)?	1–6
Q2	Which operating system do you use at home/privately?	Windows, Mac, Linux, other
Q3	What form of exam do you generally prefer for exams?	Online-assessment, on paper
Q4	Which form of exam do you prefer for this specific lecture?	Online-assessment, on paper
Q5	Please justify the answer to the previous question about the preference.	Open question
Q6	Did you feel well-informed in advance about the technical and organisational conditions of the exam?	1–6
Q7	How did you like the organisation of the exam?	1–6
Q8	How did you like the technical support during the exam?	1–6
Q9	It was easy for me to concentrate during the exam	1–6
Q10	Where there any aspects of the exam that disturbed your concentration?	Open question
Q11	How do you rate the user-friendliness of the OLAT login process?	1–6
Q12	How do you rate the usability of the download and the upload of the Excel file in OLAT?	1–6
Q13	How do you rate the design of the Excel tasks?	1–6
Q14	Did you have any technical problems when solving the Excel tasks?	Yes, no
Q15	If you had technical problems when solving the Excel tasks, please explain them.	Open question
Q16	Do you have any comments on usability in general?	Open question

(continued)

Table 1 (continued)

No	Questions	Answer options
Q17	In your opinion, how can the examination/examination process be improved?	Open question
Q18	Do you have any further comments on the exam or its implementation?	Open question

Note. This table gives an overview of the structure of our surveys. Column 1 numbers the questions, column 2 states the questions asked (translated from German to English), and column 3 shows the answer options provided. For numerical ranges, six is the highest/best rating and one is the lowest/worst rating. All survey questions provided here were asked in all surveys; however, several other questions were added or dropped in certain years

average rating of 4.91 (with 6 being the highest rating and 1 the lowest rating). In total, 76% of the students answered "very good" or "good" (ratings of 6 and 5, respectively); 20% of all students answered the question with "average" (a rating of 4); and only 4% stated that their computer skills were bad (ratings of 3, 2, and 1 respectively). The distribution of the ratings did not change significantly from 2011 to 2018, so there is no clear trend towards a higher self-perceived level of computer skill. The majority of the students (57%) personally used Windows, around 43% used Mac, and only 1% used Linux as their operating system. The exams were taken on university-owned laptops with Windows as the operating system, corresponding to the personal environments of the majority of students. We can conclude that students meet the basic knowledge and skill requirements to participate in online-assessment exams, though with a small advantage for Windows users.

5.1.2 Preferred Form of Exam

Though students may fulfil the prerequisites for participation in online-assessment exams, this does not imply that they like, let alone prefer or even see any added value in, online-assessments as compared to a traditional paper-based exam setting. Therefore, we were interested in knowing the form of exam generally preferred by students. As visualised in Fig. 5 (left pie chart), only 19% of students preferred an online-assessment exam to traditional paper-based exams. Reasons for this can be found in Q5 of the questionnaire. Some students feared that they would fall victim to technical problems affecting their exam score (e.g., being unable to save or upload data), while others said that online-assessment exams were new to them, so they did not know what to expect and felt uncomfortable in the new exam environment. Other reasons were our specific exam setting in terms of a lack of knowledge in Excel (i.e., formulae), having a better overview on paper, and being able to perform more interim calculation steps (as many as the student wants), as well as having enough space for one's own notes. Some believed that Excel exams are extremely error-prone (click on a wrong cell and delete it; mistakenly use A1 instead of A2). The reasons why various students preferred online-assessment exams included the fact that the contents of the lecture and the exam were closely aligned because of better-suited question types. The Excel exams also correspond better to the real world (i.e., how work is performed in jobs). Complex calculations in Excel are handled very conveniently by linking cells, and the focus of the exam is on the student's understanding of the material instead of their calculator skills. According to Fig. 5, online-assessments were widely favoured for these courses, with a majority of 73%. This enforces our conclusion that they are the more suitable option in this case and clearly shows that preparing students for the assessment format, as well as aligning the course contents, exercises, and exam setup, critically influence the students' preference regarding the exam setting.

The influence on student preference can also be illustrated by a single case study (see Fig. 6). In 2017 we decided to change the final exam of the course "Corporate Finance 2" to an online-assessment exam without simultaneously

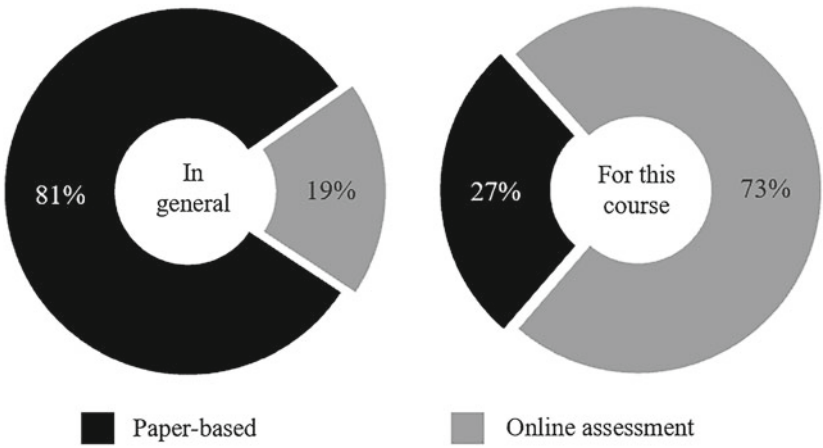


Fig. 5 Preferred form of examinations before 2020. (*Note.* These two pie charts show the average answers for Q3 and Q4 of students based on our survey from 2011 to 2018; n = 177, answers based on 7 surveys)

updating any e-learning materials in the LMS. Therefore, the course materials and the final exam were not sufficiently aligned; while the e-learning of the course focused on readings, multiple-choice questions, and a few Excel files, the final exam was held as a full Excel case study. The percentage of students favouring the online-assessment exam in this scenario was only 43% (n = 30, response rate = 51%). In the following year, we restructured the e-learning course to be more closely aligned with the online-assessment exam and enhanced the learning contents with additional Excel tasks. The percentage of students favouring the online-assessment rose to 88% for this specific course (n = 23, response rate = 22%). The alignment is therefore crucial for the acceptance of the students.

5.1.3 Design of the Exam

When we started with online-assessment exams using Excel, the layout and structure of our first case study were not perceived very positively (average rating of 3.33 on a scale of 1–6). We therefore gradually adapted the layout of the Excel files. In a first step, we adjusted the dimensions of the Excel exercises to fit the screen of the exam computers, minimising scrolling for the students. Additionally, we inserted more space for personal use to allow students to work more independently despite the fixed structure in the Excel sheets. We were able to gradually increase the score to an average of 4.57 in 2018.

5.2 Student Feedback on Personalised Remote Online-Assessments (with BYOD)

5.2.1 Self-perceived Level of Computer Skill

As opposed to the setting before the pandemic, in the remote setting, all students used their own devices to participate in the online-assessment. This resolved many issues with infrastructure, as students could work with the operating system, shortcuts, Excel version and language they had practised with. This is substantiated by the fact that we received much fewer comments from students worried about the infrastructure. However, we do not see that students feel more comfortable about their self-perceived level of computer skill, as the score dropped to 4.69 (with 6 being the highest rating and 1 the lowest rating). This could be due to the fact that students in the remote setting were responsible for their overall setup (including the internet connection, power supply, potential Zoom proctoring, and functioning Excel version), which the university previously took care of. Therefore, the level of complexity might have risen overall, leading to a lower self-perceived level of computer skill.

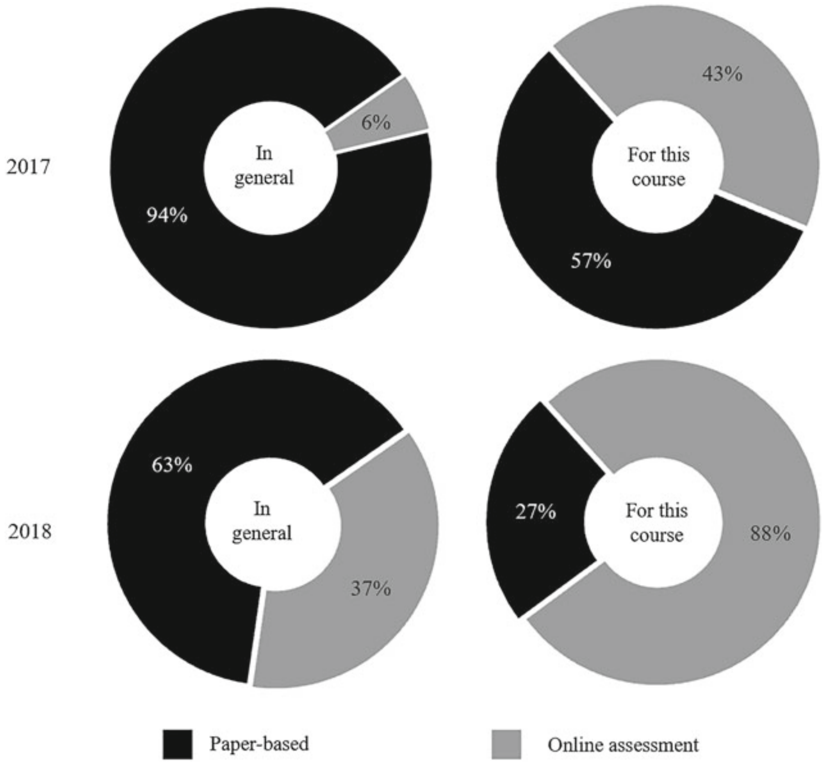


Fig. 6 Preferred form of examinations: specific case-study. (*Note.* These four pie charts show the answers for Q3 and Q4 of students of the course "Corporate Finance 2" in the years 2017 (pie charts above) and 2018 (pie charts below) based on our survey; n = 53, answers based on 2 surveys)

5.2.2 Preferred Form of Exam

Because universities and students were forced into online-assessment exams during the year 2020 and the spring semester of 2021, students grew accustomed to the remote setting and might have also experienced the advantages of online-assessment exams. As a result, we see a significant change in their perspective regarding the preferred form of exams. The majority of students (68% in general and 92% for the specific courses) prefer online-assessment to traditional Paper-based exams, as seen in Fig. 7. Reasons provided by students for and against online-assessment exams (here: Excel case study exams) can be found in the

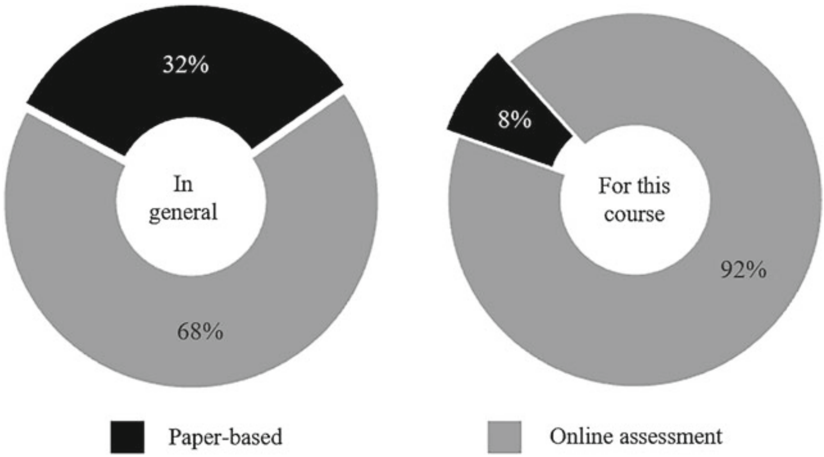


Fig. 7 Preferred form of examinations after 2020. (*Note.* These two pie charts show the answers for Q3 and Q4 of students based on our survey results in 2021; n = 177, answers based on 2 surveys)

open-ended question of the evaluation (Q8). Contrary to the exam setting before the pandemic, students preferred the familiar location at home, alone, where they were not in the middle of nervous students. They also liked the aspect of open-book exams (less learning by heart). In addition, students largely preferred the Excel exam setting, because of its focus on practical skills and because the exams can capture a broader spectrum of knowledge (i.e., Excel can do the basic calculations, so more complex exercises can be carried out). While the majority of students liked the fact that the exams focused on the "application" of knowledge, some wished for more questions focusing on theory. Thus, we might still have a remaining mismatch between teaching materials (theoretical concepts) and exercises and assessments (with mainly practical questions). Potential improvements include a more intuitive way of implementing theoretical constructs like multiple- and single-choice questions within Excel, as well as less time pressure (i.e., to allow students to finish the exam in time).

5.2.3 Design of the Exam

Throughout the years, the exam design and contents have continuously been improved based on student feedback. In the latest evaluation (fall semester 2021) of the design and user-friendliness of the exams, our Excel-based examinations received an average rating of 5.22. The integration of open-ended questions into the Excel exams was rated worst of all factors (on average, 5.0) primarily because of the limited formatting options when one is writing free-hand text in Microsoft Excel. Calculation tasks with formulas (on average, 5.25), single-choice questions (on average 5.17), and the overall design of the tasks (on average, 5.29) were similarly well-liked.

6 Critical Reflection and Discussion

This paper describes the development and use of an online-assessment software based on Excel. Over the last 10 years, this software has been implemented in an on-site online-assessment setting at the Department of Banking and Finance of the University of Zurich and was applied to several exams with more than one hundred participants each year. During the COVID-19 pandemic, we very quickly adapted to a remote, open-book exam setting. While this implementation was a success overall, some critical points require improvement in the future.

Alignment between learning activities, exam design, and the tasks that students might perform in work life ensures that students are confident that they are learning skills that can be practically applied and useful in the future. Our surveys show that this alignment is reasonably close in our case. However, there are still areas where we can improve alignment, in particular by making the learning objectives more transparent. For example, we could incorporate learning objectives into the case study format, ensuring that students have an overview of the information they should have learnt after completing a case study exercise. This is not yet done in a standardised way.

As stated and described before, we cannot rule out every possibility for cheating. Students could, for example, still discuss calculation steps (e.g., multiply price times quantity) to simplify their efforts and make grades less reliable. One possible further development is to generate values based on different distributions so that students will not know how the values behave (e.g., Gaussian distribution vs. Poisson distribution, etc.) (as seen in Rusak & Yan, 2021). A question bank could potentially further improve the above-mentioned problem of cheating. In our current exam setup, all students work based on the same exam questions, but receive their own version of these contents based on personalised numbers. A strategy often used in online examinations is the generation of personalised exams based on a pool of questions. This way, students receive an exam with different questions than their peers. While this reduces the risk of cheating, with an increasing number of questions up for selection, it also introduces the challenge of ensuring that each student receives an exam of comparable difficulty. We could investigate such a technique for Excel case studies as well, but improved approaches to personalisation (e.g., increased randomisation or shuffling) might serve the purpose equally well while ensuring that all students have comparable tasks to work on. From our standpoint, we should first work to reduce cheating possibilities through improved exam designs (described in Sect. 4.1), before focusing on the technical forms of exam supervision (which could compromise students' privacy). Based on the measures and ideas described, we are confident that attempts at cheating can be significantly reduced to a number similar to that of classical examination settings.

With regard to cheating detection, one approach includes increased analyses of the meta-level as well as the content-level. On the meta-level, we could compare pairs of students who use identical formulae (e.g., count the number of identical formulae between students). A high number of identical formulae may suggest collaboration, especially for incorrect solutions. Based on this distribution, we could extract the top-most 5% or 1% of pairs with the highest number of identical formulae. We could, additionally, incorporate the points achieved in the analysis because good students with everything correct tend to have identical formulae, too. Based on the outcome of the meta-analysis, we could manually check all the Excel submissions of those students who did not perform well on the exam and who also used many identical formulae. The threshold for these distributions has not yet been defined and would require additional investigation. However, it would also be possible to investigate further in the direction of artificial intelligence and alternative patterns, developing approaches and models in the future. When comparing the content in open text questions between different student submissions, we have so far relied on our judgement by looking for similar or identical answers. Going even further, we could, for example, automatically analyse the content of answers given to open text questions using Natural Language Processing (NLP). This may reduce the workload for the manual analysis, as unrelated answers can be filtered out beforehand. It is possible to inspect responses for the structure, language style, grammatical errors, or similar aspects of a free-text response. From our point of view, based on the measures described and possible improvement ideas, we are confident that successful cheating can be detected.

Critically reflecting on the automated correction, we have found that some errors go undetected and that there persists a tendency to award too many points for an incorrect answer. This may be a result of rounding; we normally round numbers depending on the number format (e.g., percent values to 4 digits) to avoid discrimination against students who solve the exercise using a calculator (even though this is not recommended). Another reason for undetected errors could be that the automated correction naively compares the values of the students with the sample solution and cannot critically reflect on the miscalculated outcome. This is because the automated correction mechanism first checks if the choices match the sample solution, and takes consecutive errors into account only if this is not the case. An example: A student should calculate the beta of a stock (value) and further select whether the calculated beta is (a) more sensitive to market movements or (b) less sensitive to market movements. Table 2 shows the evaluation/grades awarded of this case based on two possible outcomes.

Additionally, errors are systematised when the sample solutions used for grading are defined incorrectly, requiring control mechanisms to ensure the solutions are of the highest quality. One possible solution is to analyse the number of correct responses given by students, or the quality of the questions using item analysis. If very few students answer correctly, it may imply that the question was too difficult, or that the sample solution was incorrect. Questions with a low

Table 2 Automated correction of two consecutive questions and their evaluation (sample case)

No	Value of beta	More/less sensitive than market	Evaluation
Solution	0.8	Less sensitive than market	Both exercises are solved correctly, therefore full points are awarded
Student 1	1.2	More sensitive than market	The first exercise is not correct, therefore no points are awarded. There is a consequential error on the second question, therefore points are awarded
Student 2	1.2	Less sensitive than market	The first exercise is not correct, the second (based on the first exercise) is also not correct. No points should be awarded for both exercises. However, due to automated correction (matching the sample solution), the student gets points for the second exercise

Note. This table illustrates a scenario in which too many points are awarded for an incorrect answer. The first case shows the sample solution, while the second and third cases lay out possible solutions from students and the respective points awarded. In the third case, the student is given more points than they would have actually deserved score therefore need double-checking. We look not only for mistakes in the sample solution, but also for possible changes we could make in our teaching that would closer align the exam questions with the contents of the course.

While our online-assessment setup works well for the specific Finance courses in which it is used, the scope of the knowledge tested in our exams is concentrated on financial calculations, where the Excel format is very popular in the industry. More theory or factual knowledge-oriented domains are harder to fit into an Excel case study format, as these domains require more effort to, e.g., incorporate personalisation and other techniques that are simple to apply to numbers. While we incorporate open-ended question types for theoretical knowledge in our exams as well (e.g., explanations on the intuition of a computed number), these are relatively rare, as points are assigned manually. In future iterations of our grading tool, we plan to incorporate certain elements that aid in this process by, e.g., providing a predefined roster of point assignment options for a given set of results and/or mistakes made.

Another idea we have not yet implemented, but that is also based on the constructive alignment principle, is looking for typical mistakes that students made in an exam (e.g., figure out whether there are certain formulae that many students applied the wrong way). By looking for such frequently occurring typical mistakes, we would critically assess and improve the learning materials and teaching activities to better address typical student-errors. The examination can serve as a feedback loop when one is working on the contents for the next iteration of the course.

Additional improvements to the learning experience of our students would be to give each of them comprehensive feedback on their exams. We are thinking about generating an individual report for each student, as well as adding feedback to their Excel exams with green (correct answer), orange (incorrect value, but correct due to consecutive errors), and red (incorrect answer). Additionally, students who committed typical errors could receive automated comments in their Excel exams with a pre-defined explanation of their mistakes. If provided to students, such materialised feedback could serve as an additional learning outcome of the course, even if the course itself has ended.

A last point to critically reflect on is the collection of data from non-systematic surveys. To obtain statistically significant survey results, we would need to conduct more systematic research. We may not have recorded certain aspects of the problem. Based on future systematic surveys, we would like to determine which aspects of the e-assessment exams perform better or worse than traditional pen and paper exams. Our goal is for students to not have any disadvantages compared to a paper-based exam, and for them to prefer online-assessment exams to traditional exams due to good preparation through the learning material provided.

7 Conclusion

We strongly believe that online-assessments add value for both lecturers and students. The online-assessment practice we have developed follows the trend towards higher level examination according to Bloom's taxonomy. Exams were redesigned in terms of content, became more application-oriented, and allowed for improved testing of cross-linked knowledge. Easily adapted, the digital exam format is more resistant to crises and can quickly be upscaled to fit various subject areas in Banking and Finance. The alignment between learning activities, exam design and the tasks that students might perform in work life, ensures that students know that they are learning skills they can practically apply. Automated correction alongside personalised exams reduces human subjective judgement, improves fairness for students by limiting cheating opportunities, and allows lecturers to grade in a more structured and efficient way.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *Taxonomy for learning, teaching and assessing. A revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman.

Biggs, J. (2003). *Aligning teaching for constructing learning. Higher Education Academy, 1*(4). https://www.cardiff.ac.uk/__data/assets/pdf_file/0020/584030/Aligning-teaching-for-constructing-learning-John-Biggs-HEA.pdf.

Bloom, B. S. (Ed.). (1973). *Taxonomie von Lernzielen im kognitiven Bereich* (4th ed.). Beltz.

Emeka, C., & Zilles, C. (2020). *Student perceptions of fairness and security in a versioned programming exam*. In *ICER'20: Proceedings of the 2020 ACM Conference on International Computing Education Research* (pp. 25–35). https://dl.acm.org/doi/10.1145/3372782.3406275.

Gordon, M. E., & Fay, C. H. (2010). The effects of grading and teaching practices on students' perceptions of grading fairness. *College Teaching, 58*(3), 93–98.

Kaiser, F.-J. (1983). Grundlagen der Fallstudiendidaktik – Historische Entwicklung – Theoretische Grundlagen – Unterrichtliche Praxis. In F.-J. Kaiser (Ed.), *Die Fallstudie – Theorie und Praxis der Fallstudiendidaktik* (Vol. 6, pp. 9–34). Verlag Julis Klinkhardt.

Liu, J., & Zhang, B. (2020). *Multi-level Rasch model analysis of computer-assisted automated scoring of English listening and speaking tests*. In *2020 International Conference on*

Computer Engineering and Application (ICCEA) (pp. 632–636). https://ieeexplore.ieee.org/document/9103771.

Rusak, G., & Yan, L. (2021). *Unique exams: Designing assessments for integrity and fairness.* In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education.* https://arxiv.org/pdf/2009.01713.pdf.

Schaper, N., Reis, O., Wildt, J., Horvath, E., & Bender E. (2012). *Fachgutachten zur Kompetenzorientierung in Studium und Lehre* (pp. 1–148). https://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-02-Publikationen/fachgutachten_kompetenzorientierung.pdf.

Johanna Braun is leading the Teaching Center at the Department of Banking and Finance of the University of Zurich that focuses on online learning, online assessment and innovation in teaching. Before joining the University of Zurich, she worked in Strategy Consulting at Deloitte Consulting.

Roland Schläfli is a team lead IT and software engineer at the Department of Banking and Finance at the University of Zurich. At the department's Teaching Center, he works on innovation in teaching (e.g., classroom interaction with the KlickerUZH, gamifiction, automated grading). He also focuses on various serious game projects and seeks to foster the skills of lecturers in game-based learning.

David Schmocker is team lead Community in Educational Development at the University of Zurich. He manages various communities in the educational context, focusing on topics such as digitalization, interdisciplinarity, and internationalization in higher education. He also works as a high school teacher teaching psychology and pedagogy.

Benjamin Wilding is managing director at the department of banking and finance of University of Zurich. In his role as Lecturer, he focuses, among other things, on the creation of meaningful and practical forms of examination.

E-Examinations@Home: Adapting Large-Scale Digital Assessments for Remote Environments

Nils Hernes

Abstract

In early 2020, most German universities were faced with a major challenge: The coronavirus pandemic made it necessary for digital remote assessments to be introduced on very short notice, along with the required workflows and infrastructure. In responding to this difficult task, Freie Universität Berlin was able to draw on its longstanding experience conducting large-scale digital on-campus assessments. The present case study analyses Freie Universität Berlin's digital remote assessment strategy with regard to assessment literacy and technology, logistics and organisation, as well as legal reliability. As will become clear, assessing higher-order thinking skills, protecting assessment integrity by deploying a lockdown browser, providing extensive support structures for lecturers and students, and switching assessments to an open-book format were important cornerstones in its successful implementation. In the event, the new, highly scalable digital assessment format proved to be much more than a mere stopgap: Once the need for students to be physically present is removed, universities can effectively assess large student cohorts without having to invest in expensive infrastructure, while also enhancing their attractiveness to international students by offering digital, on-demand courses.

Keywords

Assessment literacy • Assessment logistics • Assessment technology • Digital assessments • Remote assessments

N. Hernes (✉)
Center for Digital Systems (CeDiS), Free University Berlin, Berlin, Germany

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_3

1 Introduction

Under significant pressure from the coronavirus pandemic, German universities saw an unprecedented move to digitised learning, teaching, and assessment as they were forced to shift from on-campus to remote environments in early 2020.

Whereas digital learning environments were already in widespread use prior to the pandemic, German universities introduced digital assessments at a much slower pace, although benefits such as automated scoring and scalability had long been known and successful implementations had been documented (Schulz & Apostolopoulos, 2011). During the pandemic, digital assessments suddenly became a promising, but mostly unfamiliar option for safe examinations.

In June 2020, Freie Universität Berlin (FUB) introduced *E-Examinations@Home*, a large-scale concept for digital remote assessments. FUB had been exploring the field of digital assessments since 2005, with more than 120,000 digital on-campus assessments conducted. The switch to a digital remote setting demanded extensive adaptations to FUB's existing on-campus strategy, affecting didactic approaches, technical infrastructures, as well as organisational and logistical workflows. With over 35,000 assessments completed since June 2020, E-Examinations@Home proved to be a robust and scalable solution.

This case study discusses the general implementation of E-Examinations@Home and analyses the concept from several perspectives in order to highlight aspects that were vital for the successful transformation of FUB's digital on-campus assessment practice to meet the requirements of a large-scale remote assessment scenario.

2 E-Examinations@Home: A Concept in Practice

During the development of its remote assessment concept, the E-Examinations unit at FUB could draw on ample experience in the field of digital assessments. Starting in 2005, the unit has gradually established digital on-campus assessments as an alternative to conventional paper–pencil exams (Schulz, 2017). With two dedicated digital assessment centres for on-campus assessments, over 300 computer-equipped workstations, and a wide range of didactic consultation offerings, the E-Examinations unit puts at the lecturers' disposal a scalable digital assessment environment which allows them to benefit from automated scoring, reduced logistical efforts, and automated statistical analysis, to name but a few examples.

As the pandemic brought in-person gatherings to a halt, it soon became clear that significant adaptations would be required to enable digital remote assessments to take the place of most conventional forms of on-campus assessment. Differing markedly from established assessment practices, the new examination format had wide-ranging implications for lecturers and students alike, not to mention the difficulties associated with its technological implementation. Clearly, with over 33,000 students enrolled at FUB, a remote assessment concept needed to be well structured, highly reliable, and accessible.

E-Examinations@Home is designed as a centralised service. The E-Examinations unit at FUB coordinates the individual stages of the assessment process, including the allocation of exam slots, maintenance of the assessment system, didactic consultation, technical support, as well as the vetting and conducting of exams. Throughout this process, lecturers and students follow standardised workflows set up by the E-Examinations unit.

2.1 Assessment Preparation

Lecturers must book their exam slot at the beginning of the new semester. Once the booking is confirmed, they are required to submit their exam and the list of examinees in advance using digital templates provided by the E-Examinations unit. Lecturers are offered didactic consultation to help them adapt exams to the digital remote setting and to make effective use of system features. After submission, the E-Examinations unit implements the exams into the assessment system and discusses technical or didactic adjustments with lecturers.

The E-Examinations unit provides students with information concerning their exam accounts multiple weeks in advance, along with instructions for technical preparation. Students are given the opportunity to complete a mock exam to familiarise themselves with the assessment system and its user interface. In order to do so, students must install a lockdown browser that they will also be required to use in the actual exam. Students can contact the E-Examinations unit via e-mail or telephone to sort out technical difficulties at any time before their exam.

2.2 Conducting the Assessment

At the beginning of each assessment, students, the lecturer, and technical staff from the E-Examinations unit come together in a video conference. The lecturer gives explanations concerning the exam and spot-checks students' IDs, while a member of the E-Examinations unit walks students through the assessment process and provides instructions on what to do should technical problems occur. Then, due to legislation regarding data protection and information privacy, the video conference is terminated, and students conduct the assessment without digital supervision. Students can, however, contact technical support from within the lockdown browser. Should students drop out of the exam due to technical issues, they can re-enter the assessment within a certain interval. Interruptions and resumptions are recorded in the computer-generated exam protocol.

2.3 After the Assessment

After the student group has completed the assessment, lecturers can grade the exams via the assessment system. If the assessment consists of selected response tasks only, lecturers have recourse to a computer-generated score. Lecturers can also use computer-calculated statistics to retroactively evaluate their exam with regard to properties such as selectivity and difficulty, while students can review their graded exam via the lockdown browser. In a final step, the exam documents are exported and archived on archive-grade DVDs by the E-Examinations unit, and handed over to lecturers and the responsible student records and examinations office for safekeeping and storage.

3 Transforming Digital Assessments for Remote Settings – Key Areas

The fundamental logic of digital on-campus and remote assessments is the same:

- students complete the assessment via computer accessing a web-based assessment system,
- the results are stored and processed digitally,
- and lecturers profit from features such as automated scoring.

Yet, there is also a major difference in that the assessment process involves factors that go beyond the assessment's medium—switching from on-campus to remote assessments demonstrated the significant influence of the new environment and its modalities on all components of the assessment process. The large-scale setting at FUB further amplified these effects, **as logistical bottlenecks affected**

many users. To identify transformational aspects and describe the specific characteristics of digital remote assessments, the following chapter introduces four categories for the analysis of digital assessments in general. These categories are then applied to E-Examinations@Home, highlighting several essential differences between remote and on-campus assessments.

3.1 Analytical Categories

The following categories are widely used in studies of digital assessments and cover four key areas of analysis (Friedrich et al., 2015; Schmees & Horn, 2021). Although they are by no means exhaustive (ethics and assessment psychology are only two important aspects that are excluded from this analytical model and thus from many studies), this article maintains the established fourfold division for reasons of comparability.

3.1.1 (Digital) Assessment Literacy

The term *Digital Assessment Literacy* describes the didactic competencies required to conduct digital assessments proficiently. Assessment literacy is generally understood as an educator's competency with regard to assessment principles, methods, evaluation, and feedback strategies (Sadler, 1998; Stiggins, 1995). Amongst other things, a highly developed assessment literacy permits educators to identify relevant competencies and to choose suitable methods and tasks for eliciting these competencies. Digital assessments provide new technological possibilities to conduct assessments, measure competencies, and evaluate exam items. Educators require digital-specific didactic competencies to make use of these new possibilities, which is why Eyal (2012) has proposed the term *Digital Assessment Literacy* to highlight the transformational impact of the digital medium.

3.1.2 Assessment Technologies

This category encompasses the technological components (hardware and software) that are needed to implement a given digital assessment strategy. Depending on scalability and the specific didactic use case, technological requirements can differ significantly.

3.1.3 Assessment Logistics and Organisation

This category is dedicated to the structures and workflows that allow a given digital assessment strategy to be put into practice. The subcategory of assessment organisation deals with the format's strategic orientation as a centralised or decentralised service and its structural implications, as well as the respective roles and activities of its stakeholders, while the subcategory of logistics is concerned with practical workflows and required resources.

3.1.4 Legal Reliability

This category focuses on the legal implications of deploying digital assessments. Due to their relative novelty within the German university landscape, digital assessment strategies must establish legal frameworks that comply with existing legislation (for example concerning data storage and processing) and university bylaws. The equal treatment of examinees and technical malfunctions (of both hardware and software) are further key areas of legal concern.

3.2 Assessment Literacy

Before the coronavirus pandemic, on-campus assessments at FUB were predominantly used to assess large introductory courses, a scenario which allowed lecturers to take full advantage of technological features such as automated scoring offered by the FUB's large-scale digital assessment infrastructure. Thus, lecturers mainly focused on assessing lower-order thinking skills such as reproduction and recognition by means of selected response tasks. The on-campus setting was ideally suited to the assessment of those competencies, as the risk of cheating and collusion amongst students was low due to continuous supervision during the exam.

The switch to a digital remote setting substantially altered the modalities of assessment. Not only is it difficult to supervise students during digital remote assessments, as educators can only ever monitor a fraction of students' workspaces and can hardly prevent them from communicating—in Germany, digital supervision is also subject to robust legal constraints and was difficult to deploy within most German federal states in 2020 (see Sect. 3.5). The constraints regarding supervision have had far-reaching implications for digital remote assessments: If neither the use of resources such as notes and the internet, nor collusion amongst students can be forestalled, the validity of exams is bound to decrease. In their meta-analysis, Steger et al. (2020) have demonstrated that the searchability of tasks significantly affects assessment scores. Reproduction and recognition tasks, in particular, are easily searchable, as they do not require high cognitive capacities (Bloemers et al., 2016; Diedenhofen & Musch, 2017), which means that they cannot be relied upon to validly assess competencies in an unproctored remote setting, especially when using selected response tasks.

To increase the validity of digital remote assessments, the E-Examinations unit at FUB encouraged lecturers to assign tasks that elicit higher-order thinking skills such as transfer, critical thinking, and problem solving (Brookhart, 2010), and shared with them Anderson's and Krathwohl's learning taxonomy (Anderson, 2014).

While Anderson's and Krathwohl's taxonomy has served as the theoretical foundation for teaching and learning in German higher education since the competency-oriented turn of the 2001 Bologna reform (Hilkenmeier & Schaper, 2013; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2017), the extent of its practical application in study courses and assessments was still fairly limited.

The taxonomy splits competency into cognitive processes (*remember, understand, apply, analyse, evaluate, create*) and knowledge dimensions (*factual, conceptual, procedural, meta-cognitive*). This permits educators to pinpoint which type of knowledge they want to assess and how it should be demonstrated, which in turn facilitates the practical application of abstract learning goals. Recognition and reproduction are both part of the lowest-ranking cognitive process: *remember* (Anderson, 2014). To increase validity, the E-Examinations unit advised lecturers to focus on the other five categories when constructing assessment tasks. Starting from *understand*, these categories require students to employ the knowledge and competencies they possess to solve new problems or to process new information (Anderson, 2014).

When presented with the taxonomy, lecturers were prone to two misconceptions. First, lecturers tended to treat the hierarchy of cognitive dimensions as a scale of increasing difficulty. Higher cognitive dimensions are indeed more complex than lower ones, but they are not necessarily more difficult, as the difficulty of a task is not only determined by the task's cognitive complexity, but also by the knowledge dimension addressed and by the students' familiarity with the task. For example, paraphrasing a conversation in a newly learned language can be much more difficult for students than solving a complex mathematical equation provided that students have already solved similar equations during lessons (Anderson, 2014)—yet, paraphrasing is associated with the (lower-ranking) category *understand*, whereas using mathematical formulas falls in the (higher-ranking) category *apply*. The lacking correlation between the ranking of cognitive dimensions and their difficulty has also been empirically demonstrated by Seddon (1978), amongst others.

The second misconception revolves around the suitability of selected response tasks to assess higher-order thinking skills: Their extensive use to assess competencies such as recognition and reproduction does not, in fact, rule out their application in such a context. Rodriguez (2016), Ingenkamp and Lissmann (2008), and Terhart et al. (2009) have shown that the format can be employed to assess all cognitive dimensions except *create*, and multiple empirical studies (Devitt & Palmer, 2007; Hofmeister, 2005; Tiemeier et al., 2011) have demonstrated that selected response tasks are not inferior to constructed response tasks when it comes to assessing the aforementioned categories. In order to assess higher-order thinking skills, Mayer et al. (2009) propose a task design that embeds information in new contexts or uses practical examples to which a given theory can be applied.

In practice, the change in focus from lower-order to higher-order thinking skills meant that lecturers had to commit more time to the assessment. Not only were they required to design complex tasks that were conducive to the assessment of cognitively demanding competencies, but it also took them longer to grade the exams if they opted for constructed response tasks such as essays, which must be evaluated manually.

Apart from the use of unauthorised resources, another major issue had to be addressed: the high risk of collusion amongst students, which is equally difficult to prevent in a setting with limited proctoring or even no proctoring at all. The ability to use the randomisation functions offered by the assessment system to reduce collusion therefore constituted a second important aspect of lecturers' assessment literacy.

Most assessment software enables the basic randomisation of tasks and answers to impede collusion—however, there is only a finite number of tasks, which allows students to quickly defeat the randomisation measures. In response to this problem, we encouraged lecturers to introduce higher levels of randomisation in their assessments, which can be achieved by randomising not only the order of tasks but the tasks themselves. The digital setting allows for a much more granular approach than the time-honoured strategy of distributing two or more different exam versions. Whereas students can quickly figure out their version of the exam if they can communicate with their peers, digital assessment software enables lecturers to create individual assessments by using a sizable item pool from which the system picks items at random to compile an assessment.

Yet while a bigger item pool yields a greater degree of randomisation and individualisation, lecturers must ensure that assessments are comparable with regard to thematic focus, anticipated competencies, and item difficulty. If compiled randomly, there is a strong likelihood that the assessment will violate the goal of objectively measuring students' performance. Thus, we counselled lecturers to categorise tasks by difficulty, topics, and competencies, which reduces randomness, but still allows for a didactically sound assessment that effectively impedes student collusion.

Of course, a large item pool is more time-consuming for lecturers to create. To mitigate this issue, lecturers were advised to produce different versions of the same task by cloning items, an established method of constructing test items especially in digital assessment settings (Bejar et al., 2003). One possible method of item cloning called *replacement set procedure* consists of replacing certain variables in the item stem while maintaining the overall logics and structure of the task (Glas & van der Linden, 2003)—this template-like approach to item construction can be applied to mathematical equations, for example. Cloning items gave lecturers at FUB the ability to quickly create multiple versions of the same task, which efficiently enlarged the item pool while avoiding the difficult and error-prone process of constructing entirely new tasks and distractors (Haladyna et al., 2016).

In summary, the implementation of digital remote assessment formats for use within an unproctored environment such as E-Examinations@Home required lecturers at FUB to further develop their assessment literacy. As the use of prohibited resources and collusion amongst students could not be ruled out, strategies had to be developed to address these issues. Deploying tasks that measure higher-order thinking skills, incorporating high levels of randomisation, and systematically creating large item pools were three key skills that allowed lecturers to carry out valid assessments despite the challenges associated with a remote setting.

3.3 Assessment Technologies

Digital assessments demand high technological standards, as they involve high-stake data that fulfils a gate-keeping function regarding students' futures. Technical errors resulting in malfunctioning software or data loss can have serious legal implications. Furthermore, students are oftentimes unfamiliar with and thus insecure about digital assessments—any technical difficulties can endanger students' trust in the procedure. Unintuitive software or user interfaces as well as technical problems can put students under additional stress (Elsalem et al., 2020), which correlates with reduced performance (Brodersen & Lorenz, 2020; Vedhara et al., 2000).

To conduct safe and user-friendly digital assessments, FUB relies on the assessment system “LPLUS TestStudio”. Employing a dedicated assessment system yields multiple benefits such as:

- a user interface (UI) and experience (UX) exclusively designed for conducting assessments,
- a special focus on data integrity,
- redundant data backups,
- and fast, simultaneous read and write access.

Using the built-in assessment functions of Learning Management Systems (LMS) can be problematic, as it leads to a mixing of course and assessment data. In the past, LMS were often optimised for a high number of simultaneous read accesses to course data and course files, while write accesses were not treated as a priority. Within a large-scale assessment scenario, this can lead to delays in data processing and even to data loss. When using an LMS for digital assessments, the assessment system should be operated separately from the course section, and should be optimised for a high number of simultaneous write accesses (Schulz, 2017).

Processing speeds and data integrity also depend on server computing power. Whereas in-house servers were sufficient to conduct digital on-campus assessments at FUB, the steep rise in digital assessments due to the pandemic required a switch to cloud infrastructure hosted by LPLUS in a professional data centre, a step which enables fast, continuous, and simultaneous access even when a large number of assessments is carried out simultaneously. Using servers hosted by LPLUS also gives quick and easy access to professional support and bug fixing, which would have been difficult to provide when running servers locally.

Moreover, the remote setting called for specialised software on the client side. As students utilised their private computers, their ability to share answers or copy content from the internet had to be restricted. Thus, FUB introduced *Safe Exam Browser* (SEB), an open-source lockdown browser developed for Windows and macOS by the Swiss Federal Institute of Technology in Zurich (Schneider et al., 2012). To guarantee the use of SEB, the assessment system checks the browser’s https header and matches it with a unique key deposited within the system; SEB generates the corresponding key based on the configuration provided by the E-Examinations unit.[1] When running SEB, students can only access FUB's assessment system, whereas other applications cannot be launched. Naturally, these limitations only apply to the computer used for completing the exam. While disabling copy-paste reduces the ease of exchanging answers, dishonest students can still resort to other devices to communicate or look up solutions—what becomes obvious here is that deploying a lockdown browser is only one milestone on the path to enhanced academic honesty.

Apart from locking down students' computers, SEB also standardised access to the assessment system. In comparison to digital on-campus assessments at FUB, where assessment centres are equipped with computers running the same configuration, digital remote assessments are conducted on devices that are highly heterogenous in terms of the hardware and software installed. Thus, the assessment experience can differ widely depending on the browser, its plug-ins, and the use of additional software. The introduction of SEB helped to create a similar assessment experience for all examinees even if a broad variety of endpoint devices was employed: SEB provides the same UI and UX across the many different devices used by students, and settings can be configured and distributed centrally. Reducing technological variety was also important to enable high scalability: By establishing a single point of access, the E-Examinations unit narrowed down possible sources of error, reduced the volume of support requests, and improved its capability to render assistance in case of technical difficulties. At the moment, SEB's lack of support for Linux means that Linux users must employ a non-Linux device to take the exam—when in-person gatherings resume, the E-Examinations unit will explore the possibility of giving students who do not have access to the required hardware access to computers in the assessment centres.

In brief, three technological aspects were of key relevance in the implementation of large-scale digital remote assessments: On the server side, the use of a dedicated assessment system running on scalable cloud infrastructure absorbed the computational strain of rising numbers in assessments, and secured data integrity even under high load; on the client side, deploying SEB provided a technical solution that increased assessment validity, while also canalising access and achieving the required standardisation of students' private hardware and software.

[1] Not every assessment software includes a matching tool for SEB's configuration key. Two open source plug-ins for Moodle (https://github.com/lucaboesch/moodle-quizaccess_safeexambrowser) and ELIAS (https://github.com/eqsoft/SEBPlugin) can be found via the respective hyperlinks. (Last accessed 15.07.2021).

3.4 Logistics and Organisation

Prior to the pandemic, FUB conducted about 20,000 digital on-campus assessments per year. With the onset of the pandemic, lecturers turned to digital remote assessments, and the number of digital assessments rose to 35,000 per year (as of August 2021). Thus, the introduction of E-Examinations@Home not only demanded the development of a new assessment format in general, but also the adaptation and expansion of the existing logistical infrastructure in order to support a large-scale remote environment, as well as the establishment of an organisation that was capable of effectively assisting lecturers and students during the process of transition to the new format.

The most striking logistical change from digital on-campus to digital remote assessments was caused by the shift from physical to virtual presence. While the number of students that can be assessed simultaneously on campus is determined by the physical size of the examination space and the available equipment, digital remote assessments are only limited by the capacity of the assessment system for concurrent access. Whereas before the pandemic, the E-Examinations unit had to conduct multiple subsequent rounds of assessment when assessing large groups, E-Examinations@Home can accommodate groups of up to 900 students in one sitting. In addition, the shift from one physical assessment space to students' homes also permits the simultaneous assessment of multiple student groups, thereby allowing conventional serialisation to be replaced with the parallelisation of assessments. At FUB, parallelisation is currently limited to three simultaneous assessments, as technical staff is required for each assessment's preliminary video conference and technical support during the assessment—the E-Examinations unit is in the process of exploring the use of videotaped introductions to reduce this bottleneck.

At FUB, the benefits of parallelisation were at least partly offset by the need for additional support structures: When digital remote assessments were introduced on an emergency footing, many lecturers had no previous experience with the format, leading to a lack of didactic and organisational know-how. In response to this problem, the E-Examinations unit centralised the organisation and logistics of assessments in order to relieve pressure on the lecturers, while ensuring a technologically, organisationally, and legally sound transition to remote exams.

As a first port of call, the E-Examinations unit launched a central support website featuring extensive information about the assessment process, which allowed lecturers to familiarise themselves with the new format and lowered the perceived complexity of conducting digital remote assessments. In addition, the E-Examinations unit also held virtual consultation hours, an offer that was very well received, as many lecturers were initially unsure how to approach digital remote assessments, both from a didactic and organisational angle.

Assessment preparation was standardised, with a focus on concrete steps to be taken by the lecturers and a clear delineation of their responsibilities. Lecturers had to submit their exams eight weeks prior to the scheduled date of the assessment, while the list of examinees had to be handed in four weeks in advance using templates provided via the central web hub. Requiring lecturers to design their exams two months in advance was an unusual but unavoidable measure: To save lecturers the effort of having to learn how to enter their exams into the assessment system, it was decided that the E-Examinations unit would take over the task of transposing the exam templates into digital assessments, which created a bottleneck in that enough preparation time was required to process the large number of assessments. This was accepted as a necessary evil, as the lecturers' unfamiliarity with the software could otherwise have led to a rejection of the format, not to mention a high number of support requests and serious technical flaws in the assessment. With the intense pressure of the initial coronavirus outbreak now subsiding, efforts to allow technically well-versed lecturers access to the assessment software are under way, along with the design of introductory courses designed to improve user engagement with E-Examinations@Home. Apart from giving the E-Examinations unit time to transpose the assessments, the two-month window also provided lecturers with ample opportunity to review and adapt their digital exams, which was a necessity given that many lecturers had never before assessed digitally. By and large, lecturers reported high levels of satisfaction regarding the support structures, and the collaborative, iterative approach to adapting and vetting the assessment meant that errors in the implementation could be kept to a minimum.

The new format's main stakeholders were the many students who had to be introduced to digital remote assessments. Thus, supporting students in their preparation for E-Examinations@Home was one of the key factors in the format's success. Before the pandemic, digital on-campus assessments were designed as a one-stop service: The digitised assessment was vetted extensively by lecturers and the E-Examinations unit to ensure full functionality and correct implementation; the hardware and software was checked regularly; and on the day of the exam, students were assigned their own workstation, with special stations prepared for students with disabilities. All students had to do was to arrive on time and complete their assessment.

Digital remote assessments, meanwhile, require a much higher level of preparation and involvement. The remote setting means that students must use their own computers (BYOD). As the E-Examinations unit could not manage hardware and software centrally anymore, students had to:

- ensure that the assessment software ran well on their computers,
- familiarise themselves with the assessment system,
- and be able to solve problems in case of device failure or malfunctioning software during the assessment.

Setting up hardware and software and dealing with technical problems can be particularly taxing for students. As malfunctions endanger successful participation in an assessment, students who are already stressed due to the assessment situation itself may experience additional pressure when they encounter frustrating technical difficulties, and there is a high risk of them losing trust in the assessment procedure altogether. In order to provide students with guidance throughout all steps of technical preparation, the E-Examinations unit introduced services that allow for extensive onboarding and support measures during the run-up to an exam.

As had been done for lecturers, the E-Examinations unit set up a central website that provides students with comprehensive guidelines for assessment preparation. The central website not only contains step-by-step instructions, but also installation files for SEB, video tutorials, FAQs, and troubleshooting guides. Despite all this, it was expected that some students would experience technical complications that they could not fix themselves: A BYOD environment of the sort envisaged here inevitably includes many different devices with different operating systems, individual software settings, and, in some cases, weak hardware. In order to spot and solve technical problems proactively, the E-Examinations unit provided students with exam accounts early on, which in turn enabled them to complete a mock exam and test their setup and software in the process. The mock exam also allowed students to familiarise themselves with the system, reducing insecurity during the real exam. Requiring the list of examinees to be submitted well in advance of the assessment came at the cost of some students registering after the initial submission, which necessitated iterative, time-consuming account updates. Thus, automating the registration process by connecting the assessment database to students' university accounts is one of the key goals for further development of E-Examinations@Home. To solve technical problems, the E-Examinations unit set up email and telephone support as well as consultation hours. Shortly before the assessment, students came together with the lecturer and staff from the E-Examinations unit to be briefed about the exam process and how to solve common technical problems during the assessment. While completing the exam, students could post questions in a support forum. With these measures, the E-Examinations unit sought to anticipate students' needs and provide optimal assessment conditions.

With 1000 pageviews of the central website per day on average and more than 1500 emails received in 2020, the data exemplified students' need for guidance and support. The required services—especially email and phone support—proved to be very resource intensive, and the troubleshooting of technical edge cases made it necessary for the E-Examinations unit to work on demand whenever problems arose. Nevertheless, building up these extensive support structures paid off, as on average only about 1–2% of students dropped out of assessments due to technical problems.

In summary, while the advantages of parallelising assessments were partially cancelled out by increased preparation times and the need for extensive support structures, a centralised top-down organisation enabled FUB to manage the rising number of assessments and the concomitant logistical challenges. To ensure this success, it was necessary to:

- streamline the assessment process and reduce stakeholders' need for technical competencies,
- define clear, step-by-step assignments for lecturers and students during preparation,
- and provide extensive support structures to ensure correct execution and prevent confusion.

3.5 Legal Reliability

Prior to the pandemic, not all German universities offered digital assessments, and thus did not have the necessary legal structures in place when they were forced to switch to digital assessments. Due to their specific characteristics, assessments conducted and processed via digital assessment software constitute a unique assessment format (Jeremias, 2018). As such, they require regulation in universities' bylaws (Jeremias, 2018) in order to be available as an option to lecturers and students. However, not every assessment carried out via digital means constitutes a digital assessment per se—for example, in many cases, oral assessments conducted via video conferencing tools do not significantly alter the assessment's characteristics and, therefore, operate within the borders of the format's definition in most university bylaws (Dieterich & Fischer, 2021). Dieterich and Fischer (2021) also describe an alternative: Universities can bypass the legal regulation of digital assessments by seeking students' consent for each specific digital assessment, but within a large-scale environment, the logistical and organisational effort required to provide on-campus alternatives to students refusing to take digital assessments is extremely high (in addition, some pandemic-related state regulations prohibit most forms of on-campus assessment—for Berlin, see: Senatskanzlei - Wissenschaft und Forschung, 2020). At Freie Universität Berlin (2013), digital assessments were included in the university's bylaws in 2013 after an initial explorative phase.

Apart from the issue of their inscription into university bylaws, digital remote assessments entailed further legal challenges for FUB. Whereas most digital on-campus assessments were conducted as supervised exams, supervision within a remote setting became a highly controversial topic throughout the landscape in 2020. Due to Germany's and the EU's high standards regarding personal data protection, proctoring was off the cards in most German states and, therefore, not possible at FUB.[2] As with digital assessments in general, universities were allowed to supervise students digitally with their consent, but they would run into the same logistical and organisational problems if students did not agree to be supervised and had to be offered on-campus alternatives (Hoeren et al., 2020). Thus, digital remote assessments at FUB were legally classified as open-book exams, which do not necessarily require supervision (Albrecht et al., 2021). Consequently, departments had to change their study regulations to permit open-book exams instead of supervised exams for most of their courses, a step that was necessary to enable a swift and simultaneous transition. To maintain at least some form of verification that students were taking the exams themselves, students had to make a declaration to that effect, and a random sample of examinees were asked to present their ID during the briefing. Still, unproctored digital remote assessments offer little chance of verifying who takes the exam and resemble term papers as far as the safeguarding of academic integrity is concerned.

Even though changing the legal status of digital remote assessments to that of an open-book exam made it possible to carry out assessments without the need for supervision, students' domestic environment posed legal uncertainties, too. Digital on-campus assessments at FUB had so far been conducted in two assessment centres, which provided students with optimal conditions such as an air-conditioned and soundproofed room, standardised, high-performance workstations, and stations accessible to students with disabilities. This setting complied

[2] The following laws inhibit the use of proctoring: on the national level, the Right to Informational Self-Determination (Bundesverfassungsgericht (1983)) as well as the inviolability of the home (Article 13, Basic Law); on the European level: Article 6 GDPR.

with the principle of equal treatment, which demands equal assessment conditions for all students. Within a remote environment, on the other hand, these conditions are determined by each individual's domestic circumstances. Whether students live next to a busy street, share their room with siblings, or have an unstable internet connection can all have a detrimental effect on assessment conditions. FUB has therefore decided to leave the decision to participate in digital remote assessments to the students—those who did not wish to take remote exams were able to postpone them without being penalised. Moreover, the E-Examinations unit is exploring possibilities to provide students who are confronted with unfavourable assessment conditions at home with workstations in the assessment centres once restrictions on in-person gatherings ease. Simultaneously conducting assessments on campus and remotely deprives the format of some of its advantages and presents logistical hurdles, but ensuring equal treatment and mitigating unsuitable assessment conditions are key demands that must be addressed in order to establish digital remote assessments in the long term.

Apart from problems related to the students' domestic situation, technical issues can arise throughout the assessment—batteries can die, the internet connection can drop randomly, or the computer can stop working altogether. As technical failures cannot be completely prevented by students, they should not experience disadvantages such as receiving a bad grade or even failing the assessment if malfunctions occur, which is why FUB classified any assessment terminated due to technical breakdowns under pandemic circumstances as not undertaken. Given that students commonly have only three attempts to pass an exam, this rule underscores FUB's commitment to provide fair and equal conditions to all of its students and to support them in this unprecedented crisis. It must be noted, however, that this temporary arrangement gives something of a carte blanche to students: As it is nearly impossible to retrace the exact reasons for technical failures, students can drop out of an assessment if they fear a bad grade by unplugging their ethernet cable or by disconnecting their computer from the power source. How to effectively differentiate between real or simulated technical difficulties is a question that remains to be answered once pandemic-related rules expire.

When it was faced with the coronavirus pandemic, FUB focussed on finding flexible yet dependable legal solutions that enabled the quick adoption of digital remote assessments in all departments, while simultaneously reducing potential negative effects on students. Current regulations will likely require adaptation and specification once students at FUB return to the campus and digital remote assessments lose their emergency status. Then, at the latest, it will become necessary to establish a legal framework that conclusively addresses equality issues and technical complications during assessment.

4 Making a Virtue of Necessity? Remaining Challenges and Unresolved Issues

As most conventional on-campus assessments came to a halt last year, digital remote formats rose to the occasion. Yet, the transformation of digital on-campus assessments into digital remote assessments demonstrated that even though the medium itself remained unchanged, its surrounding parameters changed significantly. Digital remote assessments require specialised didactics as well as an organisation and logistics focused on strong support structures for lecturers and students, and their legal status differs fundamentally from other assessment formats. Thus, digital remote assessments entail more than a mere shift of location—they constitute a distinct assessment format with its own strengths, weaknesses, and idiosyncrasies. The present case study has shown that lecturers, students, and administrative staff experienced a steep learning curve as they accustomed themselves to the new format. On the other hand, the university-wide move to digital assessments drew attention to their many advantages in comparison to conventional assessments (streamlined logistics, automated scoring, etc.), so that many lecturers plan to 'stay digital' even after the pandemic has subsided. In times of rising student numbers, digital remote assessments also offer an unparalleled degree of scalability, which in turn promises to relieve the logistical pressure on university administration.

Nevertheless, digital remote assessments still experience growing pains. Their emergency deployment demanded quick and unbureaucratic implementations that required a high level of flexibility from lecturers, students, and administrators. Setting our sights at the future of digital remote assessments beyond the pandemic, it is obvious that the new format still needs to mature, especially in the areas of assessment literacy and legal reliability. If digital remote assessments are to endure, higher education didactics must continue to explore the format's strengths and weaknesses in comparison with other assessment formats. Moreover, the remaining legal uncertainties will have to be cleared up—handling technical failures and their resulting legal consequences, providing equal assessment conditions to all students, and integrating effective digital supervision that adheres to data protection laws are three particularly pressing issues that are in urgent need of being addressed. This urgency is only bound to increase once emergency regulations cease to be in effect.

Digital remote assessments have the potential to be much more than a mere stopgap measure. If pursued thoroughly, the new format can yield solutions to a broad variety of problems, especially in large-scale environments such as the one found at FUB. Digital remote assessments also enable universities to explore new strategies such as offering international classes with location- and time-independent assessments, and can function as essential facilitators of innovation regarding formative assessments in an increasingly digital and transnational university landscape.

This paper was funded by the Federal Ministry of Education and Research and the State of Berlin through the Excellence Strategy of the German Government *Berlin University Alliance.*

References

Albrecht, J., Mc Grath, O., & Uphues, S. (2021). Aufsichtsklausuren aus dem Homeoffice: Datenschutzrechtliche Zulässigkeit der Online-Aufsicht von Studierenden. *Zeitschrift für Datenschutz (2),* 80–84.

Anderson, L. (2014). *A taxonomy for learning, teaching, and assessing. A revision of bloom's taxonomy of educational objectives.* Pearson.

Bejar, I. I., Lawless, R. R., Morley, M. E., Wagner, M. E., Bennett, R. E., & Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *The Journal of Technology, Learning and Assessment, 2*(3). https://ejournals.bc.edu/index.php/jtla/article/view/1663.

Bloemers, W., Oud, A., & van Dam, K. (2016). Cheating on unproctored internet intelligence tests: strategies and effects. *Personnel Assessment and Decisions, 2*(1). https://doi.org/10.25035/pad.2016.003.

Brodersen, L., & Lorenz, R. (2020). Perceived stress, physiological stress reactivity, and exit exam performance in a prelicensure Bachelor of Science nursing program. *International Journal of Nursing Education Scholarship, 17*(1). https://doi.org/10.1515/ijnes-2019-0121.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom.* ASCD.

BVerfG, Urteil des Ersten Senats vom 15. Dezember 1983, - 1 BvR 209/83 -, Rn. 1-215 (BVerfG December 15, 1983). http://www.bverfg.de/e/rs19831215_1bvr020983.html.

Devitt, P., & Palmer, E. (2007). Assessment of higher order cognitive skills in undergraduate education. Modified essay or multiple choice questions? Research paper. *BMC Medical Education, 49*(7), 1–7.

Diedenhofen, B., & Musch, J. (2017). Pagefocus: Using paradata to detect and prevent cheating on online achievement tests. *Behavior Research Methods, 49*(4), 1444–1459. https://doi.org/10.3758/s13428-016-0800-7.

Dieterich, P., & Fischer, E. (2021). Digitale Prüfungen: Ein aktueller Überblick über die prüfungsrechtlichen Anforderungen. *Forschung & Lehre, 2,* 110–111.

Elsalem, L., Al-Azzam, N., Jum'ah, A. A., Obeidat, N., Sindiani, A. M., & Kheirallah, K. A. (2020). Stress and behavioral changes with remote E-exams during the Covid-19 pandemic: A cross-sectional study among undergraduates of medical sciences. *Annals of Medicine and Surgery, 2012*(60), 271–279. https://doi.org/10.1016/j.amsu.2020.10.058.

Eyal, L. (2012). Digital assessment literacy—The core role of the teacher in a digital environment. *Journal of Educational Technology & Society, 15*(2), 37–49. https://www.jstor.org/stable/10.2307/jeductechsoci.15.2.37.

Freie Universität Berlin. (2013). Rahmenstudien- und -prüfungsordnung der Freien Universität Berlin. *Amtsblatt Der Freien Universität Berlin, 32,* 259–270.

Friedrich, J., Handke, J., Loviscach, J., Mayrberger, K., Müller-Eiselt, R., Persike, M., Thillosen, A., & Wannemacher, K. (2015). *E-Assessment als Herausforderung: Handlungsempfehlungen für Hochschulen.* Hochschulforum Digitalisierung.

Glas, C. A. W., & van der Linden, W. J. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement, 27*(4), 247–261. https://doi.org/10.1177/0146621603027004001.

Haladyna, T., Lane, S., & Raymond, M. (Eds.). (2016). *Handbook of test development*. Routledge.

Hilkenmeier, F., & Schaper, N. (2013). *Umsetzungshilfen für kompetenzorientiertes Prüfen.* HRK Nexus.

Hoeren, T., Fischer, M., & Albrecht, J. (2020). *Gutachten zur datenschutzrechtlichen Zulässigkeit von Überwachungsfunktionen bei Online-Klausuren.* Rechtsinformationsstelle Digitale Hochschule NRW, WWU Münster. https://www.itm.nrw/wp-content/uploads/RiDHnrw_11.06.20_Gutachten-zur-datenschutzrechtlichen-Zul%C3%A4ssigkeit-von-%C3%9Cberwachungsfunktionen-bei-Online-Klausuren.pdf.

Hofmeister, W. (2005). Erläuterung der Klassifikationsmatrix zum ULME-Kompetenzstufenmodell. *Berufs- und Wirtschaftspädagogik, 8.*

Ingenkamp, K., & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik*. Beltz.

Jeremias, C. (2018). Elektronische Prüfungen, Antwort-Wahl-Verfahren und der Gesetzesvorbehalt. *Juris(1),* 25–29. https://www.juris.de/jportal/cms/remote_media/media/jurisde/pdf/juris_jm/jm_2018_1.pdf.

Mayer, H. O., Hertnagel, J., & Weber, H. (2009). *Lernzielüberprüfung im eLearning*. Oldenburg Verlag.

Rodriguez, M. (2016). Selected-response item development. In T. Haladyna, S. Lane, & M. Raymond (Eds.), *Handbook of test development* (pp. 259–312). Routledge.

Sadler, D. R. (1998). Assessment in education: Principles, policy & practice. *N Education: Principles, Policy & Practice, 5*(1), 77–84. https://doi.org/10.1080/0969595980050104.

Schmees, M., & Horn, J. (2021). *Online Prüfungen.* Elan e. V.

Schneider, D. R., Volk, B., Lehre, M., Bauer, D., & Piendl, T. (2012). Der Safe Exam Browser. Innovative Software zur Umsetzung von Online-Prüfungen an der ETH Zürich. *Digitale Medien - Werkzeuge für exzellente Forschung und Lehre, 61,* 431–441. http://www.waxmann.com/?eID=texte&pdf=2741Volltext.pdf&typ=zusatztext.

Schulz, A. (2017). *E-Assessment-Center im Vergleich: Voraussetzungen und Kosten für die Einrichtung verschiedener E-Assessment-Center im Vergleich.* Technische Universität Dresden, Medienzentrum. https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-224532.

Schulz, A., & Apostolopoulos, N. (2011). EExaminations Put to Test. Potenziale computergestützter Prüfungen. *Hamburger ELMagazin, 7,* 37–39.

Seddon, M. (1978). The properties of bloom's taxonomy of educational objectives for the cognitive domain. *Review of Educational Research, 48*(2), 303–323.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2017). *Qualifikationsrahmen für deutsche Hochschulabschlüsse. Kultusministerkonferenz 16.02.2017*. https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2017/2017-03-30_BS_170216_Qualifikationsrahmen.pdf.

Senatskanzlei - Wissenschaft und Forschung. (2020, September 30). *Berliner Stufenplan für den Hochschulbetrieb unter Pandemiebedingungen* [Press release]. https://www.berlin.de/sen/wissenschaft/_assets/berliner-stufenplan-fuer-den-hochschulbetrieb-unter-pandemiebedingungen.pdf.

Steger, D., Schroeders, U., & Gnambs,T. (2020). A meta-analysis of test scores in proctored and unproctored ability assessments. *European Journal of Psychological Assessment, 36*(1), 174–184. https://doi.org/10.1027/1015-5759/a000494.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *The Phi Delta Kappan, 77*(3), 238–245. https://www.jstor.org/stable/20405538.

Terhart, E., Baumgart, F., Meder, N., & von Sychowski, G. (2009). Standardisierte Prüfungsverfahren in der Erziehungswissenschaft. Kontext, Formen, Konsequenzen. *Erziehungswissenschaft, 20*(38), 9–36.

Tiemeier, A., Stacy, Z., & Burke, J. (2011). Using multiple choice questions written at various bloom's taxonomy levels to evaluate student performance across a therapeutics sequence. *Innovations in Pharmacy, 2*(2), 1–11.

Vedhara, K., Hyde, J., Gilchrist, I., Tytherleigh, M., & Plummer, S. (2000). Acute stress, memory, attention and cortisol. *Psychoneuroendocrinology, 25*(6), 535–549. https://doi.org/10.1016/S0306-4530(00)00008-1.

Nils Hernes, is a digital assessment specialist at the Center for Digital Systems (CeDiS) at Freie Universität Berlin. He focuses on the digital transformation of assessments with regard to didactics, logistics and high-scaling workflows. Within the project "E-Assessment Alliance", Nils Hernes develops standards and modes of operation for digital assessments across the Berlin university landscape.

Student Feedback in Hybrid/Online Teaching: Relevance, Approaches and Practices

Alexandra Dorfer, Gudrun Salmhofer, and Lisa Scheer

Abstract

In student-centered teaching, students are seen as active partners of teachers in shaping and designing their learning processes. According to this view, students have to be able to contribute responsibly and constructively to courses. With their feedback they make an essential contribution to the development of teaching and learning—whether face-to-face, hybrid, or online. In this paper, the difference between formative feedback and summative evaluation is being discussed alongside guiding questions, also considering different teaching scenarios. In addition to a theoretical and empirical underpinning, we present practical examples of how feedback can be systematically implemented in teaching and learning scenarios. These recommendations are supported by various empirical studies conducted at the University of Graz.

Keywords

Course evaluation • Feedback considerations • Feedback instruments • Student feedback

A. Dorfer · G. Salmhofer (✉)
Department of Academic Services, University of Graz, Graz, Austria
e-mail: Gudrun.salmhofer@uni-graz.at

A. Dorfer
e-mail: Alexandra.dorfer@uni-graz.at

L. Scheer
Competence Center for University Teaching, University of Graz, Graz, Austria
e-mail: Lisa.scheer@uni-graz.at

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_4

1 Introduction

Classroom teaching takes place in familiar teaching and learning environments such as laboratories, lecture halls, and seminar rooms. It offers personal face-to-face contact with students and therefore the opportunity for instant verbal and non-verbal feedback and reaction to disturbances. In spring 2020, the pandemic led to the necessity of an unforeseen and sudden change in teaching and learning concepts and practices for many. Instantly, teachers had to switch to digital learning environments, often without or with little previous knowledge and competences in online teaching. This pandemic-induced shift to hybrid and online course formats led to many unexpected situations where teachers as well as students had to face challenges and deal with uncertainties. In contrast to classroom teaching, online teaching strongly reduces non-verbal feedback from students to teachers and causes communication changes: it mainly comes along with more asynchronous than synchronous communication.

Especially in situations like these, dialogue and communication between teachers and students are vital for effective teaching and learning. Not only since Hattie's meta-analysis (2013) has it been known that individual, constructive feedback can make a decisive contribution to learning. In this article we highlight the importance of student feedback for the successful implementation and improvement of (hybrid/online) teaching and learning processes. We outline differences between formative feedback and summative evaluation and point at the advantages of feedback, especially in online and hybrid settings that are (or were) new and unfamiliar. Because "[n]o evaluation is done without a mandate" (Franz-Özdemir et al., 2019, p. 5), feedback and evaluation practices can be discussed from the viewpoints of control and best possible facilitation of learning processes, governance and personal development as well as self-reflection and quality assurance/development. After tackling these meta perspective questions, we address the WHY, HOW and WHEN of feedback processes and present specific suggestions accompanied by empirical insights and experiences from the University of Graz.

2 Evaluation, Feedback and the Shift from Teaching to Learning in Higher Education

"Obtaining feedback from students is an essential requirement of reflective teaching, allowing teachers to refine their practice and to develop as professionals" (Huxham et al., 2008, p. 675). Interest in improving university teaching has grown steadily since the early 2000s—partly driven by external requirements through the Bologna Reform and requirements of external quality assurance, partly driven by institutional strategies to strengthen teaching competence and by the interest of individual teachers in good teaching and learning. The perspective shift initiated by the Bologna Reform to rethink teaching from the viewpoint of learning has brought the interactions and relations between teachers and students into stronger focus. This change in the teaching and learning culture also demands new requirements in terms of feedback.

Standardized quantitative methods essentially take a teacher-centered view by asking about student satisfaction with the performance of the teachers, teaching material and such. Such a teacher-centered evaluation process, traditionally conducted via questionnaires, tends to lead students to reflect more on the input of the teachers than on the learning processes and teachers to reflect on their own teaching behavior rather than to focus on more student-centered learning designs (Franz-Özdemir et al., 2019). While evaluations have been the means of choice at many universities in the past and students have been assessing teaching or the teachers summatively, feedback tools are now increasingly being used, researched and discussed in publications. This shift towards qualitative feedback instruments, which focus on student learning, also comes from a dissatisfaction with questionnaire-based course evaluation which has its limits regarding the joint endeavor of teaching and learning.

As universities became more autonomous during the past two decades, new support units such as quality assurance departments or centers of teaching and learning were established, which play an ongoing important role in the enhancement of teaching. To a large extent it was these units and centers that strongly guided and supported the maintenance and improvement of teaching during the COVID-19 crisis. The inevitable ad-hoc-switch to online teaching was a powerful stimulus for the digitization of teaching, which otherwise would never have received this level of importance in such a short time. Only in the coming years will it become clear how sustainably the experiences can be anchored, depending above all on appropriate strategies and framework conditions.

Before diving into the practice of collecting student feedback in hybrid/online teaching settings, we want to elaborate further on the aforementioned poles in connection with the concepts of feedback and evaluation.

2.1 What is Feedback?

“Feedback” has many different meanings and is used in various contexts—just like “evaluation”. In a broader understanding, feedback is “information provided by an agent (e.g. teacher, peer, book, parent, self, experience) regarding aspects of one’s performance or understanding. [...] Feedback is thus a ‘consequence’ of performance” (Hattie & Timperley, 2007, p. 81). Feedback can be seen as something the teacher does: giving feedback on student activities and assignments. Feedback can also be seen as something the student does: giving feedback on other students’ activities and assignments (peer feedback) or giving feedback on the instructor’s activities and the learning process (course feedback). Besides the fact that these examples already include three different understandings and practices of feedback, the use of such descriptions is problematic. It “betrays our conceptualisations of feedback” if feedback is solely viewed as something being given and/or received, because the offer of giving something “contains an undercurrent that casts feedback as a gift from you to them, rather than an active process in which they are engaged, or a dialogue between the two of you” (Winstone et al., 2021, p. 2). If teaching and learning are understood as a joint process and responsibility, this should be reflected in our communication about feedback and how it is practiced. Therefore, as Winstone et al. (2021) point out, it is necessary to check our understanding and usages of feedback: is it an information-centric or a process-centric understanding? Are we talking about feedback information or feedback processes? Consequently, it may be most useful to our discussion here to understand feedback as “a process in which learners make sense of information about their performance and use it to enhance the quality of their work or learning strategies” (Henderson et al., 2019, p. 1402). Interestingly, this definition implies that instructors are seen as learners, too.

2.2 What is Evaluation?

Commonly, evaluation is understood as the assessment of a process as well as the result of the assessment of the value of a product, a person, or a programme (Gollowitzer & Jäger, 2009). It is therefore a summative assessment of a process and/or object. In the relevant literature dealing with evaluation in the teaching/learning context, many different definitions of evaluation can be found. Despite their heterogeneity, according to Knödler (2018), most definitions have the three following components in common: (1) empirical generation of knowledge, (2) linked assessment, and (3) to make purposeful decisions with different accentuations (approaches, objectives, or objects). The purposes of evaluation mostly include the three aspects of clarification, legitimation and control. However, especially in the field of evaluation of study programmes, evaluation can also fulfil the following four functions: insight, control, legitimation and development (Stockmann & Meyer, 2014). Particularly the last function, that of development, is an important one when it comes to the improvement of teaching and the closing of the classic cycle of Plan-Do-Check-Act in quality management. The aspect of development also points at the aspect of timing: formative evaluations are carried out during the teaching process, summative ones at the end of the process. Formative evaluation has the purpose of revealing strengths and weaknesses of measures and serves to support, improve and further develop ongoing projects and programmes. Summative evaluation provides data on the completed measure and has the function of balancing, assessing and reviewing success and impact. Results of formative evaluations usually remain within the institution and its programme development. The results of summative evaluations, however, are also made available to interested parties outside the university (e.g. ministries) and contribute to a better understanding of the study programme (Knödler, 2018). A combination of both forms of evaluation is possible and sometimes the case (Stamm, 2003). Formative evaluations are often used in order to achieve better results in subsequent summative evaluations (Scriven, 1972).

2.3 How is Evaluation Regulated?

High-quality teaching is in the interest of both students and teachers, but also of a broad public and the state. These different stakeholders each have their own interests and intentions in evaluation (Döring, 2006). While course evaluations were already widespread in countries with Anglo-American education systems, such as the USA, Great Britain, the Netherlands, or Australia, the beginnings in Austria and Germany lie in the 1990s. In Austria, the compulsory evaluation of courses was legally anchored in 1993 with the Federal Law on the Organisation of Universities (UOG 1993, § 18 (1)) which was replaced by the Universities Act (UG) in 2000. This act reorganised wide areas of university law and, among other things, contains paragraph 14 on evaluation and quality assurance that obliges universities to set up a comprehensive quality management system for continuous quality and performance assurance of their entire range of services. In addition to the UG, other legal frameworks such as the University Financing Ordinance (Universitätsfinanzierungsverordnung UniFinV) or the University Quality Assurance

Act (Hochschul-Qualitätssicherungsgesetz HS-QSG) regulate the design of quality assurance measures. Interestingly, all these legal writings explicitly mention the term "evaluation", but none mentions "feedback".

Similar to Austria, the beginnings of legally required course evaluations involving students in Germany date back to the 1990s. The main drivers were the universities' increasing competitive orientation, international positioning, autonomy demands, profile-building efforts and renewed efforts to improve the quality of teaching (Auferkorte-Michaelis & Selent, 2006). Prior to that, there was a tradition of student critique of courses dating back to the student revolution of the 1968s, which related to the demand for active co-determination of students (Döring, 2006). The effort for quality teaching by teachers in Germany dates to 1971, when the "Arbeitskreis für Hochschuldidaktik (AHD)" was founded (Kröber & Thumser, 2008). The Austrian Society for Higher Education Didactics (ÖGHD) is based on similar considerations and was founded a few years later, in 1977.

2.4 What Are the Advantages of Feedback Over Evaluation?

While universities pursue different approaches and focuses in course evaluation (student competence orientation, student satisfaction, etc.), they mostly use questionnaires with Likert scales that allow summative statements at course level (Alderman et al., 2012). This methodological approach entails some shortcomings: when using standardised scales, such as the Likert Scale, methodological issues regarding bias effects have to be considered (McClain et al., 2018; Mitry & Smith, 2014;). With regard to the implementation, it must be noted that the response rates, especially with online forms and with optional student participation, are often low, which raises the question of the significance and representativeness of the evaluation results. Another practical shortcoming with regard to deriving follow-up measures is that summative data is collected, "which has little value for the students who provide the data, and which gives little impetus to instructors to take any immediate action for improvement in their teaching" (Sozer et al., 2019, p. 1004).

One advantage of feedback compared to standardized evaluation instruments is the flexibility of the former. Usually, standardized evaluation contains fixed questions and items and delivers mostly quantitative results: percentages, means, standard deviations, etc. For such course evaluation results to be useful for the ongoing development of teaching competences as well as teaching and learning concepts, instructors favor additional discussions and talks with students, often on the basis of evaluation results (Wolf & Kothe, 2020). Evaluation results, therefore, can be a stimulus and basis for a deeper reflection of the teaching and learning processes. Additionally, critical student comments and unexpected evaluation results may imply the necessity of checking back. "Teachers find the interactive discussion of the criticism in the feedback discussion relieving because it makes the criticism easier to understand or makes it clear that it is not meant personally or was expressed arbitrarily" (Wolf & Kothe, 2020, p. 188, own translation). This also points at the advantages of feedback over evaluation. Unlike standardized, summative evaluation formats, feedback instruments are suitable for accompanying the learning process, which is jointly designed by teachers and learners and should take both perspectives into account. But this can only be achieved with a thorough joint reflection, a conversation, a dialogue. Only feedback instruments represent a useful means to gather reflective, deep student perspectives and ideas on how to establish conditions for quality teaching and learning. They allow instructors to add their point of view and come to a joint understanding of the endeavor. Thus, another advantage of feedback over evaluation besides providing "the flexibility needed to identify innovative teaching" (Huxham et al., 2008, p. 676) is that it is less judgemental but rather developmental. Two other characteristics of feedback instruments that standardized questionnaires are lacking are "specificity (ability to identify aspects particular to individual tutors and classes) and discretion (allowing clear indication of what students considered most important)" (Huxham et al., 2008, p. 685). Because of the potential for very open and broad questions, feedback instruments are very useful in gathering information on unknown aspects which cannot be achieved in the same way by standardized questionnaires (Huxham et al., 2008).

The aspect of improvement or development, as described in the format of formative evaluation, is the most important feature of feedback as we understand it at the University of Graz. In our opinion, the most important effect of all measures for quality improvement in teaching is the emergence of a discourse about teaching—the initiation of a dialogue between students and teachers, which connects their mutual interest in the further development of teaching and realizes it through action. This is made possible more effectively through flexible feedback instruments than through course evaluations and requires the willingness of teachers to listen to their students' voices, to engage in a dialogue, and to implement reasonable instructional improvements. In the next section, we will discuss ways of creating such a feedback dialogue, and which challenges may be faced in the process.

3 Feedback in Hybrid/Online Teaching Scenarios

In this section we will address the WHAT, WHY, HOW and WHEN. First, we suggest that instructors reflect on a few questions before diving into feedback processes. Next, we will outline a few reasons for engaging in student feedback processes, then describe tools and questions that can be used and adapted accordingly, and finally tackle the question of when to ask students for feedback.

3.1 WHAT to Preconceive Before Initiating Student Feedback Processes?

Based on research results and academic debates we propose a few questions and reflection points that instructors should consider prior to collecting feedback. They present a meta perspective on the feedback process and put it in relation to group dynamics, teaching and learning climate, disciplinary cultures, and the Higher Education Institution (HEI) and its strategies. Among others, they include the following:

a. How reflective, communicative, etc. are my students? How well can they give constructive feedback? How well can they reflect on their learning processes and put these reflections into words?

From our experience, students are usually very communicative when invited to a feedback process. From the first year on, they are generally able to articulate their opinions and wishes. There is no need to fear online feedback, because according to Gakhal & Wilson (2019) students are not more critical when giving online feedback than paper feedback. Also, it does not seem that social media communication has negatively influenced students' course evaluation. Nevertheless, deep reflections of their learning processes, for example in longer written format, should not be expected without the respective introduction. Therefore, students need guidance on how to provide constructive feedback. They need clear orientation and should know which criteria to base it on—this has to be provided by the teacher (Gakhal & Wilson, 2019; Mojescik, 2017). This ensures transparency and offers students the opportunity to specifically prepare for it. "Providing students with guidance would encourage a constructive feedback culture in higher education institutions and support students as future employers/employees in preparing for performance appraisals" (Gakhal & Wilson, 2019, p. 486). However, a constructive feedback culture that allows for the voicing and arguing of opinions is often not part of the curriculum and therefore not that easy to implement. The following student quote[1] could be interpreted in this way: “The degree programme is far too much about learning content by heart and far too little about really understanding it. There was also far too little discourse” (University of Graz, 2020a, own translation). Therefore, it is all the more important for the lecturer to provide orientation and to clearly communicate the criteria for constructive student feedback.

b. How do I perceive the relationship between my students and me? What is the climate like? What do I trigger in my students when I ask them for course feedback?

Initiating a feedback dialogue has an impact on the teacher’s relationship with students and on students’ perception of their role and responsibility in the HEI. Feedback instruments must fit the respective teaching style and understanding of teaching in order to be authentic. Only then can students be convinced that there is real interest in their feedback and that the aim is a joint improvement of teaching and learning processes. A prerequisite for the successful use of feedback instruments by students is a trusting and appreciative learning climate. Instructors should be aware that anonymous instruments can lead to distorted, unfair feedback and create personal conflicts (Mojescik, 2017).

c. How do I deal with student feedback and suggestions for improvement? How much room and time for change do I have in my course?

Teachers should take into account that feedback is always a subjective assessment and an individual perception of a situation (Mojescik, 2017). This means that feedback can be very heterogeneous, just like the group of students. Usually, not all suggestions can and will be realized, for example because of restricted resources or because teachers refuse to make changes against their professional judgement (Flodén, 2016). If feedback cannot or will not be implemented, it is advised to disclose this and explain to the students why not, like one University of Graz law professors explains (University of Graz, 2020a).[2] Transparency in this

[1] The quote is from the anonymous graduation survey. Since 2020, alumni of the University of Graz are asked to fill it out after completing their studies.

[2] Prof. Markus Steppan, Institute of the Foundations of Law, and Assoz. Prof. Jürgen Pirker, Institute of Public Law and Political Science, both University of Graz, were interviewed by Simone Adams as part of the Centre for Digital Teaching and Learning interview series digiTales.

regard coupled with an explanation of the didactical concept is beneficial for the relationship between students and instructor as well as for students' competences. It is the basis for a relationship on a par and offers the opportunity to start a dialogue on teaching and learning.

In general, lecturers are interested in incorporating students' feedback to further develop their course designs. However, as a lecturer stated,[3] it is sometimes difficult to draw a line between constructive student feedback, i.e. justified suggestions for improvement, and a nice-to-have or unrealistic student "wish list": "In my opinion […] most professors, almost all of them, are interested in the students' opinions and also try to incorporate feedback. The question is, which feedback do I listen to? More feedback is not always better. Many students are interested in studying as pleasantly as possible, not in learning as much as possible. That this is the case is certainly a problem of prior education and the general school culture; it's hard to expect universities to smooth that out. Should I base my lectures on their feedback? I think that would be fatal in the long run […]" (University of Graz, 2020d, own translation). Therefore, lecturers do not only have to consider how much time and space can be given to student feedback, but also which limits, and where, have to be drawn from a didactic, conceptional, or organizational point of view.

The answers to questions regarding the preconditions and effects of feedback processes will lead teachers directly to the questions of purposes, how to collect student feedback, which tools to use, and which questions to ask.

3.2 WHY Initiate Student Feedback Processes?

Particularly in online or hybrid teaching, where direct, immediate contact between teachers and students is lacking, it is important to use suitable feedback instruments to obtain student feedback. One teacher sums it up as follows: "I have continuously collected feedback using different methods, several times per course. In 'regular' teaching settings, when I face my students in the lecture hall, I can see immediately whether they pay attention, whether they have understood the material, whether they are attentive, distracted or tired. Online teaching makes

[3] A series of three anonymous teaching surveys was conducted at University of Graz amongst teaching staff between March and July 2020 (University of Graz 2020a).

it necessary to explicitly ask for feedback, which I otherwise receive implicitly through facial expressions, gestures or behaviour" (University of Graz, 2021; Focus group[4]).

Collecting student feedback can serve several purposes:

a. *Mirror*

Student perspectives are an important component in the reflection of digital and hybrid learning environments. Student feedback functions as a mirror and shows how didactical approaches as well as teaching and learning methods are perceived. Another University of Graz law professor, for example, asked his students for feedback when first trying out PowerPoint slides with audio comments at the beginning of the summer term 2020 (University of Graz, 2020c). Only after receiving positive feedback he continued, as such slides are very resource-intensive.

b. *Illumination*

Student feedback can be used to gain insight into the unknown and hidden: the black boxes regarding the students, e.g. their learning habits, progress of group work, understanding of course content, course atmosphere, classroom climate or their challenges. At the beginning of his classes, one of the interviewed law professors likes to find out his students' interests, concerns and prior knowledge in order to better assess the starting point and the development of the learning processes. In the middle of the semester he checks what is going well, what could be improved, what is least understood so far and what is still wanted or needed. He then draws on these answers to design the subsequent course units, e.g. considering what should be repeated again (University of Graz, 2020b). In extraordinary times such as during a pandemic or in online teaching, it could be of special interest and necessity to ask students in these regards, since the usual opportunities to chat with them, meet them and get to know them are missing.

c. *Reflection and communication*

[4] Within the project "Workplace Health Promotion" (Betriebliche Gesundheitsförderung), two focus groups with eleven lecturers were conducted, in which they were asked, among other things, about their teaching activities since the beginning of the pandemic.

Feedback tools promote students' ability to reflect and communicate, which is a central competence both at university and in other environments. When learning is considered to be a joint endeavor, students need to have knowledge about their learning habits: they have to be able to communicate their needs and create a dialogue with others.

d. *Empowerment and agency*

Giving students a voice in the design of teaching and learning settings means to promote their empowerment and agency and to support the collaboration between students and teachers. The aforementioned second law professor invited students to share their wishes, needs and impressions regarding online teaching and learning at the beginning of the summer term 2020 (University of Graz, 2020c). The first law instructor explains that his students very much appreciate being asked and that their perspective is incorporated into the further course design (University of Graz, 2020b).

e. *Student/teacher rapport*

In contrast to summative course evaluation, which at many HEIs has to be carried out compulsorily and produces semi-visible results, the use of feedback instruments is usually voluntary and the confidential results remain with the teacher (and students). This is intended to create a positive relationship between students and teachers, a climate of openness regarding teaching/learning settings and creativity for adaptations and improvement—without fear or anxiety that critical student feedback will have negative consequences. When teaching/learning is lived as a joint process and feedback is viewed as part of this joint process, it brings students and instructor closer together and makes teaching and learning a joint responsibility.

f. *The fun factor*

Feedback tools can be playful in the form of online quizzes and games and can therefore be used to bring fun elements into the classroom. By introducing a change and diversion they can increase motivation, activity and fun in learning processes and offer opportunities to participate (University of Graz, 2020b).

After having provided answers to the question of why, we are now going to tackle the how and introduce tools for course feedback which can be used online just as much as in any other teaching scenarios.

3.3 HOW to Initiate Student Feedback Processes?

The many different feedback tools that exist can be chosen in accordance with internal and external factors, from student and instructor preferences to specific goals, interests and time resources. To ensure anonymity in hybrid/online settings, instructors could use Google Docs, wordcloud, digital whiteboards such as miro or collaborative web platforms such as padlet. In addition, learning platforms usually allow anonymous messages too.

(1) **Ad-hoc feedback**

Ad-hoc feedback tools are quick, usually little time-consuming and can be used for a variety of purposes at different times in the semester (for possible questions see Table 1). They are learner-centered, because they relate directly to the activities of the learners and may require debriefing between teachers and students.

Some examples of ad-hoc feedback tools are:

- *flash light:* When using the flash light method, a question is posed to the audience (it can be visualized for support, e.g. PPT or chat). Each participant may answer the question, but does not have to. Answers are not evaluated and there is no discussion, the teacher just listens.
- *muddiest point:* When using the muddiest point, the teacher receives indirect information on the student learning processes. At the end of a unit/topic/course, participants are asked to write down the "muddiest point", i.e. the most unclear point. In addition, teachers can ask for an explanation as to why this particular point was perceived as particularly vague.
- *one-minute paper:* The one-minute paper is a quick and targeted way to collect written feedback on the current state of knowledge, to assess the course or learning progress, or to gain feedback on questions that have remained unanswered. Students are asked to formulate feedback on (a) question(s) in written form within one minute. In our experience, one minute can be (too) short so it is advised to extend the period to around three minutes.
- *Start/Stop/Continue:* When using Start/Stop/Continue, students are asked to state what they liked and should be continued (continue), what they liked less and should not be continued in the future (stop), and what improvements/renewals they suggest (start).[5]

[5] See Hoon et al. (2015) for an evaluation of this tool.

Table 1 Possible ad-hoc feedback questions

Possible questions	What would you like to discuss in the next course session? What should/must be explained again in the next session? What was presented/discussed in the last session? What was the most important insight of the last session for you? What was the most important insight for you in today's session? What were the three most important insights of today's session for you? In what non-university context might it be relevant what you learned today? Give an example of the practical relevance of what you learned today! By … is meant (a) (b) (c) (d) By … is meant … Today's session made me think about: … After the last session, I had to think about: … After the last session, I asked myself the following question: … In the last sessions of this course it became clear to me that …

(2) **Group discussion**

A group discussion is a moderated talk between a group of students in order to collect information on learning processes. The group discussion is initiated by a stimulus or prepared question(s). A central goal is to bring up as many different facets of a topic as possible and to allow as many students as possible to express their opinion, although the group should not be much bigger than seven to eight students. Depending on the relationship between instructor and students, the discussion can be moderated internally (instructor, non-participating student) or externally (colleague, student representative). It is possible to conduct such a discussion using online conference tools, but the setting will have effects on the group dynamic and the discussion intensity will most likely differ from a group discussion among physically present students.

(3) **Teaching analysis poll (TAP)**

TAP is for teachers who want to find out how students perceive their learning processes in the middle of the semester and what could be changed to optimize student learning in the second half of the course (Frank et al., 2011). In the absence of the teacher, TAP is usually conducted by an external person at the beginning or end of a course unit. First, students are asked to answer three questions in small groups: (1) What supports your learning in this course the most?

(2) What hinders your learning? (3) What suggestions for improvement do you have for the hindering points and beyond? In the plenum, three to five majority points are collected for each question and forwarded to the teacher afterwards. In the physical classroom, the small groups put down their points on a TAP handout and the plenum results are visualized and documented on such a sheet to ensure transparency. In hybrid/online scenarios, the steps are taken accordingly digitally and in online conference tools.

(4) **Written review**

A Written Review is a feedback tool for teachers who want to encourage their students to reflect on their learning process in written form. It helps the teacher to trace the learning processes of the students in order to draw conclusions about the achievement of the teaching objectives, teaching methods, etc. (for possible questions see Table 2). It is (hand)written anonymously in classroom scenarios and in hybrid/online settings instructors have to offer a way of forwarding it anonymously, e.g. through a student representative, a collaboration tool or a learning platform.

(5) **Representatives' feedback**

Three times a semester—at the beginning, in the middle and at the end—the teacher meets with elected course representatives to jointly reflect on the course of the semester. The results and contents of these three meetings are recorded in writing, and at the end a report is prepared by the representatives, which is agreed with all course participants. A short version of this report is available for the students in the next semester ("Letter to the next semester") (Table 3).

Table 2 Possible written review questions

Possible questions	Name the learning objectives of this course and describe 1. How these objectives were met; 2. Why which objective was not met and what you take away from that; 3. What you learned apart from the original learning objectives
	Over the course of the semester, I realized that … In the course of this lecture/seminar/lab/etc., I learned (about …) that …

Table 3 Possible representatives' feedback questions

Possible questions	Meeting 1 (beginning): What suggestions do you have for improving the learning objectives I have formulated? Are the objectives understandable? What do you think of my expectations for students? Meeting 2 (middle): What are your impressions of the course so far? What supports your learning process well so far, what is still missing? Is there anything that needs to be changed in the learning objectives? How is the exchange with the colleagues going? How are you doing in your role as deputy? Meeting 3 (end): Looking back, how do you see the learning process in this course (what was learned and how)? Were the learning goals achieved? Why (not)? What was the climate and workload like? What could be done better next time (by students and teacher)?

(6) **Quizzes and questions regarding course topics**

Besides using the muddiest point, instructors can also ask direct questions on the course content to find out about (un)successful learning processes. Online tools such as Kahoot!, Mentimeter, Socrative or Slido support learning processes in a fun and entertaining way.

After presenting thoughts and ideas on why and how, the last of the three main questions regarding feedback is when to realise it.

3.4 WHEN to Initiate Student Feedback Processes?

Although many feedback instruments can be realised with little time expenditure, their integration into the time schedule of the course can be a challenge (Mojescik, 2017). Not only the implementation itself, but also the preparation and follow-up as well as, above all, the effort for any didactic/content-related improvement must be taken into account. So, when is the appropriate time to collect student feedback?

Basically, a distinction can be made between feedback carried out during a course unit and outside the classroom/online conference room. An advantage of feedback being part of a course unit is that (almost) all students who are present give feedback. Outside of the course time, the response willingness depends on the time resources and motivation of the students (University of Graz, 2020b).

Regardless of whether feedback is collected during or after a course unit, the concrete timing also plays an important role. At the beginning of the course, it makes sense to use other feedback formats than during or towards the end of the course. At the beginning of a course, the focus lies on gathering information about the students, their expectations, their previous knowledge, or their motivation. In this way, the teacher can respond to the needs of the group and prevent possible disruptions (Mojescik, 2017). Mid-term feedback is particularly suitable for obtaining concrete feedback on the teaching/learning method, the students' learning process, the delivery format, or the didactic implementation. Mid-term feedback formats create opportunities for students and teachers to engage in a dialogue about teaching and learning and to make any necessary mid-stream corrections. A study by Nancy Hunt (2012) investigated the influence of mid-semester feedback on the classroom experience. It found that both students and teachers were aware of the benefits of mid-term feedback: students appreciated the opportunity to voice their opinions at a time when mid-course corrections were still possible. Teachers indicated that communication with their students had improved, resulting in a friendlier, more open teaching and learning environment. When teachers show that they take student feedback seriously and are interested in the development of their teaching, they set impulses for an open and improved classroom experience. If formative feedback is collected towards the end of the course, on the one hand student reflection processes are stimulated. On the other hand, teachers receive information about possible conceptual adjustments which can be used for the design of future courses. In addition, further conceptual ideas can be put up for discussion and students can be asked for their opinions (Mojescik, 2017).

Now that questions on the timing of feedback have been dealt with, the next section presents an overview of possible thematic focuses and exemplary feedback questions.

4 Possible Feedback Questions

Formative feedback that is collected during the course usually contains "open-ended questions that trigger more elaborative feedback about what is going on in a class than that from end-of-semester evaluations with Likert scale-type questions" (Sozer et al., 2019, p. 1003). In addition to the feedback questions already mentioned in the HOW section above, an overview of possible questions with regard to specific topics can be found in the following Table 4. This list is to be understood as a suggestion and can of course be supplemented and adapted as needed.

Table 4 Further possible feedback questions

Learning process	What have you been missing (so far) for an optimal learning process, what could be improved? What could you as a student have done better (so far) to experience a more positive learning process? How can you as a student contribute to improving your learning in this course? What advice would you give to future participants in this course? What made learning difficult for you in this course? At what stage of your learning were you particularly challenged and why? What additional things would you have needed at this stage? What phase of your learning process do you look back on with satisfaction and why?
Teaching and learning methods	Which teaching/learning methods of the course have supported and advanced your learning (so far)? Which teaching/learning methods have helped you most (so far) to improve your subject/method/social skills? What has helped you learn about … in this course? What do you learn most from in this course (about …)? What suggestions for change that would support learning in this course do you have for the instructor? How do you assess the learning effect in this course through the measure … (Moodle, weekly feedback session, etc.)? Do you think the measure … (Moodle, weekly feedback session, etc.) would increase your learning success in this course? Why do you (not) think so? Which teaching/learning methods are good, which are not that good to support your learning?

(continued)

Table 4 (continued)

Learning objectives	What were the learning objectives of this course and how were these learning objectives achieved? What can you do—especially AFTER the seminar/outside the course attendance—to achieve these learning objectives? Why were learning objectives not met (which ones) and what conclusions did you draw from this? What did you learn beyond the original learning objectives? What lessons do you take away from the course (so far)? What was the "fuzziest/muddiest point" (the most incomprehensible/unclear/diffuse aspect) of the past course unit for you? (Give reasons if necessary) What specific issue would you like to clarify or have clarified in connection with the "title/topic of the course" in the rest of the semester? What do you think you have learned well? (Give reasons or examples) Where do you still see challenges? Where could you still learn something (important to you)? Which learning objectives are easy to achieve and which are difficult? Please give reasons why Which learning goal set by the teacher was difficult or impossible to achieve and why? What would have been necessary for you to achieve this learning goal?
Sustainability	What insights from the course will you take with you for the future? What will you continue to think about after the course? What three aspects will you take away from the course that you will still need in the future? What goals have you set for yourself as a result of taking the course and how do you plan to achieve these goals?

(continued)

Table 4 (continued)

Student centering	How open was the teacher to student feedback? Were measures already taken during the course that relate to student (interim) feedback?
In general	What did you like or dislike about the course and why? What suggestions do you have for improving this course? [keywords: *organizational aspects (room, technology, conference tool), general conditions, form of teaching, content, teaching materials (script, Moodle, videos, literature), examination performance, teacher, cooperation of the participants]*: What comes to your mind spontaneously when you hear the chosen keyword? What positive and what negative aspects can you name with regard to the keyword? What suggestions do you have for improving aspects relating to the keyword?

5 Conclusion and Implications

The COVID-19 crisis and the related lockdowns caused ruptures in regular university teaching in 2020 and the beginning of 2021. Even if not everything worked out smoothly, the experiences have shown that successful teaching/learning takes place online just as much as in face-to-face classes. As we know, the majority of teachers can imagine continuing to integrate digital elements in their classes. Thus, Thus, the opportunities offered by digitally enhanced teaching/learning will be continue to be used in the future, not as a panacea but didactically justified, technologically implementable, and organizationally integrable (Kopp, 2021; Fleischmann, 2020). In this way, the strengths of face-to-face teaching can be combined with the opportunities and advantages of digitization—on a broader and better accepted basis than before the pandemic. However, it must be emphasized that virtual as well as hybrid settings are special teaching and learning spaces (Entner et al., 2021): in face-to-face teaching, unconscious processes run parallel to consciously designed teaching and learning processes, which can have a learning-promoting effect especially in group dynamics and on the relationship level. In the virtual space, these processes do not occur automatically and must be consciously initiated. In hybrid settings, the face-to-face units should be used to trigger these accompanying processes.

In line with many scholars, we recommend engaging in feedback processes as they are a proper and common measure to support teaching and learning and its improvement, in non-pandemic times as well as when (un)voluntarily trying out something new. Feedback tools aim at improving interaction, stimulating discussion and dialogue between teachers and students, promoting communication, and thus providing valuable impulses for the further development of teaching and learning. Unlike an evaluation, feedback is a process which includes instructors and students alike. Jointly designed by teachers and learners, it is suitable for accompanying a learning process and takes all perspectives into account.

Despite providing many hands-on recommendations and therefore directly addressing instructors here, we see student feedback processes as part of institutional practices, strongly linked to a certain understanding of teaching and learning. From this viewpoint, feedback always has to be seen as part of a culture: a feedback or quality culture that tackles hierarchies and prevailing understandings of "expertise" in HEIs.

> "[…] 'quality culture' first and foremost can be a tool for asking questions about how things work, how institutions function, who they relate to, and how they see themselves. The dominant problem with quality culture as it is used today is that the concept is thought of as the *answer* to challenges, while in reality, it is a concept for *identifying* potential challenges" (Harvey & Stensaker 2008, p. 438).

Teaching is a highly demanding and time-consuming activity, and the integration of new teaching formats demands specific competencies from teachers. Although the digitization of teaching and learning has been a much-discussed topic since even before the pandemic, this has not yet led to corresponding broad impact and acceptance. The reasons for this include a lack of institutional framework conditions, (media) didactic know-how, and incentives to enrich teaching with digital elements (Schumacher et al., 2021; BMBWF, 2021; Entner et al., 2021; Kopp, 2021). It will therefore continue to be important to support teachers in their work in the future. Resources must be made available, appropriate institutional frameworks must be created, but most importantly, a strategy for teaching must be negotiated with institutional stakeholders. But it is not just teachers who need support. It should not be forgotten that students also face challenges in virtual as well as in hybrid teaching formats and are more challenged to maintain their motivation and engage in self-discipline.

References

Alderman, L., Towers, S., & Bannah, S. (2012). Student feedback systems in higher education: A focused literature review and environmental scan. *Quality in Higher Education, 18*(3), 261–280. https://doi.org/10.1080/13538322.2012.730714.

Auferkorte-Michaelis, N., & Selent, P. (2006). Feedback-Evaluation in Lehrveranstaltungen als dreistufiges Verfahren. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer, & J. Wildt (Eds.), *Neues Handbuch Hochschullehre* (I 1.2.). duz Medienhaus.

BMBWF. (2021). *Distance Learning an österreichischen Universitäten und Hochschulen im Sommersemester 2020 und Wintersemester 2020/21.* https://www.bmbwf.gv.at/dam/jcr:3db6ff5e-68f7-43d0-a31f-0e667d258d69/210701_WF048_21%20-Distance%20Learning%20an%20Unis%20und%20HS%20im%20SS20%20und%20WS20_21_bf_FINALE_VERSION.pdf.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354.

Chong, S. W. (2021). Reconsidering student feedback literacy from an ecological perspective. *Assessment & Evaluation in Higher Education, 46*(1), 92–104. https://doi.org/10.1080/02602938.2020.1730765.

Crimmins, G., Nash, G., Oprescu, F., Liebergreen, M., Turley, J., Bond, R., & Dayton, J. (2016). A written, reflective and dialogic strategy for assessment feedback that can enhance student/teacher relationships. *Assessment & Evaluation in Higher Education, 41*(1), 141–153. https://doi.org/10.1080/02602938.2014.986644.

Döring, N. (2006). Für Evaluation und gegen Evaluitis. Warum kann und wie sollte Lehrevaluation an deutschen Hochschulen verbessert werden. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer, & J. Wildt (Eds.), *Neues Handbuch Hochschullehre* (I 1.7.). duz Medienhaus.

Entner, C., Fleischmann, A., & Strasser, A. (2021). Hochschullehre im digitalen Wandel. Überlegungen zur didaktischen Gestaltung von Präsenz- und Onlinelehre. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer, & J. Wildt (Eds.), *Neues Handbuch Hochschullehre* (A 3.30.). duz Medienhaus.

Fleischmann. A. (2020). Emergency remote teaching. Pragmatische Ansätze zur Transformation von Präsenzlehre zur Onlinelehre. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer, & J. Wildt (Eds.), *Neues Handbuch Hochschullehre* (D 3.39.). duz Medienhaus.

Flodén, J. (2016). The impact of student feedback on teaching in higher education. *Assessment & Evaluation in Higher Education, 42*(7), 1054–1068. https://doi.org/10.1080/02602938.2016.1224997.

Frank, A., Fröhlich, M., & Lahm, S. (2011). Zwischenauswertung im Semester: Lehrveranstaltungen gemeinsam verändern. *Zeitschrift für Hochschulentwicklung*, *6*(3), 310–318. https://www.zfhe.at/index.php/zfhe/article/view/354.

Franz-Özdemir, M., Reimann, J., & Wessel K. (2019). Teaching Analysis Poll (TAP) – Konzept und Umsetzung einer aktuellen Methode an der Schnittstelle von Evaluation und Lehrentwicklung. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer, & J. Wildt (Eds.), *Neues Handbuch Hochschullehre* (I 1.17.). duz Medienhaus.

Gakhal, S., & Wilson, C. (2019). Is students' qualitative feedback changing, now it is online? *Assessment & Evaluation in Higher Education, 44*(3), 476–488. https://doi.org/10.1080/02602938.2018.1521371.

Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice, 20*(2), 150–169. https://doi.org/10.1080/0969594X.2012.749212.

Gollowitzer, M., & Jäger, R. (2009). *Evaluation kompakt*. Beltz.

Han, Y., & Xu, Y. (2020). The development of student feedback literacy: The influences of teacher feedback on peer feedback. *Assessment and Evaluation in Higher Education, 45*(5), 680–696. https://doi.org/10.1080/02602938.2019.1689545.

Harvey, L., & Stensaker, B. (2008). Quality culture: Understandings, boundaries and linkages. *European Journal of Education, 43*(4), 427–442. https://doi.org/10.1111/j.1465-3435.2008.00367.x.

Hattie, J. (2013). *Lernen sichtbar machen.* Schneider.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Sage Journals, 77*(1), 81–112. https://doi.org/10.3102/003465430298487.

Henderson, M., Phillips, M., Ryan, T., Boud, D., Dawson, P., Molloy, E., & Mahoney, P. (2019). Conditions that enable effective feedback. *Higher Education Research & Development, 38*(7), 1401–1416. https://doi.org/10.1080/07294360.2019.1657807.

Hoon, A., Oliver, E., Szpakowska, K., & Newton, P. (2015). Use of the 'stop, start, continue' method is associated with the production of constructive qualitative feedback by students in higher education. *Assessment & Evaluation in Higher Education, 40*(5), 755–767. https://doi.org/10.1080/02602938.2014.956282.

Hunt, N. (2012). Does mid-semester feedback make a fifference? *Journal of the Scholarship of Teaching and Learning*, *3*(2), 13–20.

Huxham, M., Laybourn, P., Cairncross, S., Gray, M., Brown, N., Goldfinch, J., & Earl, S. (2008). Collecting student feedback: A comparison of questionnaire and other methods. *Assessment and Evaluation in Higher Education, 33*(6), 675–686. https://doi.org/10.1080/02602930701773000.

Knödler, E. (2018). *Evaluation an Hochschulen. Entwicklung und Validierung eines verhaltensbasierten Messinventars zur studentischen Lehrveranstaltungs-evaluation.* Springer VS. https://doi.org/10.1007/978-3-658-25553-4.

Kopp, M. (2021). Covid-19 als Katalysator für die Digitalisierung der Hochschullehre. In J. Muckenhuber, M. Griesbacher, J. Hödl, & L. Zilian (Eds.), *Disruption der Arbeit? Zu den Folgen der Digitalisierung im Dienstleistungssektor.* Campus. [in print].

Kröber, E., & Thumser, K. (2008). Lehrveranstaltungen evaluieren. Ein Workshopkonzept zur nachhaltigen Evaluation der eigenen Lehre. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer, & J. Wildt (Eds.), *Neues Handbuch Hochschullehre* (I 1.8.). duz Medienhaus.

McClain, L., Gulbis, A., & Hays, D. (2018). Honesty on student evaluations of teaching: Effectiveness, purpose, and timing matter! *Assessment & Evaluation in Higher Education, 43*(3), 369–385. https://doi.org/10.1080/02602938.2017.1350828.

Mitry, D. J., & Smith, D. E. (2014). Student evaluations of faculty members: A call for analytical prudence. *Journal on Excellence in College Teaching, 25*(2), 56–67.

Mojescik, K. (2017). Möglichkeiten und Grenzen von Feedbackinstrumenten in Lehre und Studium. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer, & J. Wildt (Eds.), *Neues Handbuch Hochschullehre* (I 1.16.). duz Medienhaus.

Schumacher, F., Ademmer, T., Bülter, S., & Kneiphoff, A. (2021). *Hochschulen im Lockdown. Lehren aus dem Sommersemester 2020. Ergebnisse der Community Working Group Motivationsfaktoren für Dozierende zur Umsetzung digital unterstützter Lehre.* Hochschulforum Digitalisierung. https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_58_Hochschulen_im_Lockdown.pdf.

Scriven, M. (1972). Die Methodologie der Evaluation. In C. Wulf (Ed.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (Erziehung in Wissenschaft und Praxis Vol. 18, pp. 60–91). R. Piper & Co. Verlag.

Sozer, M. E., Zeybekoglu, Z., & Kaya, M. (2019). Using mid-semester course evaluation as a feedback tool for improving learning and teaching in higher education. *Assessment & Evaluation in Higher Education, 44*(7), 1003–1016. https://doi.org/10.1080/02602938.2018.1564810.

Stamm, M. (2003). *Evaluation und ihre Folgen für die Bildung. Eine unterschätzte pädagogische Herausforderung*. Internationale Hochschulschriften: Bd. 419. Waxmann.

Stockmann, R., & Meyer, W. (2014). *Evaluation. Eine Einführung.* Utb.

Svinicki, M. D., & McKeachie, W. J. (2014). *McKeachie's teaching tips. Strategies, research, and theory for college and university teachers* (14th ed.). Cengage Learning.

University of Graz. (2020a). *Graduation survey.* [unpublished document].

University of Graz. (2020b). *digiTales 11: Prof. Jürgen Pirker interviewed by Simone Adams*. [Video]. UniTube. https://unitube.uni-graz.at/play/fbe22d61-07b7-46b5-a7a6-87a0ef1b6e96.

University of Graz. (2020c). *digiTales 12: Prof. Markus Steppan interviewed by Simone Adams*. [Video]. UniTube. https://unitube.uni-graz.at/paella/ui/watch.html?id=011090be-6719-4bfc-a6a0-cda8aeb75c73.

University of Graz. (2020d). *Teacher surveys on online teaching in 2020*. [unpublished document].

University of Graz. (2021). *Focus group with lecturers within the project "Workplace Health Promotion" (Betriebliche Gesundheitsförderung).* [unpublished document].

Winstone N., Boud, D., Dawson, P., & Heron, M. (2021). From feedback-as-information to feedback-as-process: A linguistic analysis of the feedback literature. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2021.1902467.

Wolf, G., & Kothe, C. (2020). Digitalisierte Lehrveranstaltungsevaluation und Feedback – Perspektiven von Lehrenden. *Zeitschrift für Hochschulentwicklung, 15*(1), 175–194. https://doi.org/10.3217/zfhe-15-01/09.

Zhang, Z. (V.) (2021). Promoting student engagement with feedback: Insights from collaborative pedagogy and teacher feedback. *Assessment & Evaluation in Higher Education.* https://doi.org/10.1080/02602938.2021.1933900.

Alexandra Dorfer finished her studies in sociology and economics at the University of Graz where she is currently deputy manager of the Department of Academic Services with expertise in quality assurance and development in teaching. She is Arqus European University Steering Committee Member and teaches statistics at the University of Graz.

Gudrun Salmhofer After studying German philology and art history in Graz, Gudrun Salmhofer worked as a German language and literature lecturer in Poland, Slovakia, Ukraine and the Czech Republic. She took over the management of the Department of Academic Services at the University of Graz in 2005 and became managing director of the Competence Center for University Teaching in 2010. Her work focuses on quality assurance and development in teaching, educational design, study choice orientation and improving the transition from school to university.

Lisa Scheer combines sociological and educational perspectives in her current work at the Competence Center for University Teaching. Before that she was active in quality and diversity management. She is particularly concerned with issues related to diversity-sensitive teaching, social inequality, and reflexive learning and mainly teaches in the field of Gender Studies.

Digitalisation of Examination Formats in Higher Education Corona-Related Changes

Mana-Teresa Donner and Sandra Hummel

Abstract

A central element of any effective teaching and learning strategy is assessment. Closely related to evaluations of learning results is the constant challenge on the part of higher education teachers to develop assessment strategies aligned to the intended learning outcomes. The Corona-induced wave of digitation in higher education in spring 2020 opened up new forms of assessment and assessment approaches which are closely related to changing teaching and learning methods. The aim of this study is to explore how digital examination differs from those conducted in person. In order to make distinctions in the areas of examination formats, designs and strategies for examination implementation, guided interviews with 12 university lecturers at the University of Graz were conducted. Additionally, 14 examination forms from face-to-face and online teaching were analyzed. The results show that examination strategies as well as examination formats are strongly dependent on the size of the course and need to be adapted to the given framework conditions. Online examinations, comprehension and transfer questions as well as open-book examinations are increasingly used to counteract the reduced control in the online setting.

M.-T. Donner (✉) · S. Hummel
Center for Scalable Data Analytics and Artificial Intelligence, TUD Dresden University of Technology, Dresden, Germany
e-mail: mana-teresa.donner@tu-dresden.de

S. Hummel
e-mail: sandra.hummel@tu-dresden.de

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_5

Keywords

Digital assessment • Examination strategies • Online examinations

1 Introduction

With the onset of the Corona pandemic in spring 2020, the wave of digitalization in the education sector accelerated considerably (Hochschulforum Digitalisierung, 2015). The digital transformation of the didactic space altered learning and teaching arrangements and raised questions on how exams can be conducted in online environments (Hummel, 2020) and what changes are needed at the level of exam format and items (Pausits et al., 2021). In practice, the transition to digital examinations proved to be one of the most difficult aspects in this change process, as it was not only a matter of transferring traditional face-to-face examinations into an online version, but rather required the reconsideration of new examination formats considering the altered framework conditions. Despite considerable differences regarding educational objectives, types of courses or disciplines, the digital examination environments entailed the restructuring, redesigning, and reflecting of established examinations (Bedenlier et al., 2021). This reassessment simultaneously confronted teachers to review their evaluation strategies, and often led to significant changes in their assessment approaches (Adedoyin & Soykan, 2020; Rapanta et al., 2020).

Vollmer (2020) highlights the prevalence of written examinations as the primary method of assessment in the higher education sector. With a share of 85 %, this is the most frequently used form of examination in face-to-face teaching settings (Billerbeck et al., 2016). Whereas in lectures with a large number of students (>100 students; Ahn et al., 2011), questions predominantly refer to the taxonomy level "remembering" (40 %), in seminars or courses with a smaller group of students this is only 22 %, as here the focus lies on "understanding", "applying" and "analysing" (37 %) (Bloom et al., 1956; Brauns et al., 2015). In large group teaching, open exam formats are rather avoided which is due to economic implementation and efficiency reasons (Reinmann, 2014) such as the considerable time required for correcting open examination forms, as well as reduced objectivity and comparability (Billerbeck et al., 2016). Accordingly, the written examination is not only the most frequent form of testing (Brauns et al., 2015) and the simplest option (Billerbeck et al., 2016), but is also efficient in a sense of being economical to administer to large groups of students. A considerable disadvantage, however, is that written exams with closed questions are rather unsuitable for competences situated at the lowest taxonomy level

(Reinmann, 2014). However, resulting from the Bologna Process, which aimed to create greater standardization and mobility across European higher education, many universities have adopted electronic examinations. This shift towards electronic examinations represents a significant change from the traditional mode of assessment and has the potential to impact the way that students prepare for and engage with their studies. The evaluation process is fully or only partially automated (Reinmann, 2014), and open questions can be used in addition to closed questions. On the one hand, this should have the advantage of reducing costs and speeding up the correction process and, on the other hand, the overall quality is increased, for example, by eliminating manual evaluation processes (Bauer et al., 2008). Before the Corona pandemic, electronic examinations used to be conducted mainly under controlled conditions in large computer rooms with several examination invigilators.

During the COVID-19 pandemic, online examinations have become a widely used form of assessment, providing the opportunity for students to take tests remotely while maintaining social distancing and respecting remote learning requirements. Despite the benefits online assessments doubtlessly presented to ensure the continuation and completion of studies, students have had to face a wide variety of challenges. Elsalem et al. (2020) found that 32 % of the students surveyed felt more stress during online exams than during face-to-face exams. The type of examination questions can be a significant source of stress for students and may contribute to stress-related factors. For instance, closed Multiple-Choice questions may lead to feelings of anxiety as students have to select one “correct” answer from a set of seemingly similar choices. Instead, open text questions may cause stress as they require students to generate and organize their thoughts in a limited time frame. Therefore, the way examination questions are designed and presented plays a crucial role in determining the level of stress that students experience during exams. More than 75 % of students report a limitation in exam performance due to the fear of technical problems as well as the short duration of the exam. Especially in the online setting, students need sufficient information about the exam, for example transparency regarding exam duration, question format or grading modalities. The organization of examination rounds helps students to become familiar with the exam situation and is therefore considered as stress-reducing. Clarity about the procedure in case of technical difficulties is another decisive factor to reduce stress which tends to lead to poorer achievements (Brodese & Lorenz, 2020). It is therefore advisable to create an examination environment in which technical problems can be solved quickly and to provide synchronous communication channels during the exam, such as video conference rooms or the provision of email addresses, telephone numbers or availability via chat rooms. Also structured communication throughout the exam period can foster successful exam delivery in the online setting (Phillips & Phillips, 2020). Whereas in face-to-face examinations students can spontaneously ask for support, the online room is characterized by isolation. Structured communication with different communication channels allows students to ask for support during the exam. Furthermore, notifications about the successful upload of the examination forms turned out to be advisable to avoid insecurity about technical failures during the submission phase (Lipp & Dreisiebner, 2020). These unsettling factors during the lockdown confronted HE teachers with new challenges to ensure that they can accurately provide and assess their students' knowledge and skills. The aim of this article is to investigate how digital exams were finally realized and differed from those held in face-to-face settings. For better comparability, exam forms used before and after the onset of the Corona pandemic which were delivered in the same courses were analyzed. Particularly exam formats, question designs, preparation, time scope and scoring were of central interest. Furthermore, experience reports from teachers were interpreted to derive implications for the implementation of effective digital examination scenarios.

2 Research Design

The aim of this article is to investigate the extent to which the digital examinations differ from face-to-face assessments. The investigation of the research questions is a two-stage study design. In a first step, guided interviews were conducted with teachers from all faculties of the University of Graz. This should allow the identification of different forms of digital examinations, to incorporate experience reports from teachers, and to derive recommendations and implications for the design of efficient examination scenarios in the online area. In a second step, examination forms from the time before (until winter semester 2019) and after (from summer semester 2020) the Corona-related changeover were examined by means of document analyses and compared with each other to examine, on the one hand, the implementation of digital examinations in terms of examination formats and examination design from those in face-to-face courses and, on the other hand, to identify strategies regarding preparation, communication and assessment.

For the analyses, guided interviews were conducted with teachers from the University faculty. Based on the findings, a category system was developed with seven main categories: (1) examination format, (2) exam questions, (3) examination strategies with three subcategories: preparation, examination environment, and cheating, (4) communication with students, and (5) Achievements in online examinations.

Categories were formed through a mixture of a theory-guided-deductive and inductive procedure (Mayring & Fenzl, 2019). The theory-guided deductive categorisation was oriented towards the research questions and the thematic blocks of the interview guide. Three essential criteria were defined for the selection of the sample (Przyborski & Wohlrab-Sahr, 2014):

- The teacher had at least one course that was conducted as a **lecture** and concluded with a **written and/or oral lecture examination**.
- Furthermore, it was important that the teacher held the lecture **before and after the transfer to the online room** due to the Corona, since this is the only condition under which it is possible to compare the online and classroom examination.
- In addition to the type of course and the conversion to digital examinations, the **respective faculty of the lecturer** was also an important selection criterion, as a balanced ratio across all faculties was aimed for in the sample composition.

The sample for the guided interviews consists of a total of 12 teachers ($n = 12$; eight women, four men) at the University of Graz, with a length of employment of 2-36 years ($\overline{x} = 20.1$; $SD = 10.3$). Three interviewees belong to the age group 40-49 years, six interviewees to the group 50-59 years and three interviewees are over the age of 59 years. Eight of the interviewed teachers belong to the NAWI or URBI Faculty, only two to Catholic Theology and one each to the GEWI and SOWI Faculty.

In a second step, document analyses were conducted to examine the changeover from face-to-face to online examinations. The procurement of the examination questionnaires before and after the Corona-induced switch to online examinations took place simultaneously with the recruitment for the interviews via e-mail dispatch. Because both the past and current exam forms involved sensitive content, a written confidentiality agreement was reached in written form with the respective teachers to address concerns about misuse of data.

The sampling of the document analysis consists of a total of $n = 14$ examination forms, seven of which were conducted as face-to-face examinations and seven as online examinations. One examination form each was from 2016 and 2017, three examinations from 2019, six from 2020 and three examinations from 2021. The examination forms originate from the University of Graz from seven courses in lecture format. Figure 6 shows that the majority of the examinations come from the NAWI Faculty, but also two examination forms each from the Faculty of Catholic Theology, SOWI and the URBI Faculty.

For the quantitative evaluation of the examination forms, an analysis grid (see Table 1) was created based on the categories from the interviews (strategies of exam realization, preparation, creation of examination environment, exam design, exam formats, tasks, assessment criteria and feedback to students). For better clarity, both the course (marked by numbers: 1, 2, … 7) and the status of the exam before (marked by a) or after the conversion (marked by b) to online teaching were considered.

The grid includes the examination format, e.g., Multiple-Choice examination (MC), open- or closed book examination, and furthermore the format with which the examination was conducted (e.g., Moodle, Perception, Paper–Pencil). Subsequently, the table provides a breakdown of the number of examination questions, distinguishing between open-ended questions, closed-ended questions (e.g., Single-Choice or Multiple-Choice questions), and semi-open questions (e.g., assignment tasks, cloze texts). These formats are further subdivided into

Table 1 Evaluation grid for the document analysis of the examination papers

Exam	**Faculty**	**Format**	**Tool**
1a 2020	REWI	MC	Paper–Pencil
1b 2021	REWI	Open-book	Perception
Number of questions (total)	**Number of closed questions**	**Number of open questions**	**Number of semi-open questions**
30	30	0	0
15	2	10	3
Number of knowledge questions	**Number of words in question text**	**Number of comprehension questions**	**Number of words in comprehension question text**
28	13.7	2	25.6
2	12.5	13	38.4
Number of images	**Sequential presentation**	**Exam time**	
0	No	60 Min	
2	Yes	30 Min	

Note. The information in the table is entirely fictitious knowledge questions, where primarily the taxonomy levels remembering and understanding are in the foreground, and comprehension questions, where students apply and link their knowledge to real-world examples. For the knowledge and comprehension questions, the scope is recorded in terms of the number of words in the task. In addition, the number of illustrations is indicated, whether the questions were presented sequentially and how long the exam duration is. Based on these criteria the face-to-face exams are compared with the respective ones conducted in the online setting.

3 Results

3.1 Examination Format

While in face-to-face teaching the classic written examination in the form of Multiple-Choice questions and other closed question formats (e.g., cloze texts) represented the most frequently used examination form (e.g., TN3, pos. 3-8), this format was replaced in the online setting by open-book examinations.

> "Yes, before that the written exam in the course was face-to-face (...) and then in the Corona semester they switched to open-book exams online" (e.g., TN7, pos. 3-6).

Open-book exams with open-ended comprehension and application questions represent an authentic form of examination that can be particularly beneficial for those entering the workforce. Mass teaching with several hundred students in both online and face-to-face teaching makes such innovative assessment models difficult, as open-ended question formats that focus on transfer and comprehension of learning significantly increase the time required for assessment (e.g., TN4, pos. 122-126). In courses with a high number of participants, the focus is on reproduction (e.g., TN1, pos. 43-48) while smaller groups tend to focus on analysis (TN 11, pos. 54-55).

Oral online exams can easily be transferred to the online room (Goodman et al., 2021) and can be seen as equivalent to the oral face-to-face exam (TN6, pos. 9-11). In most cases, no adaptations of the examination tasks are necessary (e.g., TN2, pos. 3) and the dialogic examination discussion takes place in the form of a synchronous interaction (Persike et al., 2021) via a video conferencing system (e.g., UniMeet). Teachers also report that post exam students tell them that this format caused them less stress than oral exams usually do. There is correspondingly the advantage that the examination performance can be evaluated directly after the examination. It was evident across all forms of online exams that the move to the online space resulted in significantly higher exam participation.

> "Well, we always had a lot of students who registered, sometimes much more than in the face-to-face exams (...) sometimes we only had five or six students who took the exam, that was not the case in the Corona time, it was always a double-digit number of students" (TN7, pos. 140-146).

The analyses showed that closed-book exams in face-to-face teaching are mainly conducted as paper-pencil exams in the classroom. Open-book exams in online teaching, however, are mainly conducted via the LMS, in this study Moodle, or use digital exam software. Some teachers were limited in their testing activities using the Perception testing software, as not all item formats could be transferred to the online space.

Additionally Multiple-Choice exams ensure high standardization, assessment objectivity (e.g., TN5, pos. 26-29), and coverage of a wide range of topics. A key change in the implementation of Multiple-Choice exams in the online setting was the switch from the paper-pencil format to the use of the Perception exam software.

> "And we just copied the exam questions into Perception and we didn't change anything in the exam questions, we just copied them into Perception" (TN1, pos. 10-12).

> "And after the start, after the start of Corona, we then switched to online testing with the, what's the name of the program? Perception (…) However, I must say that I did not change the format of the questions significantly" (TN3, pos. 5-12).

The teachers mainly expressed concerns about the technical implementation as well as the functioning of the tools during the test, but this quickly became a habit after initial uncertainty.

> "So I was a little worried at the beginning, whether technically everything works. What do I have to do etc." (TN5, pos. 253-255).

> "I was particularly concerned about the technology. Whether everything would really work as expected. (...) I think we're all used to it by now (…) and everything works well that way. So that has become a habit actually" (TN2, pos. 92-98).

Different examination instruments are found in online examinations compared to examinations in the lecture hall. While paper-pencil exams are predominantly used in the lecture hall, online exams use the Moodle learning management system and Perception exam software. Although Perception allows the creation of over 18 different question types (University of Graz, 2022), the implementation of open-ended questions in the practice of online exams appears to be laborious and user-unfriendly. This could also be a reason why one teacher returned to paper-pencil exams in face-to-face classes.

3.2 Exam Questions

In addition to a well-functioning technology, a good and, above all, secure examination environment in the online setting can be created by questions that encourage critical thinking as well as the transfer of what has been learned. The shift to the online setting exposed faculty to new exam formats, but also encouraged many to return to face-to-face exams with the pre-Corona exam format after the online exam phase.

> "I probably would never have tried this without the pandemic and actually find it to be a more honest format of testing" (TN12, pos. 275-277).

> "So I relearned that, and it was actually deeply satisfying to me to see that the good old short essay [exam format in the face-to-face setting] is the best format" (TN11, pos. 201-203).

Many instructors assumed that the online examination was only a short phase and saw no need for a complete revision and adaptation of the examination format to the altered assessment conditions. Since there were no differences in grades after the first online exam was administered, the exam items or even the entire exam format (e.g., from MC exams to open-book exams) had to be adapted to meet the online requirements. Particularly, open questions resulted in a wider spread of grades. The type of exam questions in both environments is strongly oriented towards the number of students, e.g. the use of Multiple-Choice exams in mass teaching.

> "(…) it always depends on the number of participants in the course. I have this course for almost 20 years or so and it has always varied greatly in terms of size, i.e. the number of participants. At the beginning it was a small (...) group (...) and when it became a free elective, more and more students came. And with the moment also the examination mode changed" (TN5, pos. 26-34).

Closed-book exams with open-ended questions were used more frequently in face-to-face teaching prior to the Corona pandemic. Scoring open-ended comprehension questions is more subjective and more time-consuming for instructors than scoring open-ended knowledge questions because there is no simple ‘right’ or ‘wrong’ with this type of question. For many teachers, the difficulty of questions on online exams is comparable to those on face-to-face exams (25 %), while others find online exams (41 %) or face-to-face exams (34 %) more difficult. Instructors who rate their exam as more difficult in the online setting attribute this to the use of transfer questions instead of knowledge questions (e.g., TN4, pos. 184-197). When the use of learning materials during the exam was mentioned, the online exams were described as easier:

> “I have more the feeling that the online version is much easier, even with a shorter time. Most then just write something down, so there are very few who omit questions. While they were in the lecture hall and did not have the documents available, the proportion of those who simply knew nothing, or little was already significantly higher. I don’t think that the students prepare better online, but simply say yes, I’ll print out the documents and then I can look them up” (TN8, pos. 117-127).

In exam design, the use of question pools (e.g., TN1, pos. 5-6; TN8, pos. 103-106), randomization with respect to the sequences of exam questions, and mixing of answer choices are found in both exam settings.

> Face-to-face exams, e.g.: “And I had held this lecture for many years and had already built up such an examination question paper. And we had already had such an examination system with Multiple-Choice tasks that also evaluated the examinations and so on (...)” (TN1, pos. 5-8).
>
> Online exams, e.g.: “(...) I have a pool of 50 questions that I then select, so I really have the impression that the way the questions are answered, that it works quite well” (TN8, pos. 104-106).

These measures are designed to ensure that students do not answer the same exam questions at the same time. In case of online electronic exams, this is compounded by the fixed order in which students cannot click forward or backward on the questions. For face-to-face exams they usually have the option to switch back and forth between questions. Since the creation of question pools is very time-consuming, exam forms that have already gone through repeated assessment and revision processes in face-to-face classes have been adopted for online classes. Even though the process of developing question pools seems to be comparable in both settings, teachers in the online setting have an additional effort in creating new exam questions because all questions are disseminated faster due to online availability (e.g., screenshots).

> "So (...) with the online examination, the exam questions will of course go out at some point, which will normally never leave the lecture hall in face-to-face exams. So from that point of view, it is already more effort to think about how could I ask them differently this time, how could I modify the answer alternatives, sometimes it just becomes more answer alternatives, but the question is definitely kind of circulating online already" (TN5, pos. 299-304).

The document analysis also showed that the average number of words in open questions is significantly higher than in closed questions. In order not to increase the effort in assessing open questions, the scope of students' answers should be limited. Therefore, it is necessary on the part of teachers to clearly separate the task description from other irrelevant examination content. In the online setting, an increased use of graphics and illustrations can also be observed. Statistic visualizations are often included to assess interpretative, analytic and application-oriented skills.

3.3 Examination Strategies

3.3.1 Preparation

Regarding preparation, it was found that the time for students' questions regarding the forthcoming exam as well as preparatory information about exam-relevant content remained unchanged in online teaching compared to face-to-face teaching. Four teachers conducted test runs with students or colleagues, especially at the beginning of the transition, to familiarize themselves with the new format or testing system. All of the interviewed lecturers felt responsible to inform the students in good time and, above all, sufficiently about changed examination modalities, processes and procedures in the event of technical faults, as well as to make the examination link available. In addition, some instructors directly addressed the issue of plagiarism in exams and appealed to students' ethical awareness to ensure the integrity of the exam.

The preparation of online examinations depends significantly on the form of examination: While for oral online exams only the link was sent, for the alternative examination performance in the form of written papers (e.g. essays) the topics had to be determined and defined first (TN6, pos. 42-46). For online exams (e.g. open-book exams), it is possible for instructors to make changes to the exam questions via the Moodle learning management system until shortly before the exam. This is not possible with face-to-face exams in paper–pencil format due to the time-consuming preparation in the form of copying and placing the exam sheets in the lecture hall. In addition, there is no need to print out and distribute multi-page examination sheets for the examination, which not only has an ecological advantage, but also saves personnel and time resources (TN3, pos. 109-129).

3.3.2 Examination Environment

Traditional face-to-face examination environments can be described as quiet and controlled environments in which students are not allowed to use any aids. In the online setting, a good and secure examination environment can be ensured by the type of questioning, e.g. by switching the examination tasks to comprehension and transfer questions. In the case of oral online examinations, there is also the fact that both students and teachers have to turn on their cameras so that a dialogue can take place with facial expressions and gestures, identification of the students is possible and cheating can be prevented.

In addition, the examination environment in the online setting is highly dependent on functioning technologies as well as on the technical equipment of the students. As an example, one teacher (TN4, pos. 217-222) mentions the failure of the examination software and the associated postponement of the examination date. It is therefore advisable to create an examination environment in which technical problems can be solved quickly, e.g. through support from technicians during the examination.

3.3.3 Cheating

According to instructors, shifting exams to the online realm entails less control over cheating attempts.

> “So the temptation to cheat is greater because there is no control, we have also seen that, I mean I didn’t have to repeat an exam, but I know from colleagues that they had to repeat, that whole exams had to be declared invalid” (TN11, pos. 159-162).

Measures against attempted cheating range from question format to limiting processing time to increasing the number of exam questions and the use of proctoring software. The latter tends to meet resistance from faculty due to privacy issues and the logistical challenges of implementing it in mass courses (e.g., TN1, pos. 97-98; TN5, pos. 171-176).

Students are not only more likely to cheat in online exams (Janke et al., 2021), cheating is also easier for them (Aisyah et al., 2018). Teachers mention the lack of control in the case of attempted cheating as a significant disadvantage of online exams. In the case of examinations in the lecture hall, the control of possible attempts at cheating by students is carried out by the traditional examination supervisors. In the online room, there are various measures in place in this regard to curb attempts at cheating. Regarding the format of the examination tasks in the online setting, according to the instructors, only open questions would prevent students from attempting to cheat, but this could not be implemented in all courses, so other prevention measures were taken here.

Other preventive measures to avoid cheating specifically applicable to online exams target the scope, in terms of length and number of items, as well as the duration of the exam. Online exams tend to contain more questions and often prescribe a maximum number of sentences or characters. Both in online and paper–pencil exams in the lecture hall, text boxes are included to symbolize the required scope of the answer.

> "What I do, however, as a point of reference, is that I ask the questions and then put text boxes in the open, so I have prepared a word document, there are the questions, the images and then a text box, which is also colored. You have a visual suggestion, I would say, which is roughly oriented to one page per question" (TN12, pos. 113-118).

In face-to-face exams, instructors usually see if students need time beyond the scheduled exam time (TN4, pos. 209-214).

> "In face-to-face exams we were used to, if we see the exam is scheduled for an hour and everyone is still writing, then you just add five minutes. And everything is fine and no fuss. Don't worry, we'll add a few minutes until everyone has handed in" (TN4, pos. 209-214).

In the online setting, the optimal time is filtered out through pilot tests and compliance is very strictly observed: "(…) and every minute that they write about it is three percent of the points deduction or something (…) so that they are already busy, that they can manage it well, but that they also don't have time to exchange anything" (TN4, pos. 133-149). It must be considered that the increased use of transfer questions in online examinations also requires more time to answer than in closed question formats.

Teachers emphasize that exam questions that serve to transfer learning and stimulate critical thinking provide little opportunity for cheating attempts:

> "The transfer questions leave little opportunity for cheating. The answers tell me how well and comprehensibly the material was conveyed, what remained unclear and will require better explanations in the future, with examples, queries, detailed explanations and offers for discussion" (TN10, pos. 53-57).

> "(...) due to the lack of social interaction, the temptation to cheat is greater. But that also means, and maybe that's good for us (laughs) as examiners, that we have to think about examination formats where cheating doesn't make sense" (TN11, pos. 163-166).

An explicit reference to the academic honor code in combination with a warning produces the greatest reduction in attempted cheating (Bing et al., 2012; e.g. TN4).

3.4 Communication with Students

Structured communication during the exam is essential for successful completion of the exam in the online setting. While students can ask questions about the exam during face-to-face exams, the lack of spontaneous communication needs to be replaced by other forms of communication: In case of problems or important questions during the exam, students had the option of contacting the instructors by email, phone, or in a separate UniMeet room.

> "Everything worked technically without any problems (...) they could come back to this BigBlue Button room, where we waited in case of problems. And were able to ask if their exam had arrived?" (TN4, pos. 217-222).

Teachers were thus able to respond quickly to technical as well as content-related questions. In the solution of passing on the telephone number (TN8, pos. 172-175), clarification of questions only takes place between the lecturer and a student, although it could affect the entire group. This poses a difficulty in terms of fairness to the other students, since questions regarding content are usually answered for the entire group in the lecture hall.

3.5 Achievements in Online Examinations

Achievement in online exams is comparable to performance in face-to-face classes. Instructors who report lower grade dispersion (e.g., TN7, pos. 73-76) either use mostly closed-ended questions in their exams or there was no adaptation of the exam sheets. Most instructors surveyed do not remark grade variances because of the switch to online exams. One argument in favor of this is that although cheating seems to be easier in the online space, many students often do not consider how much exam time it costs to search for information. In the online setting, instructors pay more attention to independent wording to ensure that exam performance is not a collaborative outcome. In large courses, free response represents a high level of effort. As mentioned above, the effort involved in assessing open-ended questions can be minimized by specifying the word count of the answer. Another possibility is to create an evaluation grid that can be used by the students already when formulating the answers as well as by the teacher when evaluating them. An evaluation grid is a tool used to assess and evaluate the performance of individuals or groups against predefined criteria. It is commonly used in educational settings, such as in the evaluation of student work, and in professional settings, such as in performance appraisals or project evaluations. Such a grid typically consists of a table with a list of criteria along the top row and the names of the individuals or groups being evaluated along the left column. Each cell in the table represents the assessment of a particular criterion for a particular individual or group. The evaluator rates the performance of each individual or group against each criterion, usually using a numerical scale or a set of performance descriptors.

> “(...) developed an evaluation grid (...) here there are different evaluation criteria and a scale that also describes, so to speak, what does it mean when I sit at the very bottom of this scale, and what does it actually mean to get to the top level? And that helps me tremendously in the assessment, in the quick, quick assessment of the open-ended questions” (TN12, pos. 122-128).

The use of an evaluation grid allows for a systematic and objective assessment of performance, as it provides a structured framework for evaluating performance against predetermined criteria. It also facilitates the comparison of performance between individuals or groups, making it easier to identify strengths and areas for improvement. An evaluation grid can be used in a variety of contexts, such as in grading student assignments, evaluating employee performance, or assessing the success of a project. The evaluation grid can also be combined with methods of automated correction. Automated correction and grading of online assessments can have several effects, both positive and negative. The instructors emphasize that automated corrections make the lack of variation in individual items immediately visible. This allows them to adjust or delete individual exam items right away. Due to increased efficiency, automated correction can save teachers a significant amount of time compared to manually grading assessments manual scanning of the exam sheets is no longer necessary and errors due to manual evaluation steps can thus be avoided. The time savings also refer to the elimination of manual grade entry and illegible handwriting by uploading

PDF files. In addition, students can access their grades more quickly compared to face-to-face instruction. These time savings can allow teachers to focus on other important aspects of teaching and learning. Another relevance factor is consistency since automated grading can reduce deviations in applying assessment criteria and standards. Therefore, automated grading reduces the potential for subjective evaluation biases. Furthermore, automated grading can provide immediate feedback to students, allowing them to see their scores and identify areas where they need to improve. The feedback students receive however may not provide detailed feedback to students beyond a score, which can limit students' ability to understand their mistakes and improve their performance. Automated grading systems can hardly be used to accurately evaluate complex assessments, such as essays or creative projects, that require subjective evaluation. Another disadvantage are potential technical problems that can delay the grading process or produce inaccurate results.

4 Summary

The Corona pandemic posed major challenges for higher education teachers, but some opportunities for higher education have also emerged during the 'digital' semesters. This article discusses the shift to digital assessments, using a two-stage study design. The first stage involved guided interviews with 12 teachers from all faculties at the University of Graz, while the second stage involved examining examination forms from before and after the switch to online examinations due to the COVID-19 pandemic. A total of 14 examination forms were analyzed, seven of which were conducted face-to-face and seven online. The analysis was conducted using a category system that included exam format, exam questions, exam realization strategies, communication with students, and assessment of online examinations. The study found that online exams differed in exam formats, examination design, strategies of exam realization, and communication with students.

A primary goal of moving exams online was to avoid study delays and to ensure successful completion of courses (Berger, 2020). The examination environment in the online setting not only brings location independence, but also partial time savings as well as savings of ecological resources. Even though examinations in the online setting could be conducted after initial difficulties, most of the interviewees see online examinations as an additional facet, but not as a replacement for face-to-face examinations. Online exams offer the advantage of last-minute changes to the exam questions, ecological benefits and savings in personnel and time resources. In the context of preparing, conducting, and assessing online examinations, teachers prefer concrete advice, instructions, and suggestions for implementation from direct contact persons who have dealt intensively with this topic. This saves teachers the time-consuming task of searching for concrete facts in a vast number of offers. Further training seminars regarding the design of examinations in the learning management system Moodle should be adapted to the respective level of knowledge and experience of the teachers and thus also allow for more complex questions regarding one facet of online examinations. The informal exchange of experiences between colleagues ensures that the topic is viewed from different perspectives and best practice examples can be beneficial for one's own teaching. An important aspect that should also be taken into account is the urgent appeal of a teachers to provide a user-friendly examination program for conducting electronic examinations that corresponds to the latest research findings in artificial intelligence. This should also enable teachers to fully automate the evaluation of semi-open questions (e.g. cloze texts), even if words have been misspelled or similar words have been used.

In face-to-face teaching, Multiple-Choice and closed question formats are common, while open-book and oral examinations are used in online teaching. Multiple-Choice exams are preferred for their standardization and assessment objectivity. The teachers expressed insecurity about the technical implementation of online exams, but they got used to it over time. Cheating is also a major concern for instructors, and online exams offer fewer controls than traditional exams. Measures against cheating include changing question formats, limiting processing time, and using proctoring software. Furthermore, the online examination environment is dependent on functioning technologies and the technical equipment of the students. Therefore, it is advisable to create a supportive examination environment that can solve technical problems quickly, for example, with support from technicians during the examination.

Interviewees have also emphasized the importance of critical thinking and knowledge transfer in online examinations. The shift to online teaching has exposed faculty to new exam formats, but many have reverted to pre-Corona exam formats in face-to-face teaching. For further research the integration of formative approaches in online assessment as well as effective combinations of formative and summative assessments in various disciplines and course types represent interesting research perspectives.

References

Adedoyin, O. B., & Soykan, E. (2020). Covid-19 pandemic and online learning: The challenges and opportunities. *Interactive Learning Environments, 2*(1), 1–18. https://doi.org/10.1080/10494820.2020.1813180.

Ahn, H., Höfer, Y., & Kunz, L. (2011). Die Ermittlung von Lehrkonzeptionsprofilen zur Unterstützung von Berufungsverfahren. *Zeitschrift für Hochschulentwicklung, 6*(3), 135–152.

Aisyah, S., Bandung, Y., & Subekti, L. B. (2018). Development of continuous authentication system on android-based online exam application. In *2018 International Conference on Information Technology Systems and Innovation (ICITSI).* https://doi.org/10.1109/ICITSI.2018.8695954.

Bauer, Y., Degenhardt, L., Gilch, H., Kleinmann, B., & Wannemacher, K. (2008). Online exams as part of the IT-supported examination process chain. *ResearchGate.* https://www.researchgate.net/publication/228744745_Online_Exams_as_Part_of_the_IT-supported_Examination_Process_Chain/stats.

Bedenlier, S., Bandtel, M., Boom, K.-D., Gerl, S., Halbherr, T., Hebel, A.-L., Jeremias, X., Hehr, H., Mecklenburg, L., Mersch, A., Molter, K., Paffenholz, A., Reinmann, G., Riebe, K., & van Treeck, T. (2021). Prüfungen aus Perspektive der Prüfungsdidaktik. In M. Bandtel, M. Baume, E. Brinkmann, S. Bedenlier, J. Budde, B. Eugster, A. Ghoneim, T. Halbherr, M. Persike, F. Rampelt, G. Reinmann, Z. Sari & A. Schulz (Eds.), *Digitale Prüfungen in der Hochschule. Whitepaper einer Community Working Group aus Deutschland, Österreich und der Schweiz* (pp. 30–42). Hochschulforum Digitalisierung.

Berger, E. (2020). Digitale mündliche und schriftliche Prüfungen – von (fast) 0 auf 100%. In E. Berger (Ed.), *Fnma Magazin 02/2020 – E-Assessment und E-Examinations* (pp. 2–3). Verein Forum Neue Medien in der Lehre Austria.

Billerbeck, K., Barnat, M., & Knutzen, S. (2016). Kompetenzorientierung auf dem Prüfstand. Erprobung von Indikatoren für den Kulturwandel. *Die Hochschullehre – Interdisziplinäre Zeitschrift für Studium und Lehre, 2,* 1–16.

Bing, M. N., Davison, H.-K., Vitell, S. J., Ammeter, A. P., Garner, B. L., & Novicevic, M. M. (2012). An experimental investigation of an interactive model of academic cheating among business school students. *Academy of Management Learning & Education, 11*(1), 28–48.https://doi.org/10.5465/amle.2010.0057.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. The classification of educational goals. Handbook I: The cognitive domain.* David McKAy.

Brauns, J., Göymen-Steck, T., & Horn K.-P. (2015). *Lernziele, Veranstaltungs- und Prüfungsformen in erziehungswissenschaftlichen Bachelorstudiengängen. Eine vergleichende Analyse von Studienprogrammen an acht Universitäten.* IfE Publikationen. https://doi.org/10.3249/webdoc-3963.

Brodersen, L., & Lorenz, R. (2020). Perceived stress, physiological stress reactivity, and exit exam performance in a prelicensure bachelor of science nursing program. *International Journal of Nursing Education Scholarship, 17*(1). https://doi.org/10.1515/ijnes-2019-0121.

Elsalem, L., Al-Azzam, N., Jum'ah, A. A., Obeidat, N., Sindiani, A. M., & Kheirallah, K. A. (2020). Stress and behavioral changes with remote E-exams during the Covid-19 pandemic: A cross-sectional study among undergraduates of medical sciences. *Annals of Medicine and Surgery, 60,* 271–279.

Goodman, J. F., Saini, P., Straughan, A. J., Badger, C. D., Thakkar, P., & Zapanta, P. E. (2021). The virtual mock oral examination: A multi-institutional study of resident and faculty receptiveness. *OTO Open, 5*(1), 1–4. https://doi.org/10.1177/2473974x21997392.

Hochschulforum Digitalisierung. (2015). *Diskussionspapier – 20 Thesen zur Digitalisierung der Hochschulbildung. Arbeitspapier Nr. 14.* Hochschulforum Digitalisierung.

Hummel, S. (Ed.). (2020). *Grundlagen der Hochschullehre. Teaching in higher education.* Springer VS.

Janke, S., Rudert, S. C., Petersen, Ä., Fritz, T. M., & Daumiller, M. (2021). Cheating in the wake of COVID-19: How dangerous is ad-hoc online testing for academic integrity? *Computers und Education, 2,* 1–9. https://doi.org/10.1016/j.caeo.2021.100055.

Lipp, S., & Dreisiebner, G. (2020). Constructive Alignment im Kontext elektronischer Prüfungen. In E. Berger (Ed.), *Fnma Magazin 02/2020 – E-Assessment und E-Examinations* (pp. 23–25). Verein Forum Neue Medien in der Lehre Austria.

Mayring, P., & Fenzl, T. (2019). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 633–648). Springer Verlag.

Pausits, A., Oppl, S., Schön, S., Fellner, M., Campbell, D. F. J., & Dobiasch, M. (2021). *Distance Learning an österreichischen Universitäten und Hochschulen im Sommersemester 2020 und Wintersemester 2020/21.* Bundesministerium für Bildung, Wissenschaft und Forschung.

Persike, M., Rampelt, F., Budde, J., & Fath-Keiser, S. (2021). Mündliche Online-Prüfungen. In M. Bandtel, M. Baume, E. Brinkmann, S. Bedenlier, J. Budde, B. Eugster, A. Ghoneim, T. Halbherr, M. Persike, F. Rampelt, G. Reinmann, Z. Sari & A. Schulz (Eds.), *Digitale Prüfungen in der Hochschule. Whitepaper einer Community Working Group aus Deutschland, Österreich und der Schweiz* (pp. 129–134). Hochschulforum Digitalisierung.

Phillips, B., & Phillips, M. (2020). eAssessment und eExaminations im Kontext der Wirtschaftspädagogik. In E. Berger (Ed.), *fnma Magazin 02/2020 – E-Assessment und E-Examinations* (pp. 53–55). Verein Forum Neue Medien in der Lehre Austria.

Przyborski, A., & Wohlrab-Sahr, M. (2014). *Qualitative Sozialforschung. Ein Arbeitsbuch* (4th ed.). Oldenbourg Verlag. https://doi.org/10.1524/9783486719550.

Rapanta, C., Botturi, L., Goodyear, P., Guárdia, L., & Koole, M. (2020). Online university teaching during and after the Covid-19 crisis: Refocusing teacher presence and learning activity. *Postdigital Science and Education, 2,* 923–945. https://doi.org/10.1007/s42438-020-00155-y.

Reinmann, G. (2014). Kompetenzorientierung und Prüfungspraxis an Universitäten: Ziele heute und früher, Problemanalyse und ein unzeitgemäßer Vorschlag. *Gabi-Reinmann.* https://gabi-reinmann.de/wp-content/uploads/2014/10/Artikel_Berlin_Okt_14.pdf.

Universität Graz. (2022). Fragentypen. *Universität Graz.* https://epruefungen.uni-graz.at/de/tool/fragentypen/.

Vollmer, A. (2020). Wintersemester 2020/21. *Humbold-Universität zu Berlin.* https://www.cms.hu-berlin.de/de/dl/e-assessment_old/pruefungskultur/zahlen/ws2020_21.

Mana-Teresa Donner finished her master studies in *Psychology* and *Adult and Continuing Education* at University of Graz. Her research focuses on learning analytics (project: Learning Analytics—Students in Focus), higher education teaching and on self-regulated learning. She is a Ph.D. candidate (Educational Science) and investigates the learning psychology perspective of AI-based mentoring in the Research Group *Situating AI-based Mentoring* at ScaDS.AI (TU Dresden).

Sandra Hummel is an educational scientist and research group leader at the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) at Dresden Technical University. Her research focuses on human-centered AI-based teaching and learning. She is the coordinator of several EU projects concerned with designing AI tools and technologies that are responsive to the unique needs, goals, and abilities of heterogeneous learners.

Developing Questions for Digital Assessments – Approaches and Reflections from a Didactical Point of View

Julia Hense

Abstract

Digital exams were completely uncharted territory for many Higher Education institutions and university lecturers before the corona pandemic. Even after two semesters of online-learning, many university lecturers and education experts still debate how digital assessments can be organised technically and designed according to accepted didactical and administrative standards. Digital assessments do not only have to follow the requirements of the examination regulations. The technical requirements and possibilities of the university also play a role. And last but not least, from a didactic perspective, the question arises, which exam format should be chosen and is compatible with the imparted learning content and according learning objectives and, above all, how the precise exam questions should be designed. The paper presents various approaches to planning and designing exams and exam questions in digital space and formulates initial principles for designing exam questions in digital surroundings.

J. Hense (✉)
Society for Media and Competence Research mbH, MMB Institute, Essen, Germany
e-mail: hense@mmb-institut.de

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_6

Keywords

Constructive alignment • Digital assessments • Designing exam questions • Learning principles

1 Introduction

With the beginning of the corona pandemic, the topic of digital exams at universities and higher education institutions became relevant almost overnight. A topic that had previously been a side issue suddenly became the focus of the debates about the digitization of teaching and learning and, above all, the question: How can we digitize examination formats at universities and higher education institutions in an intelligent and helpful way? What aspects have to be considered?

Many classic ways of designing assessments in higher education quickly reach their limits in the digital space. Depending on the exam format, the question arises how we can verify that the person to be examined is actually taking the exam and not someone else. How can we ensure that only the permitted resources are being used? And how can we prevent entire answers to exam questions from being copied online (García-Peñalvo et al., 2020)?

Since then, these questions have been discussed both in teaching practice and from a theoretical point of view. Still, they have not yet been answered and reflected sufficiently. In particular, didactic issues in planning digital exams are often not addressed consistently. Therefore, this article aims at offering an approach for planning digital exams from a didactic point of view and to present it for discussion.

In order to be able to do this in a well-founded manner, the first step is to show the range of possible digital examination formats, to present methods and approaches and to spot at advantages and disadvantages. Subsequently, this paper will show ways how different digital assessment approaches can be well embedded in courses at university and in higher education institutions. The paper closes with a discussion of the current challenges and open questions for further discourse and empirical research.

2 Planning Digital Exams: Determining Factors

Like traditional exams, digital exams should be planned considering the determining factors and their interdependencies in order to meet the requirements of universities as well as the aims of the lecturers and the needs of the students. Three main questions should thus be addressed in order to find an answer to the overall question of how to design digital exams:

(I) ***The technical perspective: what is technically possible and feasible?***

The digitization of exams at universities and higher education institutions is far from being trivial. Technically, there are many different options available for realising digital assessments, but they all have different advantages and disadvantages. In addition, there is the question which basic digital equipment is available at a certain university or higher education institution, i.e. which digital infrastructure is available and as a result which digital examination formats can or cannot be realised. This basic infrastructure creates the framework in which digital assessments have to be developed and implemented.

(II) ***The regulative perspective: which examination regulations have to be followed?***

In addition to technical (im-)possibilities, the question arises as to which requirements should be followed with regard to examination regulations at universities and higher education institutions. Specifically, this does not only include examination and study regulations of faculties and their individual studies, but also and in particular the question of legality and data security of digital exams. Here, too, ways have been found and have increasingly been tested in practice over the past fifteen months. Nevertheless, this question arises with a certain urgency and must—ideally at an early stage—be taken into account and be discussed.

(III) ***The didactical perspective: what is didactically reasonable?***

Each discipline has its own logics and traditions, which are also reflected in the practices of assessing, which now need to be translated into digital exam formats. From a didactic perspective, the type of knowledge that has to be tested is also decisive concerning the choice of various digital assessment formats. To be precise, didactically, the type of knowledge decides on the best mode of testing and assessing the learning outcomes. When it comes to the pure reproduction of knowledge, exams in digital formats should be structured differently compared to assessments that shall focus on the ability to apply knowledge or solving practical tasks.

All three perspectives should be included when planning digital examination formats in university and higher education institutions and are mutually dependent. Therefore, they should be viewed as a unit and not seen separately from one another. In practice, the question of what is technically feasible and the question of legal certainty of examinations are often in the foreground, as the first step of implementing digital exams often is to create the framework that enables and legitimizes digital examination formats in the first place.

What is logical from a higher education policy perspective does not go far enough from a didactical perspective. Technical standards that have been set and been implemented, affect the possibilities of the didactical examination design. Hence in the following, various common digital exam formats—e.g. a digital written test, oral exams via video conference or even the use of a proctoring system—shall be introduced including remarks on advantages and disadvantages and didactical implementations.

3 Digital Exams – Methods and Approaches

In this chapter, various possibilities and approaches of digital exams at universities and higher education institutions are presented. The list does not claim to be complete, but rather is intended to show the range of possibilities. In 2013, Krüger and Schmees carried out a basic categorization of assessment types at universities. They differentiate between five assessment types that become relevant at different points during the course of study. Figure 1 shows an overview of the different assessment categories. In the following, all types are briefly presented and classified in the context of digitization.

3.1 Type 1: Advisory Assessments

This type of assessment is usually used to support and facilitate decision-making concerning e.g. one's study programme. Classically, it is about the choice of classes and here specifically about the question of whether one's own skills and goals fit a certain course, or else which course suits your own goals and skills. This type of competence-based assessment can also be digitized. The Pennsylvania State University, for example, works with several digital competence, skill

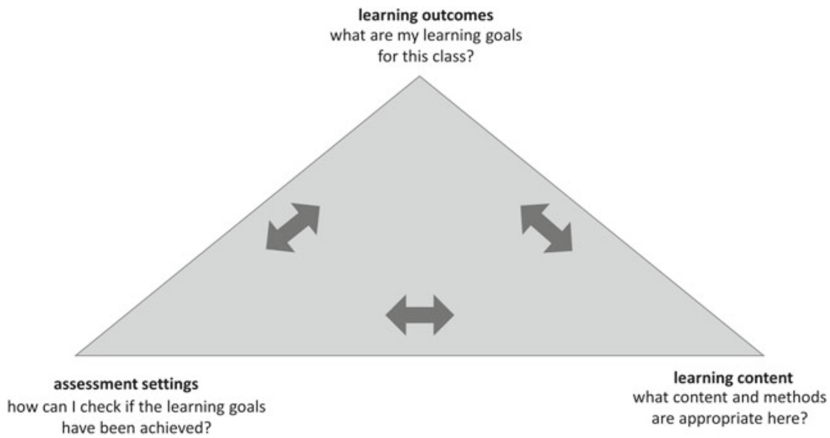


Fig. 1 Model of constructive alignment according to John Biggs (1996)

and strength assessments, which can be used to assess one's personal dispositions in order to find the study programme most suitable (University of Pennsylvania, 2021). (Potential) Students have the chance to choose between different standardised questionnaires, e.g. based on the Myers Briggs Type Indicator. After answering the questions online, they will receive the results instantly. The focus is not on excluding options generally, but rather on creating a personal profile of strengths, weaknesses, interests and personal values, which will help to choose a personal career path.

3.2 Type 2: Diagnostic Assessments

These assessments are e.g. surveys to determine whether someone has reached an intended level of knowledge or to determine the level of knowledge in a certain topic in a group in advance, e.g. to be able to adapt the content of a course to the needs of the participants. Probably the best-known diagnostic assessment in the education sector in general is the ICILS study (Eickelmann et al., 2019). The so called International Computer and Information Literacy Study (ICILS) is a widely acclaimed international comparative research study in the field of school performance research. The aim here is to identify and measure computer-related competencies of eighths grade students in different countries, to compare the skill levels internationally and to identify – on a national level – changes in skill levels both between different school types and in the course of time, as the ICILS-study is iterated every five years. If diagnostic assessments are used to prepare course content, it is important to work out the course aims and the assessment questions well in advance in order to ascertain what content might be appropriate to help the students achieve the course goals (Csapó & Molnar, 2019).

3.3 Type 3: Formative Assessments

Formative assessments are used at different stages during a learning process and accompany it. The formative assessment is not about finally evaluating a level of learning, but rather using previously defined criteria to determine whether or not someone has achieved the defined goals in their current learning process and how, if necessary, adjustments should be made in order to support learners to achieve the specified learning goals. So, this is less about evaluating, but more about navigating through a learning process. In digital settings, this can be done through feedback, which comes directly on the basis of a previous performance, such as an instant feedback, e.g. whether a given answer in a quiz or a chosen approach for solving a task has been correct or not. Many serious games work with such an instant feedback. Formative assessment can also be provided in the form of an accompanying chatbot, for example, which at the given time stimulates deepening or reorientation in the learning process with questions, for e.g. by giving questions to foster reflection on a learning topic or to help with stimulating questions, when a student has problems solving a certain task (Smutny & Schreiberova, 2020). Recommender systems can also be understood as a form of formative assessment, provided that they make personalized suggestions for the next learning steps, based on the learning goals set and the learning deficits analysed from the performance and behaviour of a learner.

3.4 Type 4: Summative Assessments

In the discourse about digital exams, references are often made to summative exams – both at the end of the semester in relation to a course and final examinations as part of an overall study programme. These are classic summative assessments that evaluate a final learning outcome based on criteria, the classic performance assessment. Summative assessments are often written exams at the end of a semester, a final presentation or an oral exam (Miller, 2020). Summative assessments can also be used in a formative way, when the results of a summative assessment are used to help students guide them through their learning processes in order to ensure they achieve a set learning goal. This requires a series of smaller assessments during the semester instead of one final exam at the end of a semester and content that can be clearly subdivided into different topics. In digital settings, even a small digital quiz at the end of a certain topic can be a summative assessment. If using digital summative assessments at the end of a semester to test the overall achievements of a student in a course, several options are available, such as electronic Portfolios, digital multiple choice exams or written essays.

3.5 Type 5: Quality Assurance Assessments

At university and higher education institutions, quality assurance assessments involve the evaluation of classes (Schmidt & Tippelt, 2005). The aim is to improve teaching based on student feedback on the type and structure of a course, but also on the performance of the lecturers and the content and methods used. Many universities already have a digital system that automatically sends an email at the end of the semester to students who have registered for a course with the request to fill out the evaluation form for the event and submit their feedback digitally. One of these systems in use is e.g. evasys, an online tool that is being used by several German universities and higher education institutions in order to evaluate course quality (evasys, 2021).

3.6 Different Scenarios for Digital Exams

The assessment types that have been described above, show the wide variety of aims that can be addressed with the various approaches. However, there are still many options to transform these assessment types into digital exam settings. The following scenarios provide a good range of what is possible and is already widely used.

3.6.1 Digital (Multiple Choice) Exams

The digital exam is a classic among digital exam formats. Such an exam is often used at the end of the semester to either query the content of a course or as a final exam for an entire module. This makes the digital exam a classic setting that can be assigned to the summative assessments. The questions are mostly constructed as multiple-choice questions. A digital multiple choice exam can be taken by all students at the same time in the same place – provided there are enough devices available for processing. However, it can also be written at different times, e.g. when a course is being split into cohorts that complete the exam at different times. It is also possible to have the exam done completely self-paced and completely unsupervised. However, this places special demands on the formulation of the exam questions and the construction of the overall exam. This is discussed in more detail in Chap. 3.

3.6.2 Proctoring

Proctoring as such does not actually represent a separate examination format, but should nevertheless be considered separately here, as this topic in particular has sparked controversial discussions. Proctoring is a technical variant of the supervision of digital exams. With the help of digital technology, it is supervised that the person registered is actually taking the exam, that no cheating is possible and no unauthorized aids are being used. To make this possible, however, technical interventions are necessary, which are controversial in terms of data protection and the abidance of GDPR (Cuijpers, 2021). For example, the webcam of the students' computer is accessed externally. The movements of the mouse pointer and the open windows of a web browser can also be viewed externally. Both is ethically critical and borderland within GDPR. The use of proctoring in examinations at universities and higher education institutions enables the decentralized execution of examinations on the one hand. On the other hand, the limits to monitoring the students are fluid here and the extent to which the use is following data protection acts must be assessed on a case-by-case basis. In fact, proctoring services are already being used in exams at various universities and higher education institutions around the world, e.g. by several American universities such as the University of Houston (2021) or the University of Toronto (2021). But also European universities, such as the Università degli Studi di Milano (2021), the Tilburg University (2021) or the Hasso Plattner Institute (2021) offer proctoring services to their students.

3.6.3 Oral Live Exam

Another variant of performing a classic university exam scenario digitally is the live video exam. This can both be offered as a formative or summative setting that is relatively easy to digitize. In contrast to the classic exam, the examinee, examiner and recorder meet via video conference software in the digital room and can carry out the exam here. Per se, there is still the possibility of accompanying a live exam with Proctoring in order to rule out attempted fraud. More practicable and didactically clearer, however, is the formulation of tasks and questions in a way that complicates attempts of deception, which is still possible in live video exams, e.g. through the parallel use of information from the Internet. The possibility to cheat can also be reduced by choosing the questions for the exam wisely and in an application-oriented way, instead of merely focusing on the reproduction of knowledge.

3.6.4 Open-Book Exam

So-called open-book exams or take-home exams are digital exam formats that are not carried out under supervision, but are processed by the students independently. It is a written exam that can be completed anywhere. They can be offered both as summative and formative assessments. The tasks are given to the students on a specific date and must have been completed by the students at a specific deadline, e.g. 24 h later. As a rule, the tasks are set digitally and the solutions are transmitted electronically. The use of aids is usually allowed. And it is individually regulated and determined in advance, which resources may be used. The faculties can determine their own regulations on helpful resources, the processing period for an exam and the scope of the exam tasks (Universität Hamburg, 2021). Open-book exams should therefore be designed in an application-oriented way and allow the use of aids of various kinds. Since the questions are announced at a set point in time and the processing time is usually tight, this limitation also makes fraud attempts more difficult, so that open-book exams are a good alternative to digital multiple-choice exams.

3.6.5 E-portfolio

E-portfolios are another way of digitizing exams at universities and higher education institutions. The study portfolio is actually a classic formative examination approach in which students usually create a portfolio for a specific task over the course of a semester. In the case of an e-portfolio, this is done digitally, e.g. on the university's own learning platform. The task for which a portfolio is created is defined at the beginning of the semester. It can be the same task for everyone or can be chosen individually by each student as agreed upon with the lecturer. The scope and goal of the portfolio must be clear from the beginning. Theoretically, an e-portfolio can also be part of an open-book exam, but in this case, it must be ensured that the type and scope of the task is meshed with the time given for the completion of the task. The strength of the e-portfolio lies in the fact that its extensive collection can be created for an (independent) question and later made accessible to others. Due to the longer editing time, discussions in between with the lecturers are also possible so that any questions can be clarified and, if necessary, the course of the learning process can be corrected (Hericks, 2020).

3.6.6 Self-assessments and Peer Reviews

Other formative-oriented exam settings are the so-called self-assessment or peer review procedures, which are particularly suitable for digital use. The students of a course independently process one or more short work assignments according to specifications given beforehand and make their results digitally available to the other students in the course. This is often done through a university's learning management system. The students' contributions can range from short written tasks up to a set number of discussion contributions in the class forum. The discussion can be completely self-directed or managed by the lecturer. In the second step, the students assess the contributions of their fellow students with regard to predetermined criteria. Compliance with feedback rules is particularly important here. Because all students produce their own contributions and have to rate a fixed number of third-party contributions, each of the students receives both one – mostly several – ratings or feedback and at the same time deals intensively with a topic (Katzlinger et al., 2018).

3.6.7 Semester Projects, Business Games Etc.

Similar to e-portfolios, several other approaches can be integrated into classes as formative digital assessment formats, such as semester projects, simulation games, the development of individual concepts on a certain topic or the design and creation of one's own media content. In any case, students usually develop their own project over the course of a semester. The goal here does not necessarily have to be the creation of a portfolio, but can also be a finished concept, an independently conducted workshop, the participation in a simulation game, a written reflection on a practical activity or the creation of different types of digital media content (e.g. instructional videos, a podcast, a photo project etc.). In particular, the creation of digital media can also be very well linked with the "flipped classroom" approach, in which students are specifically involved in order to first develop content themselves and then convey it to fellow students, e.g. via instructional videos that they have created themselves (Bühner & Sommer, 2020). In the digital space there is the possibility to work together decentrally and asynchronously on the development of a content and at the same time to create your own knowledge base.

These approaches already are practiced at universities and higher education institutions. Of course, mixed forms are conceivable and possible, so that the transitions between the settings outlined here are fluid. The subject area, topic and examination specifications constitute the choice of the appropriate exam setting. In some departments, for example, exams are prescribed, in other departments there is more room for manoeuvre in the choice and design of the exams. As a result of the corona pandemic, freedom has arisen in many places in the last few months that allow and enable a different approach to examinations, that permit and also make new ways of examination possible.

This results from the decentralization and asynchrony of many digital exam settings as well as from the simplified possibility of cheating with digital aids. Both aspects can be cushioned by a clever design of the examination tasks and, if done so, also have a didactic benefit. In this respect, digital assessments can lead to a strengthening of competence-based exams, provided this is done correctly. Hence in the following, some considerations on the didactic design of digital exam questions will be carried out.

4 Didactical Considerations for Formulating Digital Exam Questions

An assessment understood as a knowledge query is always inseparably linked to the preceding transfer of this knowledge. Accordingly, the design of the examination must already be considered when designing the content and methodology of a course. For the planning of courses in universities and higher education institutions that will end with a digital exam, a few points must be considered, which will be explained in more detail here.

4.1 The Biggs Model of Constructive Alignment

In practice, the concept of constructive alignment has proven itself for planning courses in higher education (Biggs, 1996). The approach was developed in the 1990s by the Australian psychologist John Biggs and was lately received and discussed in the higher education sector for questions of higher education didactics in particular. The model is based on the conceptual design of classes on the basis of three dimensions. The focus is on the design of the learning objectives (1) and the exam settings (2) in order to check the learning objectives previously communicated in the course. The teaching and learning methods (3) as well as the specific content of the course are only defined when the assessment settings and the learning objectives are clearly defined (Biggs, 1996).

Critics accuse the model of restricting the learning process itself to testable knowledge and reducing the success of a learning process to achieving the set learning objectives. There is a risk that one's own creative learning and solution paths will be ignored and valued less. In a constructivist tradition of thought, there is no way to actually check whether the teaching goals set by the teacher are actually translated one-to-one into personal learning goals and furthermore learning outcomes.

Undoubtedly, these are relevant objections that must be considered in the further theoretical discussion, but do not apply exclusively to the model of constructive alignment. They generally highlight the dilemma of checking knowledge and learning success.

For the practical question of the design of examinations in a university and higher education context, the orientation towards constructive alignment can be a helpful approach to design classes and to synchronize them with the possibilities of digital examinations at a certain university or higher education institution. Therefore, the model should be taken as a basis here and will be explained in more detail below.

4.1.1 Dimension of Learning Objectives

In the constructive alignment model, all considerations are based on the definition of learning objectives. The point here is to determine which learning objectives should or must be conveyed, e.g. based on the requirements of the study regulations, descriptions in the module handbook, etc. where the learning objectives are also specifically formulated. Learning objectives and the content to help achieve the learning objectives, are strongly linked. An example: Constructivism as an epistemology can be a learning content. One learning goal here can be, for example that at the end of the class, students can explain precisely what constructivism is and who its most important representatives are. At this point it is not yet necessary to break down the material and the learning objectives into individual seminar sessions. This only takes place in the third planning step.

4.1.2 Dimension of Examination Methods

In the model of constructive alignment, the determination of the learning objectives is followed by the determination of the examination methods. The questions and the method for checking whether the learning objectives have been achieved are derived directly from the learning objectives defined for a course—even before the structure of the course is precisely planned.

The examination regulations often leave room for several settings and accordingly the focus here should be on how the achievement of the defined learning objectives can be checked best. Tests can be carried out in a wide variety of completely analogue settings, but also in completely digital and also in hybrid forms, meaning mixed analogue and digital settings. At this point the decision should be made in which way and for what purpose digital media shall be used in the exam.

The decision is based on how exam questions are specifically formulated in comparison with the learning objectives and which work assignments are actually given in an exam (see Sect. 3.3).

4.1.3 Dimension of Teaching and Learning Methods

Only after setting learning goals and examination methods, the teaching and learning methods, which shall help convey the learning content and achieve the learning objectives, will be chosen. Usually certain knowledge should be imparted or presented in an adequate way, a logical order, at a certain point in time, etc., so that the planned examination can be carried out. Depending on the requirements of the study and examination regulations, limitations may have to be taken into account, such as a given examination format. In principle, however, the choice of teaching and learning methods should be based on three criteria:

4.1.4 Teaching and Learning Methods According to the Class Content

First and foremost, it is a matter of choosing the teaching and learning method that is best suited to convey the material and support the achievement of the learning objectives defined for a class. Here you can orientate yourself on learning objective taxonomies. Learning objective taxonomies indicate on which abstraction level a learning objective is to be located, whether it is primarily about learning content by heart and reproducing it, or whether it is also about analysing content, about (practical) application, etc.

4.1.5 Teaching and Learning Methods According to the Workflow

The workflow of a class also has a structuring effect on the selection of teaching and learning methods. If, for example, practical phases are part of a class, they will probably not take place at the beginning of a seminar, but rather in the middle of the semester or at the end. Sometimes, practical phases and phases of theoretical considerations can also alternate. If certain content or a special procedure has to be conveyed by then, this must be done before the practical phase. This does not only have to be scheduled in terms of time, but also methodically.

4.1.6 Teaching and Learning Methods According to Learning Preferences

Ideally, teaching and learning methods should be varied anyway. Not everyone learns well by listening or reading. Many people need visual support or they learn best when they apply and try out new knowledge. This should be considered when choosing the teaching and learning methods for a class as well. Ideally, different options are made available in parallel here, so that all learners have the chance to choose the learning method that suits their needs best.

4.2 Find and Set Learning Objectives as the Basis for Formulating Examination Questions

Formulating learning objectives clearly and unambiguously is an important prerequisite for being able to develop examination formats that work properly in terms of assessing the level of achieving the learning goals. Here, one of the oldest didactic principles applies: from the easy to the difficult, from the precise to the conceptional. This means that content should build on one another, that the basics are taught first and only then the specifics. Of course, this must also be considered when setting the learning objectives for a class.

Anderson and Krathwohl have presented a revised learning objective taxonomy based on Bloom's learning objective taxonomy in 2001. While Bloom's taxonomy of learning objectives, dating back to 1956, is primarily aimed at the acquisition of cognitive skills (Bloom et al., 1956), i.e. the pure acquisition of knowledge, Anderson and Krathwohl also include the acquisition of competencies. They expand Bloom's learning objective taxonomy by adding a process dimension. Figure 2 illustrates this.

The original taxonomy according to Bloom provided six consecutive dimensions for the formulation of learning goals, which Anderson & Krathwohl adapted in nuances and called "the cognitive process dimension". This includes: remember, understand, apply, analyse, evaluate, create. The simple reproduction of knowledge is therefore on the most basic level of the taxonomy. Understand means the understanding of this knowledge based on it. Apply means to be able to apply the newly acquired knowledge. Analysis means to penetrate the newly acquired knowledge in its depth. Evaluate stands for developing your own well-founded attitude towards the new knowledge. And ultimately, create means to further develop the newly acquired knowledge independently.

the knowledge dimension

metacognitive knowledge and awareness of cognition, both one´s own and in general	identify	predict	use	deconstruct	reflect	create
procedural knowledge of criteria and application of methods and techniques	recall	clarify	carry out	integrate	judge	design
conceptual awareness of the interrelationships between different elements	recognize	classify	provide	differentiate	determine	assemble
factual basic factual knowledge to be acquired with discipline	list	summarize	respond	select	check	generate
	remember knowledge is recallable, even after a long time	**understand** deduce meaning from the instructional material	**apply** use knowledge and methods in relevant situations in an appropriate way	**analyze** deconstruct knowledge and determine the relations between the parts	**evaluate** assess an issue according to relevant criteria	**create** rearrange and reorganise elements to something new

the cognitive process dimension

Fig. 2 Learning objective taxonomy according to Anderson and Krathwohl (2001), model adapted by Iowa State University (2012)

Anderson and Krathwohl added a second axis to the model, which deals with the knowledge dimension, precisely with the type of knowledge that is classified in the cognitive process dimension. The model distinguishes between four sub-categories: factual, conceptual, procedural, metacognitive. Factual refers to numbers, data and facts. Conceptual means knowledge that includes ideas, concepts, constructs etc. Procedural refers to process-oriented knowledge, patterns of action, etc. and metacognitive knowledge means abstract knowledge, knowledge on the meta level, etc.

The result is a matrix that helps to classify the knowledge itself and its complexity factor when it comes to learning processes – from easy to difficult, from precise to conceptional. This gives a very good overview of the levels at which learning objectives can be classified, which level of difficulty prevails in the planning for a certain class and where, if necessary, additional information is required to make it possible to achieve all learning goals within a class.

4.3 Derive Exam Questions from Defined Learning Objectives

The learning objectives of a class also form the basis for the formulation of questions and tasks for later exams. The point is to check properly and adequately whether the learning objectives have been achieved or not. Therefore, in the logic of constructive alignment, after the formulation of the learning objectives, the examination questions and tasks are first determined before the content and teaching methods are selected.

When determining exam questions and tasks, it is advisable to look at each learning objective individually and to consider how it can be determined: Are theoretical questions suitable here? This is usually the case with learning objectives that aim at factual knowledge. Is it about applying practical skills? Then another exam setting might make more sense, as the associated learning objective is likely to be aimed at procedural or metacognitive knowledge.

Table 1 is a helpful instrument to both formulate learning objectives and assessment questions for (digital) assessments. For each learning objective, an examination question or examination task should be specified. In addition, attention should be paid to whether there are overarching learning objectives that need to be considered.

The University Rectors' Conference has developed a guide for the formulation of learning objectives, which gives examples of how learning objectives can be formulated well and clearly and how they can be assigned to the individual levels

Table 1 Categorization of assessment types in university teaching according to Krüger and Schmees (2013)

Type	Moment	Purpose	Decision
Advisory	Before beginning studies	Orientation, course guidance	Recommendation on study programmes
Diagnostically	Before starting a learning process	Categorisation, admission	Choosing the right class
Formative	Within the learning process	Reflect the current learning outcome	Manage the learning process
Summative	At the end of the learning process	Reflect the final learning outcome	Performance assessment
Quality control	At the end of a course	Improve teaching	Suggestion for improvement

Table 2 Set learning goals—table based on the taxonomy of Anderson and Krathwohl (2001)

Dimension of cognitive processes							
Dimension of knowledge		Remember	Understand	Apply	Analyse	Evaluate	Create
	Factual						
	Conceptual						
	Procedural						
	Metacognitive						

of the learning objective taxonomy. The defined verbs that describe the cognitive processes linked to the several levels of the learning taxonomy, e.g. *describe, summarize* or *present* as verbs that can be used to formulate learning objectives in the second level of the learning taxonomy, which aims at *understanding*. (Hochschulrektorenkonferenz, 2015). This can also be applied to the formulation of exam questions and used as a control mechanism to check to what extent the exam questions also cover the defined learning objectives (Table 2).

4.4 Develop Exam Questions for the Digital Context

The formulation of exam questions is not a trivial task, even in classic higher education teaching and learning settings. In addition to content-related and formal requirements, when digitizing exams, the implications of these technical components with regard to the formulation of exam questions must be considered. There are a few challenges to bear in mind:

4.4.1 The Choice of the Exam Setting Determines the Possible Type and Range of Exam Questions

The different digital assessment scenarios have already been shown in Chap. 2. Depending on the examination requirements of a faculty and the degree of freedom in exam design for each lecturer, the general regulations for digital exam setting limit the options for assessment question design, or in each case specifies how exam questions can and should be designed.

A digital multiple choice exam, for example, demands to formulate clear and well-defined questions, including key phrases both in the question itself and – in case of open answering formats – in the sample solution as well. For multiple choice questions, the lecturer needs to develop several answer options for each question, of which only one – or, depending on the type of task, several – are correct for each question without overlapping.

4.4.2 Phrase the Questions Based on the Learning Objectives

Yet this is not enough to set a good assessment task. Here it is much more necessary to take up the key terms from the learning objective formulations in the examination questions and in the answer options. The precise formulation of the questions is important here and should also be chosen in accordance with the learning objective taxonomies in the sense of a competence-oriented examination. Table 3 shows how this can work (Vogt & Schneider, 2009).

Here you can see the direct link between an assessment task chosen and the learning goal specified. Thus, before planning the tasks precisely, the lecturer should have a clear idea of what is intended to achieve with choosing one or another type of question or setting and check with the help of the learning objectives if this corresponds. Thus, the constructive alignment model can be helpful, when planning and designing a course. It allows a very structured and well clear way to do so. Many lecturers start planning their classes by choosing the content they want to convey and the methods to deliver it. By starting endwise with defining the learning objectives and as a next step planning and opposing the examination tasks, the whole course will finally be more consistent and target the learning objectives more accurately.

Table 3 Question types relating to learning objectives according to Vogt and Schneider (2009)

Question type	Learning objective
Multiple choice	• Recognize terms • Understand concepts • Recognize connections • Evaluate settings
Allocation	• Relate terms • Assign concepts to each other • Recognize hierarchies
Sorting	• Analyse processes • Reduce historical developments
Short texts	• Reproduce terms • State mathematical results numerically reproduce numbers
Long terms	• Describe the solution of a problem define a position
Cloze	• Understand the structure of a text • Reproduce terms

4.4.3 Develop Application-Oriented Questions

As soon as the learning objectives of a class go beyond the pure imparting of knowledge, it is advisable to ask application-oriented examination questions, which is possible in all digital assessment settings. This is also especially helpful for digital examination formats. The application orientation actually allows a check to be made as to whether an issue can be learned and transferred to other situations and also makes any attempt to cheat more difficult. Application-oriented examination questions will mostly be reflected in the upper levels of the learning objective taxonomy according to Anderson and Krathwohl (2009) on the process dimension and should hence be phrased according to this.

Application-orientation can be realised both in oral or written exam settings, in assessments focussing on theoretical and practical knowledge and in almost any examination setting. The same rules for phrasing the questions and considering learning goals apply here. Yet the questions aim at the upper levels of the learning taxonomy according to Anderson and Krathwohl, which means – in practice – that most application-oriented questions will aim at evaluating a certain matter or creating own concepts or content based on the course topics.

4.4.4 Benefit from the Technological Possibilities

In digital exam settings in particular, the time at which the knowledge is applied and the test to determine whether the knowledge has been applied correctly or not can differ. Hence digital exam settings are often asynchronous settings. In a digital live exam, both will still take place synchronously. If, e.g., an e-portfolio approach is chosen instead, the point in time at which knowledge is applied falls into a different period than when it comes to determining whether the knowledge has been acquired and applied correctly.

This must be taken into account when designing the tasks: How much time does the examinee have available to solve the task? Can the task be solved in the given time? Which tools can be used? Does the examinee have definite access to it? Especially with digital exams that are structured in an application-oriented manner, it must always be considered that there must be access to sources that are permitted and required, and that there is also the possibility that someone can gain access to sources that are actually not approved for an examination.

4.4.5 Minimize the Risk of Digital Fraud by Phrasing the Examination Tasks

Regardless of the exam format itself, the formulation of the exam questions enables you to minimize the risk of fraud in digital settings. If, for example, a certain content or tasks are taught early on in a class that are explicitly referred to later in the examination, this can already help to prevent attempts at fraud, as the relevance of the subject matter in detail is only revealed later. If a course is planned on the basis of constructive alignment and learning objectives and examination questions are developed and related to one another based on a learning objective taxonomy, this fact arises almost organically, even in digital examination formats.

In addition, e.g. for digital exams that are written by different examinees at different times, a randomized procedure can be used. In this case, several exam questions are formulated for each learning objective. However, not every question is asked for every exam date, but the selection is made randomly. It is only ensured that each learning objective is checked. So, the solutions to questions cannot be passed on so easily and attempts at fraud in this direction are made more difficult.

4.4.6 Use Interaction-Oriented Exam Questions

In principle, attempting to cheat application-oriented exam questions is more difficult, but of course still possible. Therefore, in digital exam settings, it is important to interweave application-oriented tasks with the seminar content and to set tasks in an interaction-oriented and discursive manner. In this way, self-reflection is stimulated and it can hardly be copied and copied by others. The more personal contributions an exam setting contains, the lower the risk of attempting fraud.

Accordingly, questions would not be asked with the aim of reproducing or reflecting on an object of knowledge, but rather go further in the sense of one's own application of the knowledge to a (also self-chosen) topic with the aim of further development. This type of construction of examination questions naturally requires a different way of thinking on the part of teachers than has often been the case up to now. This kind of questions usually require the class content as a basis for developing an individual, applied solution on a defined problem.

In the past, mostly, exam questions were asked with a view to the reproduction of knowledge from memory, possibly with a tendency towards application orientation. In the digital space, the application orientation becomes even more important not only in order to prevent fraud, but furthermore to achieve a real examination of an object of knowledge – from a didactic point of view this is most welcome.

5 Challenges and Open Questions

The arguments and explanations in this article show that the formulation of exam questions in digital exam settings is not only decisive for making attempted fraud significantly more difficult or even prevent it. It is also decisive from a didactic perspective. Once again it becomes clear that the debate about the digitization of learning processes is only superficially a technical one. Basically, it always comes down to focusing on originally pedagogical and didactical questions: What does it take to stimulate and shape a good learning process? Bearing this in mind, the debate on digital exams in universities and higher education institutions can lead to new perceptions and open up possibilities for development. The following thoughts offer starting-points for such a discussion.

5.1 Rethink Exam Settings in Universities and Higher Education Institutions

In particular, when designing digital exam settings, there are many opportunities to rethink and redesign exams at universities and higher education institutions and to focus even more on the didactic components of a higher education exam. The prerequisites here are, of course, technical possibilities and conditions on the one hand, but even more the openness of the faculties to recognize and systematically establish new, more innovative exam settings.

5.2 Consider Solutions for Data Security Issues

There are of course several questions that have not yet been discussed and clarified. Again, and again, the issue of legal certainty of various digital exam formats is in the room. There is certainly still some need for reflection and adjustment here, both with regard to fraud security and with regard to data protection issues.

5.3 More Empirical Research on Digital Learning Methods and Digital Exam Settings

Of course, more educational research is needed that deals with the different digital exam formats and starts here in a concrete and application-oriented manner. Close interlinking between educational, media and information technology research would be helpful here in order to find out more about which settings works best for whom, and where further technical support may be needed, etc.

5.4 Foster Diversity in (Digital) Exam Settings

Last but not least, the digitization of university and higher education examination formats has resulted in a trend that can generally be found again and again in the debate about digital learning: the trend towards more formalization of learning and examination processes and the trend towards increased monitoring, such as is possible through proctoring. Both are to be viewed critically. The debate about digital learning tends to first focus on the technical component. This seems to be promoted by the attempt to translate analogue settings one-to-one into digital settings. The problems that arise as a result are counterbalanced through more mechanization and more surveillance.

Here it would be important to remain open to use the possibilities and advantages of digitization by developing new ideas and new ways to shape learning processes – there are already approaches and practical experiences! And to remain critical of excessive formalization and to enter into the discourse here. Ultimately, it's never about technology as an end in itself. It's always about making learning processes stimulating and helpful.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Pearson Education.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *High Education, 32,* 347–364. https://doi.org/10.1007/BF00138871

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. The classification of educational goals. Handbook I: The cognitive domain.* David McKay.

Bühner, B., & Sommer, J. (2020). Der Inverted Classroom – eine Königsdisziplin der digitalen Hochschullehre? In M. Friedrichsen & W. Wersing (Eds.), *Digitale Kompetenz. Herausforderungen für Wissenschaft, Wirtschaft, Gesellschaft und Politik* (pp. 129–134). Springer. https://doi.org/10.1007/978-3-658-22109-6_13.

Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology, 10,* 1522. https://doi.org/10.3389/fpsyg.2019.01522

Cuijpers, C. (2021). Online proctoring put to the test. In E. Aarts, H. Fleuren, M. Sitskoorn, & T. Wilthagen (Eds.), *The new common. How the Covid-19 pandemic is transforming society* (pp. 53–58). Springer. https://doi.org/10.1007/978-3-030-65355-2_8.

Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., & Vahrenhold, J. (Eds.). (2019). *ICILS 2018. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking*. Waxmann. https://kw.uni-paderborn.de/fileadmin/fakultaet/Institute/erziehungswissenschaft/Schulpaedagogik/ICILS_2018__Deutschland_Berichtsband.pdf.

Evasys. (2021). *Befragungs- und Prüfungssoftware*. https://evasys.de/.

García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande-de-Prado, M. (2020). Recommendations for mandatory online assessment in higher education during the COVID-19 pandemic. In D. Burgos, A. Tlili, & A. Tabacco (Eds.), *Radical solutions for education in a crisis context* (pp. 85–98). Publisher. https://doi.org/10.1007/978-981-15-7869-4_6.

Hasso-Plattner-Institut. (2021). *Die Online-Aufsicht*. https://open.hpi.de/pages/proctoring_details.

Hericks, N. (2020). *Das Portfolio als Prüfungsinstrument in der Hochschullehre.* Erprobung des ePortfolios im Rahmen der Lehrerbildung. Goethe-Universität Frankfurt am Main. https://doi.org/10.4119/hlz-2730.

Hochschulrektorenkonferenz. (Ed.). (2015). *Lernergebnisse praktisch formulieren.* https://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-02-Publikationen/Lernergebnisse_praktisch_formulieren_01.pdf.

Iowa State University (2012). *A 3-dimensional model of bloom's taxonomy*. https://www.teachthought.com/critical-thinking/3-dimensional-model-blooms-taxonomy/.

Katzlinger, E., Herzog, M. A., & Franz, L. (2018). Digitales Peer Review in der Hochschullehre. Wie eine Evaluationsmethode den Lernprozess direkt unterstützt. In S. Harris-Huemmert, P. Pohlenz, & L. Mitterauer (Eds.), *Digitalisierung der Hochschullehre. Neue Anforderungen an die Evaluation* (pp. 57–70). Waxmann.

Krüger, M., & Schmees, M. (Eds.). (2013). *E-Assessments in der Hochschullehre. Einführung, Positionen und Einsatzbeispiele*. Internationaler Verlag der Wissenschaften.

Miller, A. (2020). *Summative assessment in distance learning*. https://www.edutopia.org/article/summative-assessment-distance-learning.

Schmidt B., & Tippelt R. (2005). Lehrevaluation. In I. Gogolin, H. H. Krüger, D. Lenzen, & T. Rauschenbach (Eds.), *Standards und Standardisierungen in der Erziehungswissenschaft. Zeitschrift für Erziehungswissenschaft* (Vol. 4, pp. 227–241). VS Verlag. https://doi.org/10.1007/978-3-322-80769-4_15.

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education, 151*. https://doi.org/10.1016/j.compedu.2020.103862.

Tilburg University. (2021). *Online proctoring with Proctorio*. https://www.tilburguniversity.edu/students/studying/exams/online-proctoring/proctorio.

Università degli Studi di Milano. (2021). *Online proctored quiz exams*. https://www.unimi.it/en/study/bachelor-and-master-study/following-your-programme-study/sitting-exams/distance-exams/online-proctored-quiz-exams.

Universität Hamburg. (Ed.). (2021). *Take-home-exams*. https://www.uni-hamburg.de/elearning/methoden/online-pruefen/take-home-exams.html.

University of Houston. (2021). *Proctoring services*. https://uh.edu/online/students/proctoring-students.php.

University of Pennsylvania. (2021). *Self-assessments*. https://careerservices.upenn.edu/self-assessments/.

University of Toronto. (2021). *Online proctoring*. https://teaching.utoronto.ca/ed-tech/online-learning/online-proctoring/.

Vogt, M., & Schneider, S. (2009). *E-Klausuren an Hochschulen*. http://geb.uni-giessen.de/geb/volltexte/2009/6890/pdf/VogtMichael-2009-02-20.pdf.

Julia Hense is an expert in digital learning, educational counseling and educational policy. She holds a doctoral degree in educational science and has been dealing with questions related to digital instructional design for many years. After various positions in the research sector, e.g. at the DZHW in Hanover and the University of Bielefeld, Julia Hense worked for the Bertelsmann Foundation for several years. Here she set up the "Monitor Digitale Bildung", a large-scale nationwide study on the situation of digital education in Germany. At mmb Institute, she is responsible for the AI-Campus, an online learning platform on the subject of artificial intelligence.

Trust and Cheating: From Assessment Literacy to a Literacy of Practice for Digital Remote Assessments

Alexander Schulz

Abstract

The outbreak of the coronavirus pandemic in early 2020 and the subsequent lockdowns struck at the very heart of the operational activities of higher education institutions (HEIs) worldwide. Almost from one day to the next, teaching and learning needed to be radically reconsidered. Many HEIs either had to massively reduce on-campus teaching for hygienic reasons or were forced to replace it altogether with emergency remote formats in which students participated from home (Hodges C, Moore S, Lockee B, Trust T, Bond A. The difference between emergency remote teaching and online learning. Educause Review, 2020). As HEIs struggled to carry on providing academic education under the most extraordinary circumstances, not only teaching and learning, but also the practice of assessments had to be rethought from the ground up. During the search for ad-hoc actionable options, it became blatantly obvious that the macroscopic relationships and dependencies associated with remote digital assessments went far beyond issues of didactical and psychometrical preparation. Given that the legal reliability of assessments is of utmost importance for HEIs, operational measures in four key fields of actions had to be applied: (1.) law and regulations; (2.) technology; (3.) didactics and psychometrics; and (4.) organisation and logistics. Following the trail of assessment literacy (Wollersheim H-W, Pengel N. Von der Kunst des Prüfens – Assessment Literacy. HDS. Journal – Perspektiven Guter Lehre, (2), 14–32, 2016) and bearing in mind the specific legal situation in Germany (Fischer

A. Schulz (✉)
Center for Digital Systems (CeDiS), Freie Universität Berlin, Berlin, Germany
e-mail: alexander.schulz@cedis.fu-berlin.de

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_7

E, Dieterich P. Prüfungsrecht in Zeiten der Coronavirus-Pandemie. NVwZ - Neue Zeitschrift Für Verwaltungsrecht, 10, 657–666, 2020), the present article aims at outlining a basic framework for a literacy of practice for digital remote assessments.

Keywords

Assessment literacy • Digital remote assessments • Literacy of assessment practice • Legal reliability • Quality criteria

1 Introduction

Even under extraordinary circumstances, higher education institutions (HEIs) have the obligation to enable students to graduate, and to continuously develop applicable educational scenarios while maintaining academic integrity. In the case of the coronavirus pandemic, meeting this obligation entailed the implementation of emergency remote teaching (ERT) (Hodges et al., 2020) and emergency remote assessment (ERA) scenarios on extremely short notice and on a previously unimaginable scale.

Scope of the Article

The present article focuses on summative assessments conducted remotely. Even before they had to be carried out in this way, summative assessments were in a league of their own in German HEIs due to their purpose of *legally certifying* the level of a student's achievements. Since failing summative assessments (named "Klausuren" in Germany) repeatedly may result in relegation from the university and hence in a restriction of the freedom of career choice, precise legal regulations apply.

Divided into three parts, this article intends to approach the topic from a practice-oriented perspective. Aimed at extending the field of assessment literacy (Wollersheim & Pengel, 2016) to a *literacy of practice for digital remote assessments*, it first provides a brief introduction into the basics of assessment validity, reliability, objectivity (Krebs, 2008), and practicality (Compos, 2020), and summarizes some of the main empirical findings concerning academic misconduct during the coronavirus pandemic (Janke et al., 2021; Bilen & Matros, 2020). The second part of the article discusses the fundamental differences between digital on-campus assessments and digital remote assessments. In order to lower the entry threshold into the topic, we assume that on-campus assessments share similarities with laboratory experiments in that both are conducted under controlled conditions. Remote assessments are markedly different in that regard, as off-campus environments (including, for example, the homes of examinees) are outside of HEIs' direct control. This perspective allows us to identify at least three fundamental research questions:

I Which conditions can be controlled within on-campus and remote environments?
II Which legal, technical, didactical, and psychometrical as well as organisational and logistical measures are available for controlling conditions during on-campus assessments, and which measures can substitute for the lack of supervision in remote settings?
III What effects can these measures be expected to have on the validity and reliability, and thus the overall quality of digital remote assessments?

As we attempt to answer these questions, the main differences between familiar digital on-campus assessments and novel digital remote assessments are sketched out, while actionable measures are identified and theoretically evaluated for efficacy with the help of quality criteria. The third and final section of the article draws a conclusion by outlining the benefits of, and challenges for, the sustainable establishment of digital remote assessments in German HEIs.

1.1 From Assessment Literacy to a Literacy of Practice for Digital Assessments

The concept of assessment literacy is mainly focused on didactical and psychometrical aspects. We therefore propose to extend it to a comprehensive *literacy of practice for digital assessments* that encompasses the assessment process as a whole. As a starting point, we use Webb's definition:

> "**Assessment literacy** is defined as the knowledge of how to assess what students know and can do, interpret the results from these assessments, and apply these results to improve student learning and program effectiveness" (Webb, 2002, p. 1)

and transpose it to a macroscopic level:

> **Literacy of practice for digital assessments** is defined as the knowledge of how to apply (1.) legal, (2.) technological, (3.) didactical and psychometrical, and (4.) organisational and logistical measures in order to facilitate digital assessments.

Fields of Action for Digital Remote Assessments

Taking a macroscopic view of digital assessments leads to the realisation that the various measures belong to distinct fields of action and are thus carried out by different departments of a HEI (Schulz, 2021). It should be noted that there may be significant differences regarding the ease (and thus also the speed) with which these measures can be brought to bear in the various fields – for example, legal measures (e.g., changing the study and examination regulations) can be implemented at a much slower pace than organisational or logistical measures (e.g., allocating more rooms or personnel). Given such a vastly different processing speed, it proved to be advisable to determine which requirements in one field of action might be met by carrying out measures in alternative fields. The following table exemplifies the fields of actions and their objectives (Table 1):

Table 1 Fields of action for digital assessments

Field of Action	Objectives	Responsibility	Examples
Laws and regulations	Maintaining legal validity and reliability by implementing measures assuring both equal opportunities and privacy protection	Legal departments	• Study regulations • Examination and assessment regulations • Data protection regulations
Technology	Translating legal and didactical requirements into technical measures	IT service units	• Stability and security of the digital assessment platform • Provision of randomised task order
Didactics and psychometrics	Creating quality-assured assessment tasks from intended learning outcomes	Lecturers	• Selected Response Tasks • Constructed Response Tasks
Organisation and logistics	Designing scalable work processes for assessments by monitoring all measures and by translating legal, technical, and didactical requirements into organisational and logistical measures	Administration	• Holistic description of work processes, fallback and emergency plans • Providing personnel

The fields of action outlined above are to be understood as a low-threshold proposal for a generalised framework. Necessarily simplified, it may not be applicable to every HEI – for instance, in some cases, organisational and logistical measures can also be carried out by IT service units. Moreover, "Organisation and Logistics" constitutes an overarching field of action, as it comprises, controls, and monitors measures to be implemented in all other fields. It should also be noted that the logistical aspects listed here are geared towards on-campus assessments, and therefore do not adequately represent the requirements for digital remote assessments.

Quality Criteria: Validity, Reliability, Objectivity, and Practicality

In the pursuit of a literacy of practice, extending the definitions of quality criteria to encompass all fields of action is an indispensable step. In what follows, we will generalise Krebs' didactical and psychometric definitions of *validity*, *reliability,* and *objectivity*, as well as Compos' definition of *practicality* to render them applicable to the fields of action described above. First, we take Krebs' definition of validity:

> **Assessment validity** "asks whether the process in question really measures what is intended" (Krebs, 2004, p. 30, own translation).[1]

By adding Munzer's juridical definition, where validity is understood as the legal validity of legal rules and laws (Munzer, 1972, p. 3), we extend Krebs' definition to match the most important field of action—the laws and regulations which are in force or have yet to be created to facilitate digital assessments:

> **Legal validity of digital assessments** is defined as a quality criterion that asks how to comply with legal requirements while conducting digital assessments, or how to create legal conditions or frameworks that facilitate digital assessments.

Second, we then perform a similar generalisation regarding Krebs' definition of reliability, which "asks how precisely a given characteristic is measured" (Krebs, 2004, p. 30, own translation)[2]:

> **Reliability of digital assessments** is defined as a quality criterion that asks how precisely an assessment can be repeatedly conducted.

[1] "Validität: Gültigkeit. Prüfungsqualitätskriterium, das danach fragt, ob das betreffende Verfahren wirklich das misst, was beabsichtigt ist" (Krebs 2004, p. 30).

[2] "Zuverlässigkeit. Prüfungsqualitätskriterium, das danach fragt, wie genau ein Merkmal gemessen wird (..)" (Krebs 2004, p. 30).

As far as the legal requirements are concerned, reliability of the assessment process means that participants need to be enabled to achieve comparable assessment results under comparable conditions. Here, irregularities are unacceptable, but could occur if participants are not treated equally or if environmental conditions vary greatly. Regarding the technical preconditions, this means that the infrastructure and the equipment used by each participant must be technically stable. The same applies to the overall organisational (and logistical) processes: They should not deviate from case to case.

Third, we take Krebs' definition of objectivity, which stipulates "the independence of the measured results from the examiners" (Krebs 2004, p. 30, own translation),[3] and extend it as follows:

> **Objectivity of digital assessment procedures** is understood as the independence of the assessment results from the participants' technical equipment, the environmental conditions they are subject to, and the persons invigilating the assessment.

As the fourth and last quality criterion, practicality is required. For Compos (2020), "practicality in assessments means that the test is easy to design, easy to administer and easy to score. (..). It is economical to deliver. (..) The layout should be easy to follow and understand" (Compos, 2020, p. 1). Regarding the abovementioned fields of action, we arrive at the following definition:

> **Practicality of digital assessments** is understood as the facility with which a digital assessment can be conducted by participants, examiners, and other personnel. For it to be achieved, conditions must be controlled by applying measures in the legal, technical, didactical, and organisational fields of action.

1.2 Academic Integrity: Current Findings of Misconduct in Digital Remote Assessments During the Coronavirus Pandemic

The outbreak of the coronavirus pandemic in early 2020 and the subsequent lockdowns forced HEIs worldwide to massively reduce on-campus capacities or entirely switch over to remote teaching and assessment. As this momentous shift had to be performed with little or no time for preparation, the result was the introduction of what Hodges et al. (2020) have referred to as emergency

[3] "Im Zusammenhang mit Prüfungen wird unter Objektivität die Unabhängigkeit der Prüfungsergebnisse von den Untersuchern verstanden" (Krebs 2004, p. 30).

remote teaching (ERT), a scenario which stands in marked contrast to meticulously planned and implemented online teaching and learning. Following Hodges et al. (2020), we refer to digital remote assessments adopted on an ad-hoc basis as emergency remote assessments (ERA).

If we want to evaluate the quality of potential measures concerning digital remote assessments across all fields of action, we must first consider possible negative effects associated with ERAs. Two recent empirical studies of *academic misconduct* during the coronavirus pandemic are particularly instructive in this respect: Bilen and Matros (2020) present ways to technically detect cheating in digital remote assessments, while Janke et al. (2021) have investigated whether academic misconduct occurred more often in the summer semester of 2020 compared to pre-pandemic levels by questioning 1608 German students about their behaviour. As a starting point, we adopt Janke et al.'s definition of

> **academic dishonesty and misconduct** "as the intentional breaking of academic rules for personal gain" (Janke et al., 2021, pp. 3–4).

Second, we specify the opposite of academic dishonesty by following Wikipedia Contributors' definition whereby

> **academic integrity** can be defined as a moral code and an ethical policy that "supports the enactment of educational values through behaviours such as the avoidance of cheating (..), as well as the maintenance of academic standards (..)" (Wikipedia Contributors, 2021).

Besides moral and ethical aspects, academic dishonesty may have serious legal consequences, since assessment participants who engage in academic misconduct by cheating run the risk of being expelled. However, the fact that ERA had to be implemented immediately meant that it was hardly possible to take sufficient preventive measures in all fields of action. Since students were aware of the provisional status of digital ad-hoc solutions, it was inevitable that some would classify the risk of fraudulent behaviour being detected as low or non-existent. In so doing, however, they underestimated the technical possibilities afforded by data analysis carried out after the completion of digital remote assessments. Bilen and Matros (2020) describe several technology-based ways to detect cheating retroactively. Like other digital services, digital remote assessments leave traces in the assessment system's log files. This enables the identification of participants who, for example, answered questions correctly but implausibly fast ("in under thirty seconds per question", p. 11) or who submitted the very same *incorrect* answers as their peers, both of which is highly indicative of illicit cooperation.

But if students know about the legal risks, why do they engage in cheating anyway? With an eye to current research on academic integrity in higher education (Daumiller & Janke, 2019; Douglas et al., 2015), Janke et al. (2021) conclude that students "engage more frequently in (..) behavior" like cheating under certain circumstances, especially "when they feel that it will go unnoticed or at least unpunished by authorities such as their instructors" (p. 4). The authors of the study also suggest that "students may feel less accountable and less likely to be punished during online exams" when instructors are unable "to monitor the physical testing environment" (p. 5). Some additional information may be helpful here to classify these findings: Janke et al. (2021) questioned 1608 German students (68.7% female, 0.8% diverse), out of which 385 (78% female) reported that they had participated in "online exams" (which, for the purposes of this article, is translated as "digital remote assessments"). Another 349 students reported that they had participated in both "online exams" ("digital remote assessments") and "on-site exams" ("on-campus assessments"). Looking at the data (pp. 27–28), we can see Janke et al. (2021) used a seven-point scale (from 1 = "never" to 7 = "very frequently"). Amongst the 385 students who had participated in "online exams" ("digital remote assessments") only, the average frequency of cheating was at $M = 2.57$ ($SD = 1.81$), whereas the average for "on-site exams" ("on-campus assessments") was considerably lower ($M = 1.33$, $SD = 0.81$), as reported by 874 students. The effect of possibly biased samples seems to be low, as the average frequency of general academic dishonesty is comparable in both groups (on-site exams: $M = 1.43$, $SD = 0.42$ and online exams $M = 1.51$, $SD = 0.49$). Although their sample is very likely not representative, based on the available data we second the authors' interpretation that there is a strong indication that students "cheated more frequently in online than in on-site exams" (p. 2) during the summer semester 2020, thus putting academic integrity and the legal reliability of digital remote assessments at risk.

2 Comparison and Implementation

This section of the article aims at determining the fundamental differences between digital on-campus assessments and digital remote assessments. Bearing in mind our previous discussion where we emphasized that digital assessments in general are both limited by *conditions* and facilitated by *measures* within different fields of action, we now take the three research questions listed in the introduction to this article as our basic framework: In what follows, we will outline the different conditions that obtain in on-campus and remote environments, identify possible measures for maintaining control under these conditions, and evaluate the *effects* of said measures on the validity and reliability of digital remote assessments.

2.1 Digital On-Campus and Digital Remote Assessments Conditions, Measures, and Procedures

Generally speaking, assessments must be valid, reliable, and objective. The assessment process must also be practical for examiners, or in other words, "(..) permit inferences to be drawn concerning the skills of examinees" (Case & Swanson, 2002, p. 10).

As mentioned above, we assume that on-campus assessments share certain similarities with laboratory experiments. Conducted under supervision in a fixed setting, both allow for the establishment of stable and controlled environmental conditions, which in turn facilitate quality assurance. Experiments in laboratories must be internally valid in the sense that they must measure what they are intended to measure, and the results of lab experiments must be reproducible and reliable, thus requiring controlled conditions as well as strict abidance by designated procedures and workflows. The same holds true for on-campus assessments: Here, the knowledge and skills of students are measured in lecture halls, which facilitate invigilation and environmental control. Assessment procedures and conditions such as rules of admission, exam duration, and permitted resources are fixed and apply to every participant, thereby ensuring equal treatment of assessment participants as well as the comparability, validity, and reliability of assessment results. Digital on-campus assessments go one step further in that they are often conducted in computer rooms or digital examination halls (CRDEH).[4] Within a CRDEH, additional invigilation can be implemented, as the assessment process is digitally recorded for every participant.

If we apply this perspective to remote assessments, we realise that they share more similarities with field experiments which are conducted in natural environments. Here, control over procedures and environmental conditions is necessarily reduced – if possible at all, supervising them would require far too much effort

[4] Two examples of this are the "Testcenter" at the University of Bremen (http://www.eassessment.uni-bremen.de/testcenter.php) and the "E-Examination Center" at FUB (https://www.e-examinations.fu-berlin.de/pruefungsraum/index.html).

to be practicable. This is also the case with remote assessments: Environmental conditions are difficult or impossible to control, and *physical* supervision is simply not a viable option in this scenario. Moreover, while some procedures and conditions of digital remote assessment such as assessment duration can be subjected to digital control, provisions like a ban on auxiliary resources is much harder to enforce.

Likewise, with quality assurance being much more complex to implement in digital remote assessments, the legal reliability of remote assessments per se is at stake. Therefore, in order to obtain objectively reproducible and reliable results within digital remote assessments, meticulous planning of procedures and workflows would have been essential. In the actual event, however, the intense time pressure created by the coronavirus outbreak made such thoroughness very hard to achieve.

2.1.1 On-Campus Environments and Measures

In Germany, a “Klausur” constitutes a summative examination that *must* be conducted under invigilation to carry legal weight. Within on-campus environments, legal compliance can be assured in a fairly uncomplicated way, as the exam and study regulations are, since time immemorial, designed for on-campus exams. In this context, issues of legal compliance are mainly focused on the principle of equal treatment of participants. Due to their technical peculiarities, additional normative regulation may be required for digital on-campus assessments in Germany (Niehues et al., 2014), where they are viewed as a type of performance assessment that differs fundamentally from a written assessment. Equal treatment of participants is ensured by the HEI through measures such as physical admission control and supervision during the assessment, as well as the provision of rooms (organisation and logistics), exam devices and technical infrastructure (technology), and carefully designed exam questions (didactics and psychometrics).

Procedures, Conditions, and Measures

Traditional paper-and-pencil based on-campus assessments mostly follow strict procedures devised from a *didactical* view. In the course of didactical preparation before the semester, intended learning outcomes (ILOs) for teaching and learning are defined. From these ILOs, exam questions are then derived quite easily, an approach that has been described with the didactical concept known as “constructive alignment” (Biggs & Tang, 2011). With regard to digital on-campus assessments, it is advisable to take a more macroscopic view, as the setting up of CRDEHs tends to be complex and cost-intensive, and is therefore unlikely to be achieved solely by lecturers in faculties without access to the expertise and resources provided by technical departments or IT units. An essential part of the preparation of digital on-campus assessments is the implementation of technically reliable and scalable solutions, but not all requirements associated with digital on-campus assessments can be met by just throwing technology at them. Instead, we must first determine which conditions for the successful deployment of digital on-campus assessments cannot be satisfied within their respective field of action, and will therefore have to be supported by suitable measures in other fields of action in order to achieve reliability and scalability for large cohorts. For digital on-campus assessments, this means that organisational preparation involves defining standardised processes and workflows, and assigning specific responsibilities (Table 2):

While paper-and-pencil exams usually take place in lecture halls, digital on-campus assessments are often conducted in CRDEHs. In a nutshell, the advantage of such a setting lies in the fact that all *technical* systems in the facility (including lighting, technical devices, and the network infrastructure) are under the control of the examiners (Schulz, 2017), and can hence be prepared for different scenarios like classical assessments or resource-rich assessments (Halbherr et al., 2019) which allow for the measuring of competencies by using realistic or authentic assessment designs. This ensures technical stability and reliability, and thus also legal reliability, as the principle of equal treatment is translated here into appropriate technical measures and specifications. The technical preparation of assessments is sometimes carried out by specialised IT units, who at times also migrate assessments into the digital assessment system and activate features like randomisation of questions, and/or administer the CRDEHs.

Table 2 Examples for defining standardised processes and identifying the relevant fields of action

Process category	Questions	Field of action
Preparation of exams	How does the technical implementation of the assessment into the digital assessment system (DAS) work?	Technology
	Who is responsible for the technical implementation?	Organisation
Conducting of exams	How does the admission to the assessment room work?	Logistics
	Who is responsible for admission control and technical supervision during the assessment?	Logistics
	Who is didactically supervising the assessment?	Organisation
	What needs to be done if technical issues occur?	Technology

A downside of CRDEHs is their limited logistical capacity, which makes thorough *logistical* planning indispensable. Large cohorts may have to be split up into more manageable sub-cohorts, which are assessed in series rather than in parallel. This in turn has an impact on *didactical* preparation, as for *legal* reasons, each serially assessed sub-cohort must be given a unique set of exam tasks to rule out the passing on of information between sub-cohorts, which would violate the principle of equal treatment.

2.1.2 Remote Environments and Measures

As the coronavirus pandemic hit universities, practical measures that allowed educational operations to be maintained had to be found rapidly. For remote teaching, the use of video conferencing tools was an obvious choice. Finding similar solutions for assessments, by contrast, was much less straightforward. Some HEIs decided to retain on-campus assessments with drastically reduced capacities, while many others instantly began designing digital remote assessment scenarios. Intense time pressure notwithstanding, even HEIs who could not draw on previous experience with digital assessments soon became aware of the macroscopic relationships and dependencies that characterise digital assessments beyond issues of didactical preparation. A multitude of questions were in urgent need of answers: Which scalable remote assessment scenarios could be implemented and rolled out quickly? How much time would it take to set them up? Which fundamental conditions in the abovementioned fields of action had to be satisfied to stay as close to the concept of on-campus "Klausuren" as possible (Fischer & Dieterich, 2020)?

Conditions, Procedures, and Measures

In contrast to dedicated on-campus facilities, students' private residences do not permit physical supervision by the HEI. Moreover, examiners have no control over the environmental conditions in students' workspaces (e.g., noise) or the technical properties of their devices (system performance, configuration, and stability, available bandwidth, etc.). In effect, we know next to nothing about the setting in which the assessments take place.

Taking the mechanisms of control and supervision that are used in on-campus assessments as a starting point, we can identify several essential conditions that must be met within digital remote assessment scenarios: (1.) The *identity* of the participants must be verified before and during the exam; (2.) *cheating* within assessments by exchanging answers amongst participants or using unauthorised resources (e.g., searching the internet) must be mitigated; (3.) the assessment must be capable of *validly*, *reliably,* and *objectively* measuring the skills and knowledge of the participants; and (4.) the entire assessment process must be *practically* implementable and easily manageable for examiners with reasonable effort.

Technically, remote "Klausuren" that meet the legal requirement for supervision could be implemented with the help of digital remote proctoring, a form of oversight which uses a webcam to monitor examinees during the exam. Proctoring solutions range from transmitting video of the examinees to a live examiner in the manner of a video conference over monitoring the examinees and their screens to using artificial intelligence instead of a live examiner to detect unusual behaviour by analysing eye movements and keystrokes (White, 2020).

Hylton et al. (2016) have argued that digital remote proctoring can reduce the probability of cheating, making it one of the first technical choices to maintain the legal reliability of assessments within remote environments. But although digital remote proctoring theoretically enables supervised "Klausuren" in a remote setting, its introduction in Germany and other countries is impeded by serious *legal* obstacles in the field of privacy protection, as supervising examinees at home may in practice be tantamount to monitoring their private living quarters. The legal hurdles for intruding into the privacy of citizens are high in Germany, and the country's federal system means that the individual states ("Bundesländer") all have their own legislation in place. Generally speaking, Germany's federal states have proved highly reluctant to introduce normative regulations that would allow for digital remote proctoring.

Where digital remote proctoring is difficult or impossible to implement, measures must be devised that can compensate for a lack of invigilation. *Didactical* measures to mitigate unallowed cooperation between examinees include creating tasks targeting higher levels of *cognitive taxonomies* and designing more than one set of questions aimed at the same learning outcomes. With this approach, students are required to evaluate and compare cases or even create new hypotheses, thus effectively having to apply their knowledge and competency to new situations. This form of skill training and measuring is in line with the intentions behind the Bologna reform (EMHE, 1999), as it encourages thinking outside the box and prepares students for future challenges. However, in some degree programmes such as the natural sciences, the setting of tasks targeting higher levels of cognitive taxonomies only makes sense if students have already grasped the basics and are capable of reproducing them—which, in terms of cognitive complexity, is a fairly low-level task. Yet if assessments target lower levels of cognitive complexity, this increases the risk of cheating, as basic facts and information can easily be found on the internet. Well-known examples of cognitive taxonomies are those by Bloom et al. (1956) and Anderson and Krathwohl (2001). The digital revision by Mayer et al. (2009) entitled

"Computer Supported Evaluation of Learning Goals" (CELG), in particular, is frequently cited due to its simplicity and practicality, as it already includes suggestions for task type assignments aligned to certain levels of cognitive taxonomies.[5]

Where the targeting of higher levels of cognitive taxonomies is impracticable, creating multiple sets of questions targeting the same learning outcomes may be an alternative. The disadvantage of this approach is that lecturers will likely need more time to prepare assessments, and the amount of time required to technically implement different sets of questions is also bound to increase.

The use of genuine digital assessment systems (DASs) like the commercial LPLUS TestStudio[6] or the open-source software Dynexite[7] can be subsumed under *technical* considerations, as these systems offer detailed assessment process protocols which permit the a posteriori detection of fraudulent activities via assessment analytics. DASs allow for randomisation of task order, randomisation of distractors within closed response tasks, and randomised selection of equally difficult tasks, thus increasing the time examinees need to cooperate illicitly, while also giving examiners control over exam duration. Additionally, DASs can be combined with a secure digital exam environment that students must install on their devices prior to the exam to be able to participate – the Safe Exam Browser (SEB)[8] developed by ETH Zurich is an example of such a tool. Available for multiple operating systems, it establishes a controlled technical environment on students' exam devices which blocks access to unauthorised material and prevents the use of copy and paste. However, as it is not a proctoring tool, it cannot trace the use of other devices or material.

The use of video conferencing tools prior to and after the exam may help to reduce the probability of academic misconduct. Since the use of video conferencing technologies for lectures and seminars is legally allowed, obligatory *procedures* can be implemented which require examinees to visit a virtual conference room before the exam as a digital equivalent to physical admission requirements. At this stage of the process, identity checks can be carried out. Procedures such as these can also be combined in a mixed assessment, where examinees are obliged to return to the virtual conference room immediately after the written exam for an oral examination.

[5] During the coronavirus pandemic, assessments and tasks targeting higher levels of cognitive taxonomies have often been confused with "open-book" tasks. Open-book tasks simply allow students to use certain resources—the label says nothing about the level of cognitive competency examinees must have reached to complete them.

[6] https://lplus.de/home.

[7] https://dynexite.rwth-aachen.de.

[8] https://safeexambrowser.org/about_overview_en.html.

This scenario is likely to have a deterrent effect regarding fraudulent activities, as the risk of being found out during the oral examination is significant.

Another very simple measure derived from findings in the field of social psychology should also be considered: As Shu et al. (2010) have demonstrated, priming participants' awareness of honesty standards by making them "read or sign an honour code significantly reduce[s] unethical behavior" (p. 330). Within DASs, this is easy to implement: Agreeing to a code of honour can be made a precondition to participation in the assessment.

2.2 Evaluating Measures for Quality Assurance Within Digital Remote Assessments

In a final step, we will evaluate some of the previously discussed measures for digital remote assessments by using the quality criteria defined earlier. Here, three things are worth mentioning: First, not all quality assurance aspects may be applicable to all measures; second, while we have touched upon several legal aspects from a specifically *German* perspective, we are not lawyers – reliable legal counsel should be sought, and legal departments should be involved in the setting up of digital remote assessments from the very outset; and third, as already mentioned, the field of action entitled "Organisation and Logistics" is originally derived from on-campus scenarios. In the case of digital remote assessments, this field intersects with all other fields of action, as it has a profound impact on the entire digital assessment process. Hence, sampled measures are evaluated for the first three fields of action only (law and regulations, technology, didactics and psychometrics).

2.2.1 Measure: Using a Human Examiner for Digital Proctoring and/or Identity Checks Prior to the Exam

Field of Action: Laws and Regulations

Legal validity requires digital proctoring to be in accordance with the applicable laws. This must be established first, as privacy protection is concerned. Legal reliability of results is high due to supervision as per exam regulations. Legal objectivity is linked to didactical objectivity and may depend on the type of digital proctoring (in-house human examiner, external human examiner, AI examiner). Legal practicality is likely low, as privacy protection is concerned. Changing laws and regulations takes time.

Field of Action: Technology

Technical validity is likely high, as digital proctoring systems are designed to be fail safe. Technical reliability may be reduced since the proctoring tool consumes technical resources on the examinees' devices. Technical objectivity is likely high, as digital proctoring is carried out for all participants. Technical practicality is low if commercial proctoring providers cannot be used for legal reasons.

Field of Action: Didactics and Psychometrics (DAP)

Didactical and psychometrical validity is likely high, as digital proctoring simulates supervision in on-campus assessments. For the same reason, didactical and psychometrical reliability is also likely to be high, as is didactical and psychometrical objectivity. Didactical and psychometrical practicability is linked to technical practicality, and thus likely low if commercial proctoring providers cannot be used for legal reasons.

2.2.2 Measure: Conducting Oral Exams Immediately After the Written Exam

Field of Action: Laws and Regulations

Legal validity may have to be established first, as this measure is likely to require changes to existing exam regulations. Once legal validity is established, legal reliability is high. Legal objectivity is likely connected to didactical objectivity. Legal practicality may be low, as legal validity must be established first.

Field of Action: Technology

Technical validity is likely not applicable in this case. Technical reliability might be reduced if examinees must switch between the digital assessments system (DAS) and a video conferencing tool. Technical objectivity is likely not applicable in this case. Technical practicality might be reduced if examinees must switch between the DAS and a video conferencing tool.

Field of Action: Didactics and Psychometrics (DAP)

Didactical and psychometrical validity, reliability, objectivity, and practicality are likely high, as the tasks have already been set in the written exam immediately prior to the oral examination; only additional time is required.

2.2.3 Measure: Using Digital Assessment Systems (DASs)

Field of Action: Laws and Regulations

Legal validity of using dedicated assessment systems must be created by documenting what kind of personal data is stored and where (BSI baseline protection principle). When legal validity has been established, legal reliability of using dedicated DASs is likely high. Legal objectivity is connected to didactical objectivity. State-of-the-art DASs do not influence didactical measuring, which means that legal objectivity should be high. Legal practicality is low if DASs are not already in use, as IT security concepts in line with BSI baseline protection principles must first be established.

Field of Action: Technology

Technical validity is likely high, as genuine DASs are designed to be fail safe. Technical reliability is high, as all participants must install the same tool and configuration. Technical objectivity is high because the DAS must be used by all participants. Technical practicality is high, as DASs allow for standardised workflows.

Field of Action: Didactics and Psychometrics (DAP)

DAP validity and reliability is likely high, provided that known ILOs are measured. DAP objectivity is likely high because the DAS must be used by all participants. DAP practicality is likely high if a DAS is already in use.

2.2.4 Measure: Installing and Using Safe Digital Exam Environments on Students' Devices

Field of Action: Laws and Regulations

Legal validity of using safe digital exam environments needs to be created by documenting what consequences are associated with installing such a tool. When legal validity has been established, legal reliability is high, as all participants must install the same tool and configuration. Legal objectivity is likely connected to didactical objectivity. State-of-the-art digital exam environments do not influence didactical measuring, which means that legal objectivity should be high. Practicality is low if safe digital exam environments are not already in use, as IT security concepts in line with BSI baseline protection principles must first be established.

Field of Action: Technology

Technical validity is likely not applicable for this measure. Technical reliability is likely high, as the tool creates a standardised environment on the examinees' devices. Technical objectivity is likely not applicable for this measure. Technical practicality might be reduced, as tutorials must be created and support for the installation process provided.

Field of Action: Didactics and Psychometrics (DAP)
DAP validity and reliability is likely high, provided that the use of the tool has been explained and practised before. DAP objectivity is likely high, as the tool must be used by all participants. DAP practicality is likely connected to technical practicality. It might be reduced, as tutorials must be created and support for the installation process provided.

2.2.5 Measure: Randomisation of Task Order or Distractors Within Tasks

Field of Action: Laws and Regulations
Legal validity is high, as long as tasks do not build on results of other tasks that are randomly torn apart by the measure. Legal reliability is high, provided that legal validity is ensured. Legal objectivity is likely connected to didactical objectivity but is nonetheless expected to be high. Legal practicality is high, as this measure is a standard feature of dedicated assessment systems (DASs).

Field of Action: Technology
Technical validity is likely not applicable for this measure. Technical reliability is likely high, as this measure is a standard feature of DASs. Technical objectivity is likely not applicable for this measure. Technical practicality is likely high, as this measure is a standard feature of DASs.

Field of Action: Didactics and Psychometrics (DAP)
DAP validity and reliability is likely increased, as cheating by collaborating is made more difficult. DAP objectivity remains high. DAP practicality is connected to technical practicality and thus likely high, as this is a standard feature of DASs.

2.2.6 Measure: Random Selection of Tasks Within the Same Difficulty Level

Field of Action: Laws and Regulations
Legal validity is likely connected to didactical validity. A comparable difficulty level is required to ensure a level playing field. Legal reliability is high if legal validity is ensured. Legal objectivity is likely connected to didactical objectivity.

Legal objectivity and practicality are likely high, as this measure is a standard feature of DASs, which do not influence didactical measuring.

Field of Action: Technology

Technical validity is likely not applicable for this measure. Technical reliability is likely high, as this is a standard feature of DASs. Technical objectivity is likely not applicable. Technical practicality is likely high, as this measure is a standard feature of DASs.

Field of Action: Didactics and Psychometrics (DAP)

DAP validity and reliability is increased, as cheating by collaborating is made more difficult. DAP objectivity remains high. DAP practicality is likely connected to technical practicality and thus high, as this measure is a standard feature of DASs.

2.2.7 Measure: Using Assessment Analytics for Post-Exam Cheat Detection

Field of Action: Laws and Regulations

Legal validity is highly dependent on data protection. The use of assessment analytics for post-exam cheat detection likely requires legal regulation. Legal reliability is high if legal validity is ensured. Legal objectivity is high provided that all required data is available. Legal practicality might be low, as legal validity must be established first.

Field of Action: Technology

Technical validity is likely not applicable for this measure. Technical reliability is likely high, as protocols are a standard feature of dedicated DASs. Technical objectivity is likely not applicable for this feature. Technical practicality is likely high, as this measure is a standard feature of DASs.

Field of Action: Didactics and Psychometrics (DAP)

DAP validity, reliability, and objectivity are likely not applicable for this measure. DAP practicality is likely connected to technical practicality and thus high, as this measure is a standard feature of DASs.

2.2.8 Measure: Using Tasks Targeting Higher Cognitive Taxonomy Levels

Field of Action: Laws and Regulations

Legal validity, reliability, objectivity, and practicality are likely connected to their didactical counterparts. As long as examinees are not overwhelmed by the assigned tasks, the legal quality criteria are likely fulfilled.

Field of Action: Technology

Technical validity, reliability, objectivity, and practicality are likely not applicable, as this is mainly a didactical measure.

Field of Action: Didactics and Psychometrics (DAP)

DAP validity and reliability is likely increased, as cheating by collaborating or searching the internet is made more difficult. DAP objectivity likely remains high. DAP practicality depends on how much knowledge and competencies a student has acquired. Students in the early stages of their studies must first learn the basics, which may complicate the application of this measure.

2.2.9 Measure: Using a Code of Honour for Priming

Field of Action: Laws and Regulations

Legal validity is likely not concerned provided that the code of honour does not discriminate against anyone. Legal reliability is likely high if legal validity is ensured. Legal objectivity is likely not concerned. Legal practicality is high, as the measure is easy to implement via the DAS.

Field of Action: Technology

Technical validity is likely not applicable for this measure. Technical reliability is likely high, as customising the entry procedure to exams is a standard feature of DASs. Technical objectivity is likely not applicable for this measure. Technical practicality is likely high, as this measure is a standard feature of DASs.

Field of Action: Didactics and Psychometrics (DAP)

DAP validity and reliability is bound to increase given that academic misconduct is likely to be reduced (Shu et al., 2010). DAP objectivity is likely not applicable for this measure. DAP practicality is connected to technical practicality and thus high if dedicated DASs are used, as this is one of their standard features.

3 Conclusion and Outlook

Digital remote assessments represent a major challenge for all higher education institutions (HEIs), especially those accustomed to conducting on-campus assessments as paper-and-pencil exams. Yet, the coronavirus pandemic forced all HEIs to instantly design emergency remote assessment (ERA) scenarios and implement them on an unprecedented scale. What this massive roll-out has demonstrated is that while digital remote assessments are manageable and highly scalable, they resemble field experiments more than laboratory experiments in that they take place under largely uncontrolled conditions and hence require meticulous planning and the taking of appropriate measures for academic integrity to be maintained.

As we have shown, effective quality assurance in the field of digital remote assessments can only be achieved by rethinking and extending the concept of *assessment literacy* – failure to do so means that the probability of academic misconduct is bound to increase.

This article is therefore intended as a first proposal for a *literacy of practice for digital remote assessments.* With regard to digital assessments in general, we have proposed a model that considers four fields of action, namely (1.) law and regulations; (2.) technology; (3.) didactics and psychometrics; and (4.) organisation and logistics. The latter constitutes an overarching field of action, as it comprises, controls, and monitors measures to be implemented in all other fields. Taking legal reliability as our point of departure, we have also reconsidered didactical quality criteria and extended them to be applicable to measures in all fields of action.

Even without a pandemic looming in the background, digital remote assessments have a number of distinct advantages: Travel times to and from the HEI are no longer an issue, they are family-friendly, and ease the logistical burden on students from abroad. From an economical point of view, the mass scalability of digital remote assessments is highly attractive to the steering committees of HEIs, as it has the potential to significantly reduce the logistical capacities required for on-campus assessments.

However, in many German federal states, the sustainable implementation of digital remote assessments will depend on changes to the existing legal framework. Given that measures aimed at remote proctoring, for instance, have a direct impact on fundamental freedoms such as the students' right to privacy, this is likely to be a difficult and protracted process.

This paper was funded by the Federal Ministry of Education and Research and the State of Berlin through the Excellence Strategy of the German Government Berlin University Alliance.

References

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

Biggs, J. B., & Tang, C. K. C. (2011). *Teaching for quality learning at university: What the student does*. McGraw-Hill.

Bilen, E., & Matros, A. (2020). Online cheating amid COVID-19. *Journal of Economic Behaviour & Organization, 182,* 196–211. https://doi.org/10.1016/j.jebo.2020.12.004.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* David McKay Company.

Case, S. M., & Swanson, D. B. (2002). *Constructing written test questions for the basic and clinical sciences – Third edition, national board of medical examiners*. Researchgate. https://www.researchgate.net/publication/242759434_Constructing_Written_Test_Questions_For_the_Basic_and_Clinical_Sciences.

Compos, J. (2020). *Reliability, validity and practicality*. EnglishPost. https://englishpost.org/reliability-validity-and-practicality/.

Daumiller, M., & Janke, S. (2019). The impact of performance goals on cheating depends on how performance is evaluated. *AERA Open, 5*(4), 1–10. https://doi.org/10.1177/2332858419894276.

Douglas, D. M., Paullet, K., & Chawdhry, A. (2015). Student perspectives of cheating online classes. *Issues in Information Systems, 16*(4), 215–223. https://doi.org/10.48009/4_iis_2015_215-223.

EMHE – European Ministers in charge of Higher Education. (1999). *The Bologna Declaration of 19 June 1999: Joint declaration of the European Ministers of Education*. https://www.eurashe.eu/library/bologna_1999_bologna-declaration-pdf/.

Fischer, E., & Dieterich, P. (2020). Prüfungsrecht in Zeiten der Coronavirus-Pandemie. *NVwZ - Neue Zeitschrift Für Verwaltungsrecht, 10,* 657–666. https://beck-online.beck.de/?vpath=bibdata/zeits/NVWZ/2020/cont/NVWZ.2020.657.1.htm.

Halbherr, T., Lehner, H., & Kapur, M. (2019). Resource-rich versus resource-poor assessment in introductory computer science and its implications on models of cognition: An in-class experimental study. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*.

Hylton, K., Levy, Y., & Dringus, L. P. (2016). Utilizing webcam-based proctoring to deter misconduct in online exams. *Computers & Education, 92–93,* 53–63. https://doi.org/10.1016/j.compedu.2015.10.002.

ICAI—International Center for Academic Integrity. (2021). *The fundamental values of academic integrity: Third Edition*. ICAI. https://academicintegrity.org/images/pdfs/20019_ICAI-Fundamental-Values_R12.pdf.

Janke, S., Rudert, S. C., Petersen, Ä., Fritz, T. W., & Daumiller, M. (2021). Cheating in the wake of COVID-19: How dangerous is Ad-hoc online testing for academic integrity, *Preprint*. https://doi.org/10.31234/osf.io/6xmzh.

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). *The difference between emergency remote teaching and online learning*. educause review. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning.

Krebs, R. (2004). *Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung*. Universität Bern, Institut für Medizinische Lehre. https://www.iml.unibe.ch/attachment/7/download/mc_anleitung.pdf.

Krebs, R. (2008). *Multiple Choice Fragen? Ja, aber richtig*. Universität Bern. https://docplayer.org/docview/37/17842354/#file=/storage/37/17842354/17842354.pdf.

Mayer, H. O., Hertnagel, J., & Weber, H. (2009). *Lernzielüberprüfung im eLearning*. Oldenbourg Verlag.

Munzer, S. (1972). *Legal validity*. Springer Verlag.https://doi.org/10.1007/978-94-011-9271-2.

Niehues, N., Fischer, E., & Jeremias, C. (2014). *Prüfungsrecht* (6., neubearbeitete Auflage). *NJW-Praxis: Vol. 27, 2*. Beck.

Schulz, A. (2017). *E-Assessment-Center im Vergleich – Voraussetzungen und Kosten für die Einrichtung verschiedener E-Assessment-Center im Vergleich.* Technische Universität Dresden. https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-224532.

Schulz, A. (2021). Handlungsfelder digitaler Prüfungen: Recht, Technik, Didaktik und Organisation. In M. Bandtel, M. Baume, E. Brinkmann, S. Bedenlier. J. Budde, B. Eugster, A. Ghoneim, T. Halbherr, M. Persike, F. Rampelt, G. Reinmann, Z. Sari, & A. Schulz (Eds.), *Digitale Prüfungen in der Hochschule* (pp. 18–24). Hochschulforum Digitalisierung. https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_Whitepaper_Digitale_Pruefungen_Hochschule.pdf.

Shu, L. L., Gino, F., & Bazerman, M. H. (2010). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin, 37*(3), 330–349. https://doi.org/10.1177/0146167211398138.

Webb, N. (2002). *Assessment Literacy in a Standards-Based Education Setting, Paper Presented at the American Educational Research Association Annual Meeting*. WCER Working Paper No. 2002-4. https://www.researchgate.net/publication/251923627.

White, N. (2020). *'Creepy' universities are recording students and monitoring their eye movements to stop them cheating as they're forced to take tests in their bedrooms because of Covid-19*. Daily Mail Australia. https://www.dailymail.co.uk/news/article-8243637/Creepy-software-used-stop-university-students-cheating-online-exams-amid-coronavirus.html.

Wikipedia Contributors. (2021, May 16). *Academic integrity*. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Academic_integrity&oldid=1023446467.

Wollersheim, H.-W., & Pengel, N. (2016). Von der Kunst des Prüfens – Assessment Literacy. *HDS.Journal – Perspektiven Guter Lehre,* (2), 14–32. http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-212985.

Alexander Schulz is the coordinator of Freie Universität Berlin's (FUB) E-Learning and E-Examinations Unit at the Center for Digital Systems. The focus of his work is in the field of large-scale digital learning and assessment technologies. He is project leader for FUB's part-projects within the two large scale German joint AI-projects 'tech4compKI' and 'IMPACT', funded by the Federal Ministry of Education and Research. He is coordinator of the 'E-Assessment Alliance' project within the 'Berlin University Alliance' (BUA) as part of the German Universities Excellence Initiative.

Online Assessment from a Broader Perspective with Practical Applications

Hale Ilgaz and Denizer Yildirim

Abstract

Assessment and evaluation can be one the most challenging aspects of online learning processes. With the current global pandemic in particular, the wave of digitization has become a more prominent issue for all level educational institutions. Online assessment is part of the instructional design process and cannot be considered independently of this process. The type of content, learning objectives, and expected outcomes play an important role in determining these assessment methods. In this context, the main aim of this chapter is to explain online assessment and evaluation approaches and to present sample applications for different content areas and which tools and techniques can be better for instructors during the designing of the assessment process. Within this aim, assessment tools and strategies have been presented for the fields of education science, medical education, and legal education, with specific usage cases.

Keywords

Assessment design • Moodle LMS • Online assessment • Practices

H. Ilgaz (✉) · D. Yildirim
The Faculty of Open and Distance Education, Ankara University, Ankara, Turkey
e-mail: hilgaz@ankara.edu.tr

D. Yildirim
e-mail: dyildirim@ankara.edu.tr

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_8

1 Introduction

Assessment is an important component of teaching and learning for all levels of education. For online learning, assessment is also one of the most challenging parts of the learning process. The recent global Coronavirus pandemic has hastened a wave of digitization for all levels of educational institutions. While even in traditional face-to-face learning environments assessment types and techniques play a critical role, it is more challenging in online learning environments. From a general perspective, the main goal of assessment is the process of gathering and interpreting data by instructors for grading and for tracking learning (Arnold, 2014; Boud, 2000). The assessment process comprises several dimensions that affect the entire process, including finding the most efficient method, security, and reliability.

Online assessment is part of the instructional design process and cannot be considered independently from this process. The type of content, learning objectives, and expected outcomes play an important role in determining these assessment methods. The extended use of information and communication technologies and learning management systems (LMS) have necessitated integrated tools for assessment activities. It has been observed that traditional assessment methods such as multiple-choice tests, true/false questions, or drag-drop questions are preferred in systems with many assessment tools (Lourdes et al., 2017; Stödberg, 2012). The assessment process consists not only of grading students, but also the evaluation of the entire process, including the instructor and the program itself. Therefore, there are no one-size-fits-all solutions in an effective online assessment process. For an effective assessment process, instructors should plan to combine both formative and summative assessment approaches. Related to these approaches are several tools, each of which has advantages and disadvantages. Combining various assessment tools can compensate for their individual disadvantages.

In this chapter, assessment strategies and sample cases that can be effective for different fields are presented. The presented cases serve as examples of the methods that can be included in an online course.

2 Theoretical Approaches

Assessment and instructional design cannot be considered separately from each other. Instructional design models such as ADDIE, ASSURE, Dick & Carey Model, and the Kemp Design Model are widely used in the design of e-learning processes. Although named differently, each model contains assessment stages for both the process as a whole and for learning outcomes (Edmonds, Branch & Mukherjee, 1994). Therefore, independently determining the techniques to be used in the assessment process from the content and teaching method leads to problems in the implementation process.

There are two main approaches for online assessment: formative and summative. Formative assessment can be defined as "assessment for learning", which evaluates student learning and allows for the next step to be planned. Summative assessment can be defined as "assessment of learning", and systematically evaluates student achievement, mainly at the end of the semester (Harlen & James 1997). Despite appearing different from each other, the two assessment approaches share a relationship, and their combination can provide a sound and reliable assessment process for online learning environments (Fig. 1).

The assessment approach widely used in traditional learning environments can be defined as a summative approach to assessment. In general, it is based on obtaining a grade, usually consisting of the average scores of the final exam and other exams related to the course. Formative assessment on the other hand is a feedback-based process that aims to improve learning within the process rather than at the end (Guerrero-Roldán & Noguera, 2018). Both of these assessment approaches are also widely used for online learning environments. Although

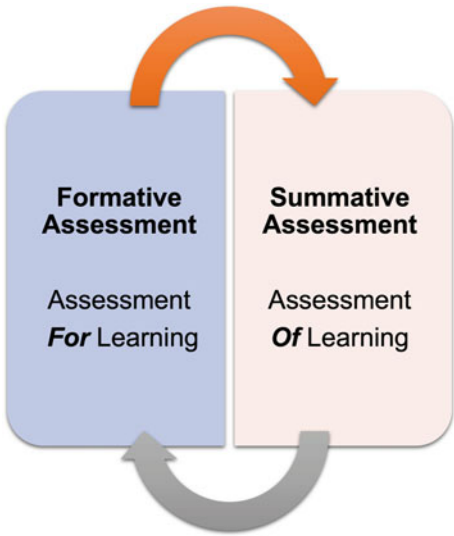


Fig. 1 Formative and summative assessment approaches

formative assessment seems to be one of the most effective approaches in the assessment process, situations do arise in which little to no improvement is seen in students, despite regular and clear feedback. It has been reported that these poor outcomes are due to deficiencies in the instructional system associated with the formative assessment (Sadler, 1989).

Online assessment is not an easy process and challenges arise when an effective online assessment process is desired. The two main challenges are instructor workload (Dunn et al., 2003; Sheridan, 2006) and cheating (Alruwais et al., 2018; Fask et al., 2014). When comparing the processes of online and traditional learning environments, online learning requires more effort than traditional learning environments. In parallel, a well-designed assessment process, regardless of online or face-to-face learning environments, is expected to set clear expectations with a reasonable workload and provide opportunities for students to learn, rehearse, practise, and receive feedback on their own. For this reason, it is important that the activities used in the assessment process match the expected learning outcomes (Ragupathi, 2020). While cheating is one important element that should be controlled regardless of whether learning is occurring online or face-to-face, it causes more concern in online learning environments due to the lack of control. Turning this disadvantage into an advantage can be done through an effective assessment design. Research has shown that open-book or open-web-exam environments in particular, with their open resource access, support deep learning (Myyry & Joutsenvirta, 2015).

To evaluate the process in online assessment, presented content should be diversified and the assessment process should be structured using different activities. Students who interact more with the learning activities and tasks in the system will be more advanced in terms of learning and proficiency. Therefore, it will not be effective to conduct formative assessment in a course design when students are not actively participating in the process and are inhabiting the role of the passive listeners (Spector & Park, 2017). For this reason, instructional design and assessment design are two components that interact and are directly interconnected. In this chapter, the authors will provide a holistic implication set for three different fields utilizing real-life cases. The presented frameworks will be applicable in other fields as well (Fig. 2).

Online Assessment Design in a Learning Management System

Learning management systems (LMS) contain a large number of content creation and assessment tools. Open-source LMSs, which are open to public development, allow for access to many plug-ins free of charge on the web. Systems such as Moodle, Canvas, ELMS Learning Network, Open Edx and Forma LMS are widely used by

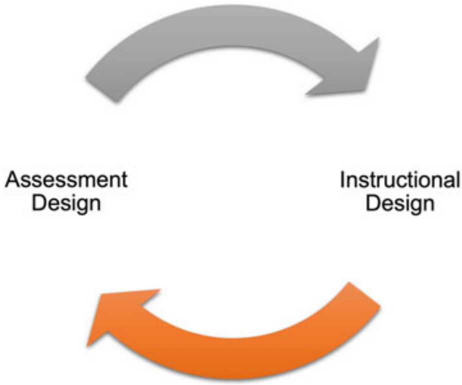


Fig. 2 The instructional design and assessment design relation

educational institutions. Although each LMS has advantages and disadvantages, Moodle is one of the most widely used open-source LMS. The most important aspect that makes Moodle stand out from other LMSs is that, due to its widespread use, many developers offer its plug-ins and share open-source solutions to problems encountered. Many commonly used LMSs share similar tools in terms of content creation and assessment activities. The tools used in the context of assessment activities and their areas of use are given in Table 1.

In general, the main assessment tools included in LMSs can be categorized in this way. All these tools can be used for courses from any field according to the course design and the instructors' planned objectives. However, using all available tools does not lead to an effective and efficient evaluation process, while selecting and using some of the most useful tools in supporting each other will provide more effective results in assessment design.

When it comes to conducting both theoretical and practical courses through online learning, the first option that comes to mind is to design courses in which theoretical knowledge is transferred or reinforced or where theoretical knowledge is reflected in practice. Therefore, from a general perspective, three different course designs can be presented, focused on content, discussion, or collaboration. These design options are presented in Table 2. Online assessment activities in the context of these course designs offer various opportunities to measure the acquisition of theoretical knowledge, the analysis and synthesis of theoretical knowledge, and the reflections of the achievements after application.

Table 2 summarizes the online assessment tools that can be functional for three different course designs focused on content, discussion, or collaboration. In this context, first consider a case in which all or some of the units of a course are designed

Table 1 Assessment tools, usage areas and functions

Assessment Tools	Aim	What can be done?
The reports for content view and uptime reports	• Assessment of behavioural engagement • The number of viewing learning resources • The duration of online study in the system	Giving additional points to students with high participation, warning students with low participation periodically via e-mail or in live virtual sessions
Peer assessment	• Assessment of the product (a report, assignment, or project output) from the perspective of other learners • Pre-assessment before the instructor's assessment • Supporting peer learning through the assessment process • Provide more frequent feedback through peers	Assessment of midterm exams through homework and peer assessment in the first stage of homework assessment, then instructor's evaluation in the second stage
Rubric assessment	• Standardizing the assessment process for the stakeholders who will make the assessment • Freeing the assessment process from subjective evaluations • Conducting a fair assessment	While using midterm exams through homework or project outputs, • Determining the tasks expected from the student in the homework or product, • Preparing a categorical or Likert grading table that is suitable for the determined processes, It is more effective if the first stage of the assessment is done by peer assessment and approved by the instructor in the second stage

(continued)

Table 1 (continued)

Assessment Tools	Aim	What can be done?
Rubrics enriched with learning analytics	Giving students points according to the frequency of their behaviour in the discussion activities	Scoring can be made according to all or a few of the criteria listed below: • The number of different people interacted with at discussion activities • The number of posts or conversations in discussion activities • The number of student comments on discussion activities
Activity completion reports	Whether the student has fulfilled the minimum tasks required in the discussion activities	Activity completion settings can be made according to all or a few of the criteria listed below: • Has the student shared at least one comment by a specific date? • Has the student responded to the comment at least once? • Did the student get an average score of at least 70 or more at the end of the peer review?
Examinations	The quiz add-in includes different types of questions: essay, multiple-choice, true/false, drag and drop (image or text), short answer, numerical, calculated, and matching type questions. Adding extensions allow for access to different question types such as chemical formulas, etc	Preparation of a question bank consists of different question types with different difficulty levels Providing the opportunity to repeat the test by giving the information whether the answer is correct or incorrect without giving the correct answer

Table 2 Course design and related tools

Course Design	Content Tools		Assessment tools		
	Synchronous	Asynchronous	Formative		Summative
			Analytics	assessment activity	
Content-oriented	Virtual session	Watching recordings Reading documents (student-content interaction)	View reports Embedded tests in interactive videos	If the number of students is large; Multiple-choice question (MCQ) tests, peer assessment If not; Assignments with the rubric, Essays	MCQ tests Assignments with rubric Essays
Discussion-oriented	Virtual sessions (LMS tools that will enable students to interact with teachers and teachers with students)	Following the discussion activity Commenting in the discussion activity (LMS tools that can be used for a group or all students to discuss a topic together)	Activity completion reports Rubrics enriched with learning analytics	Peer assessment	Essays, rubric evaluation
Collaboration-oriented	Virtual meetings with various online conferencing tools	Grouping in LMS (tools to be used within the LMS to facilitate the collaboration of a group of students)	Activity completion reports Rubrics enriched with learning analytics	Essays Peer assessment	Project presentations, Rubric evaluation Essay

with a content-oriented approach. In this design, LMS tools (for example, virtual classroom tools like Zoom, Collaborate, Teams etc.) are used to enable students to interact with the teacher simultaneously. Asynchronous study resources such as videos and presentations/documents are shared on the course page for students to gain knowledge.

In a discussion-oriented design, LMS tools can be used to enable instructor-student interaction, both synchronously (e.g.: virtual sessions) and asynchronously (e.g.: discussion board). Students can discuss a specific topic with both their teachers and peers through various web conferencing tools. On the other hand, it can be ensured that a group or all students can discuss a topic together by using asynchronous tools such as forums. Thus, it is possible to reinforce theoretical knowledge.

In a collaboration-oriented design, students can plan a project study in virtual sessions using various web conferencing tools. Students may be asked to make a video of a product or project process they have designed to observe whether they reflect the theoretical knowledge they have acquired. These videos can be evaluated by peers both within the group and in other groups.

In the next section, course designs for different educational fields and assessment activities related to these designs are presented.

3 Suggested Designs for Specific Fields

The global health pandemic has created a learning emergency necessitating the alteration of learning and assessment strategies (Adedoyin & Soykan, 2020). This emergency has brought about meaningful changes in the assessment approaches of higher education faculty members (Rapanta et al., 2020). For example, the use of certain assessment strategies, including essay exams, oral exams, project work, practical assessments, and portfolios, has decreased (Pandya et al., 2021). In addition, Pandya et al., (2021) determined that while there was a significant difference in the teaching methods and readiness of the faculty members before and during the pandemic period, there was no significant difference in the course content and technological support. This finding shows that faculty members are trying to adapt to the emergency distance education process by changing the learning and assessment methods by increasing their technical skills required for online education without making any changes to the content.

However, in the emergency remote teaching process, it seems difficult to expect digital transformation alongside this pedagogical transformation (Flores &

Gago, 2020; Iivari et al., 2020). In this context, supporting teachers with in-service training is the first solution that comes to mind. For example, a report prepared by the World Bank stated that in-service training should be supported both educationally and technologically (Beteille et al., 2020). However, we do not yet know the implications of teachers' experience for the effectiveness of online assessment design at the end of a process where solid theoretical knowledge can be obtained, and this knowledge is supported by practice. For example, Flores and Gago (2020) stated that ideally presented scenarios do not correspond to real application contexts.

However, solutions to be offered for online assessment should be designed taking into account a controlled workload for instructors and students.

3.1 Case 1: Online Assessment Methods and Techniques that Can Be Used in the Field of Education

Education faculties are institutions in which learning activities are examined in-depth, both theoretically and practically. In addition, the field of educational technology focuses on the design and development of appropriate technology to achieve learning goals. Therefore, with the dynamism of the field of educational technology, it can be expected that education faculties will be better equipped to meet emergency education needs. In this context, it seems possible that online assessment designs proposed for education faculties would be more diverse than those of other fields. This chapter, however, focuses on content-oriented course design in its examples of online assessment processes.

In content-oriented course design, the resource viewing reports can be in the online assessment process, if the number of students enrolled the course is high, multiple choice question tests can be structured using assessment tools such as peer assessment or assignments with rubrics and essays. Since individual student work is more important in content-oriented design, information can be obtained on how long participants stay on the course page (Fig. 3).

By examining participation reports, it is possible to identify in advance learners who are likely to have limited interaction with the content, and feedback can be provided to these learners through various tools in the Moodle LMS. For example, reports can be obtained on which resource a student has viewed and when and information messages can be sent through the system to learners who do not interact with the content of the relevant week (Fig. 4). 172.

If students want to be assessed through homework, Moodle LMS provides functional tools for teachers to score homework. The first of these tools is rubrics,

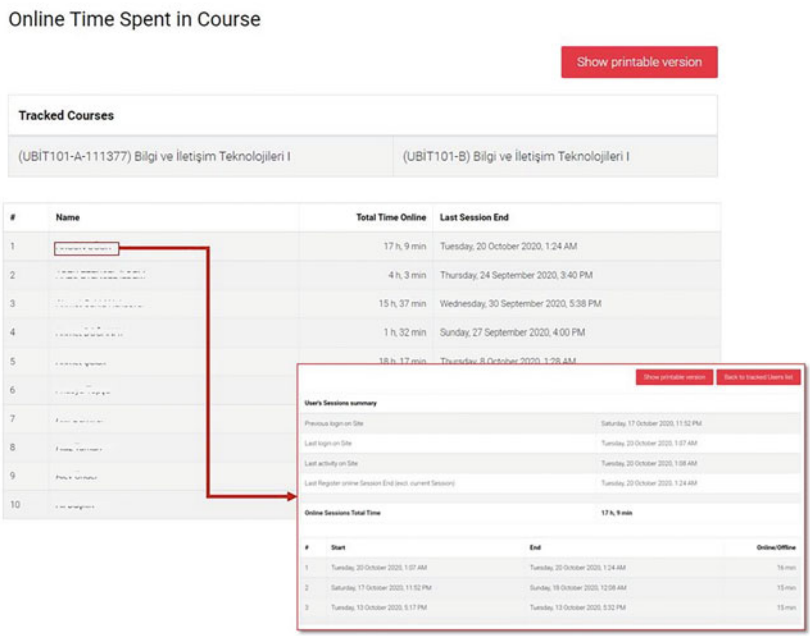


Fig. 3 Online time spent in a moodle course (Moodle plugins directory: Attendance Register, 2017)

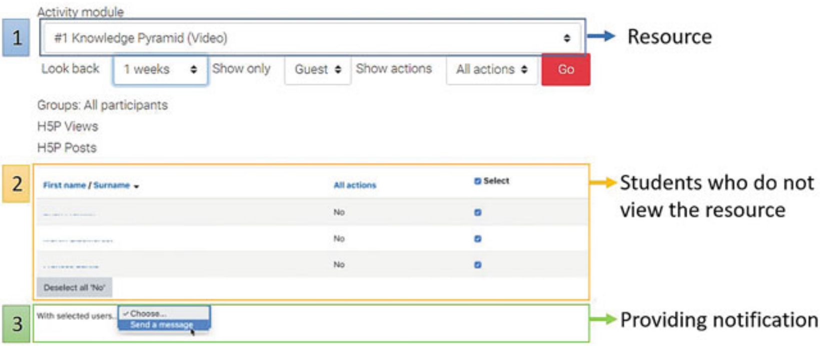


Fig. 4 Participation report in moodle LMS (Participation report—MoodleDocs, 2019)

which allow teachers to make categorical scoring of assignments by considering predetermined criteria. A fairer assessment of an assignment or project may be possible using rubrics within the Moodle LMS (Fig. 5).

With activity completion reports in a discussion-oriented lesson or unit design, it is possible to assess whether students fulfil the minimum tasks required from them in the forum (activity completion reports) or to assess their performance based on their activities in the discussion boards (rubrics enriched with learning analytics). When criteria are created on variables such as the minimum number of discussion views, the number of views shared in the discussion forum, or the average score given to the views shared in the discussion, activity completion reports can be used to check whether the student fulfils the basic requirements (Fig. 6).

The use of a component such as the discussion board also provides tools to facilitate the peer review processes. For example, it can be ensured that other students or teachers give points to each discussion comment, and the average of the given points is shown below the comment. Therefore, peer evaluation is automatically employed in this process (Fig. 7).

Peer assessment processes can be carried out more easily in a collaborative course design. Using tools within the LMS, students can be grouped randomly or according to a certain systematic criterion (such as the number of students in the group), making it easier for a group of students to work together. Thanks to these tools, each group can have a separate discussion activity or separate task (Fig. 8).

3.2 Case 2: Online Assessment Methods and Techniques that Can Be Used in the Field of Medical Education

Medical education, along with other training areas where practice and clinical skills are at the forefront, is one area facing difficulties in the remote online learning process. The need for hands-on practice is one of the biggest limitations to online education in the medical field, where practical experience with the human body is an integral part of the learning process. This greatly affects the needs in the design of the course. For this reason, these handicaps can be avoided by choosing a blended model instead of completely online education in areas in which clinical skills are at the forefront. Since medical education is both a theoretical and practical course, the content that can be presented theoretically can be transferred online. In addition, with the video-based structuring of the departments in which clinical skills are involved, students will be provided

Problem of Practice	The Problem of Practice is not identified or defined. ***0 points***	The Problem of Practice is not clearly defined or does not line up with the project scope. ***1 points***	The Problem of Practice is not clearly defined or aligns with the project scope. ***2 points***
Project Overview	The Project Overview is missing or incomplete. ***0 points***	The Project Overview is broadly defined or is missing some necessary details. ***1 points***	The Project Overview gives a clear picture of how this Capstone will work. ***2 points***
Artifacts/Examples	There are no artifacts or examples of any components of the Capstone. ***0 points***	The provided artifacts/examples do not provide enough information to give an understanding of how the Capstone will work in practise. ***1 points***	The provided artifacts/examples provide sufficient information to give an understanding of how the Capstone will work in practise. ***2 points***
Reflection	The reflection plan is missing or incomplete. ***0 points***	The reflection is broad, is missing some necessary details or is not related to the Capstone. ***1 points***	The reflection gives clear and useful feedback on the Capstone, including strengths and areas for growth. ***2 points***

Fig. 5 Rubrics example for assignment/project in moodle course (Rubrics—MoodleDocs, 2019; Rubrics and Grading Guides, 2021; Rubrics and Grading Guides, 2021)

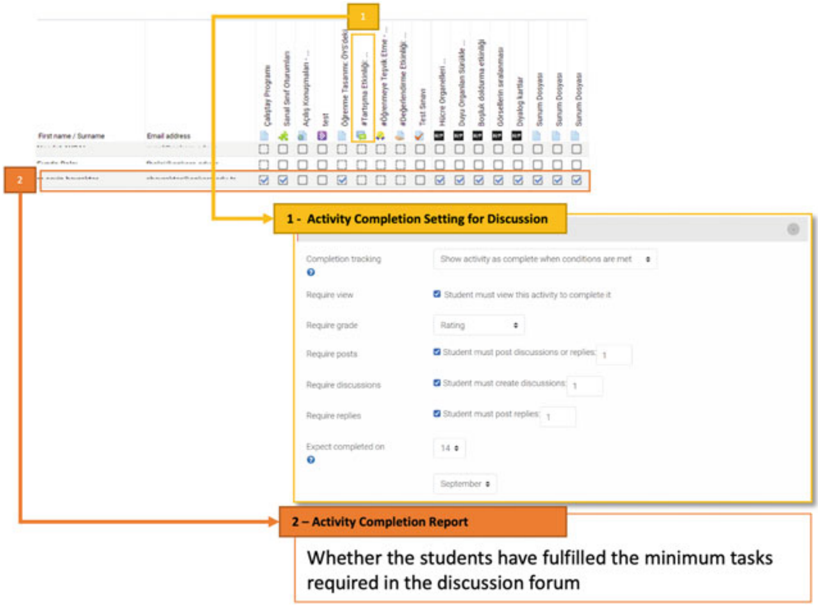


Fig. 6 Activity completion in moodle (Using Activity completion—MoodleDocs, 2021)

with the opportunity to repeat it many times and when they come together in a face-to-face environment, students will be provided with a preliminary knowledge of these skills. This will enable the transfer of the skill to be explained in a shorter time and the time during the lesson to be allocated to the questions of the students and the development of their practical skills. Of course, this process is reflected in the evaluation dimension in the same way. Structured Clinical Examinations (OSCEs), which are used to assess practice-based skills, or assessment processes conducted with volunteer patients are the main assessment elements that are interrupted in the fully online learning process (Fuller et al., 2020).

For this reason, a video-based course design in the field of medical education supports the achievement of learning goals. Asynchronous videos, lecture notes or handouts, face-to-face sessions, and external resources can be included in a lesson to be designed in this context.

Asynchronous videos can be used to increase the readiness level of students, especially before the lesson. In this context, important terms and topics specific to the related subject should be explained and theoretical information presented

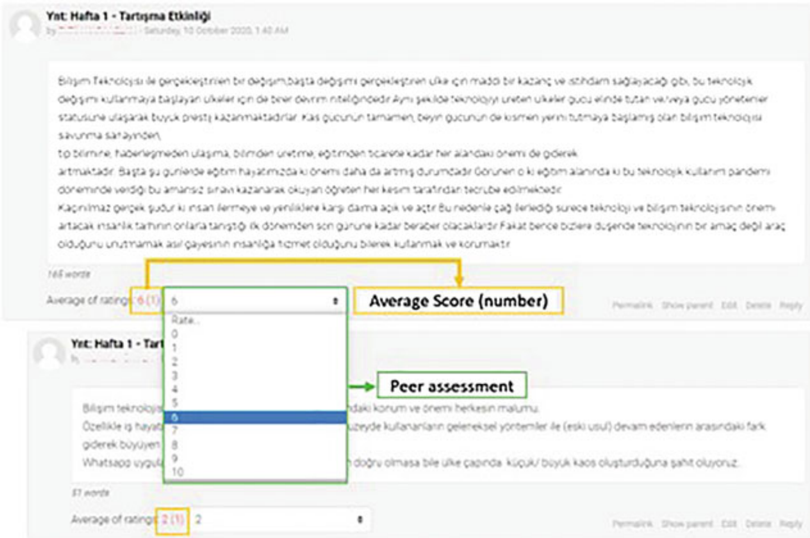


Fig. 7 Peer assessment in a discussion board (Forum activity—MoodleDocs, 2021)

in these videos. If necessary, clinical skills should be explained with a shot taken during practice in a second video. Points to be considered in asynchronous videos are limiting video length to 5-6 min and the instructor's adopting of an energetic and lively narration style during the lecture (Costley et al., 2017; Ilgaz, 2019). Lecture notes or handouts also serve to enrich the course by providing a guide for topics that include theoretical knowledge or practical skills. These articles, best practice videos, or important news can be presented as external resources.

In the assessment process, in accordance with the topic presented each week, peer assessment, homework evaluation, rubric evaluation, question banks enriched with equivalent questions, essays based on clinical case interpretation, tests, usage of rubrics enriched with learning analytics, Q&A sessions, video reflection blogs, video recordings, and group posters events can be given.

In such blended learning environment designs, it is important that the students do the pre-work and come to the lesson ready. For this reason, it may be preferable to include the students' viewing of videos, lecture notes, or external resources in the system in the course grading, as this will affect participation. Data usage logs can be utilized to obtain activity completion reports (Fig. 9).

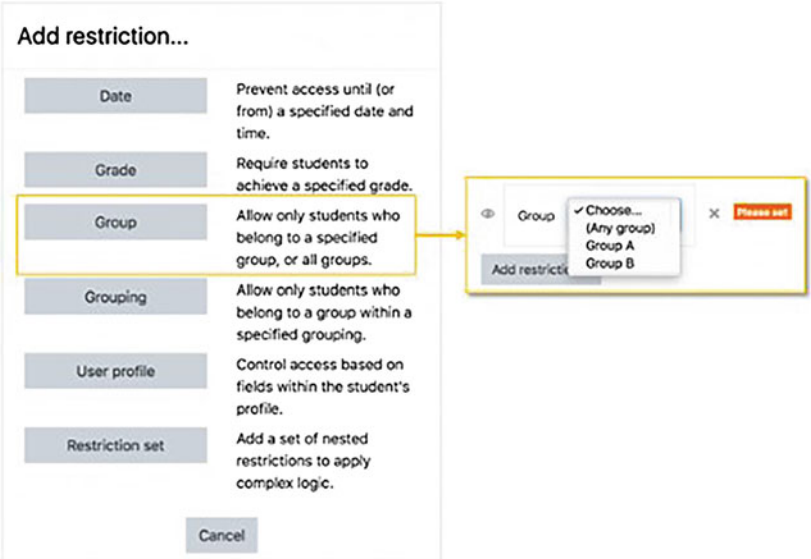


Fig. 8 Restrict access to a resource or an activity by the group (Restrict access settings—MoodleDocs, 2021)

Clinical practice videos uploaded by the students with the rubric to be prepared by the instructor can also be evaluated by the students in a peer assessment activity. For this purpose, the workshop or the peer work add-in can be used. With the workshop add-in, students can assess other uploaded videos according to the assessment criteria prepared by the instructor (Fig. 10). Optionally, it is possible to assess their own videos. Another effective aspect of this assessment method is that students should be able to self-assess according to the given criteria.

In a course design implemented as described, it can be observed whether each group has completed the assigned tasks and how much progress they have made thanks to the activity completion reports. In this way, separate feedback can be provided for each group. Reflecting experiences (such as micro-teaching) in the context of applied courses can be possible with the peer study tool. With this tool, group members can score their contributions to a project or product (micro-teaching video) prepared by the group (Fig. 11).

In addition to these data essay based clinical case, interpretations can also be used in medical education. The handicap in this type of assessment activity

Fig. 9 Activity completion report (Activity completion report—MoodleDocs, 2021)

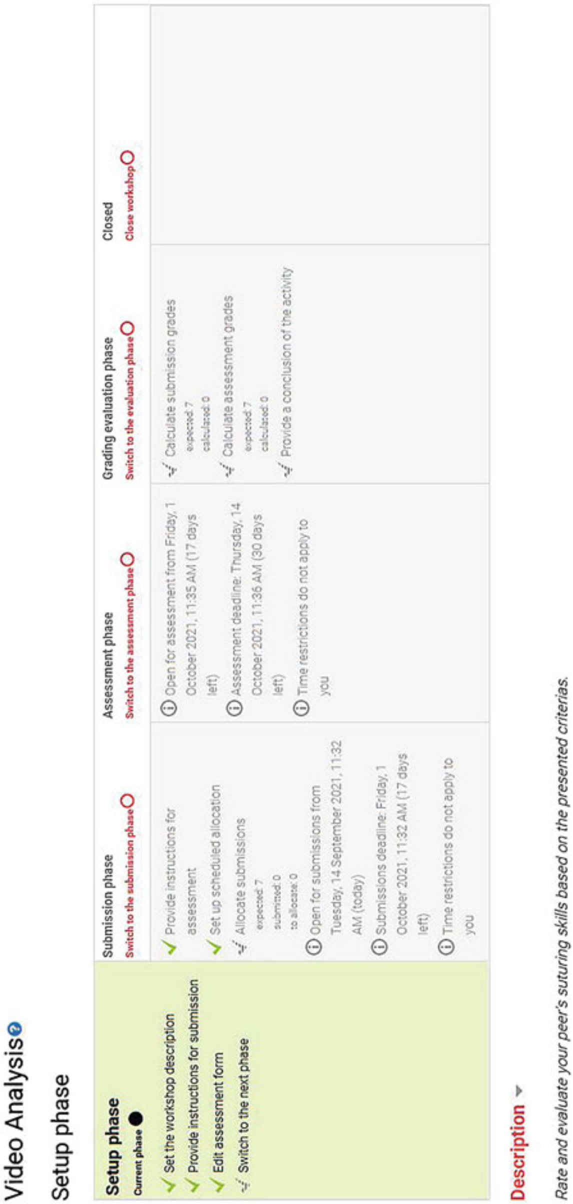


Fig. 10 Workshop activity (Workshop activity—MoodleDocs, 2021)

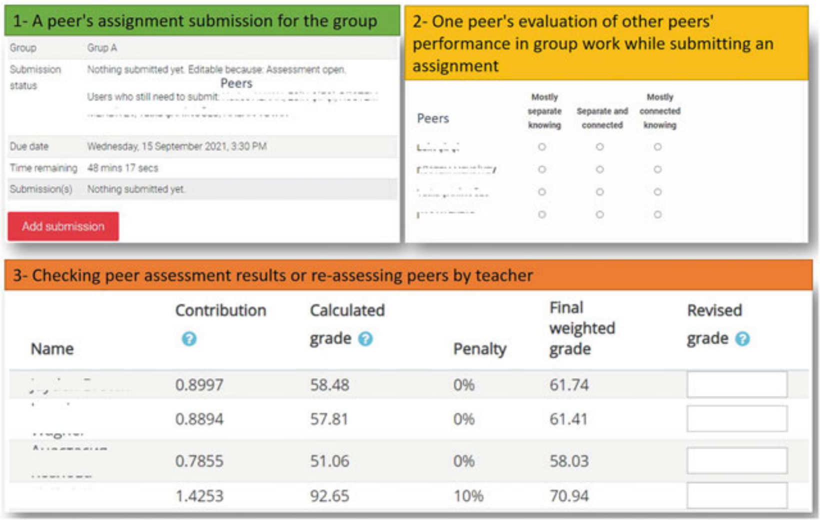


Fig. 11 Peer work in moodle (Peerwork Activity—MoodleDocs, 2020)

is that it requires manual grading. If there is a large group of students, more than one evaluator may take part in the process. Assessments can be made by the instructors with the help of a rubric containing assessment criteria. It is also possible to give detailed feedback to the students in this activity (Fig. 12).

3.3 Case 3: Online Assessment Methods and Techniques that Can Be Used in the Field of Legal Education

Like all other discipline areas, legal education is also affected by digitalization. While legal education does not consist of psychomotor abilities, it requires decision-making abilities. The use of simulations (Barton, & Maharg, 2006), asynchronous videos (Lacey, 2021), and online learning (Dutton et al., 2019) are not new methods in the legal education process. Engagement is one of the important elements in legal education, where case studies and decision-making processes are examined intensively (Gerken, 2021). It is important for legal education to include case studies of different qualities that will increase engagement during the course process. In course design, a more text-based design can be

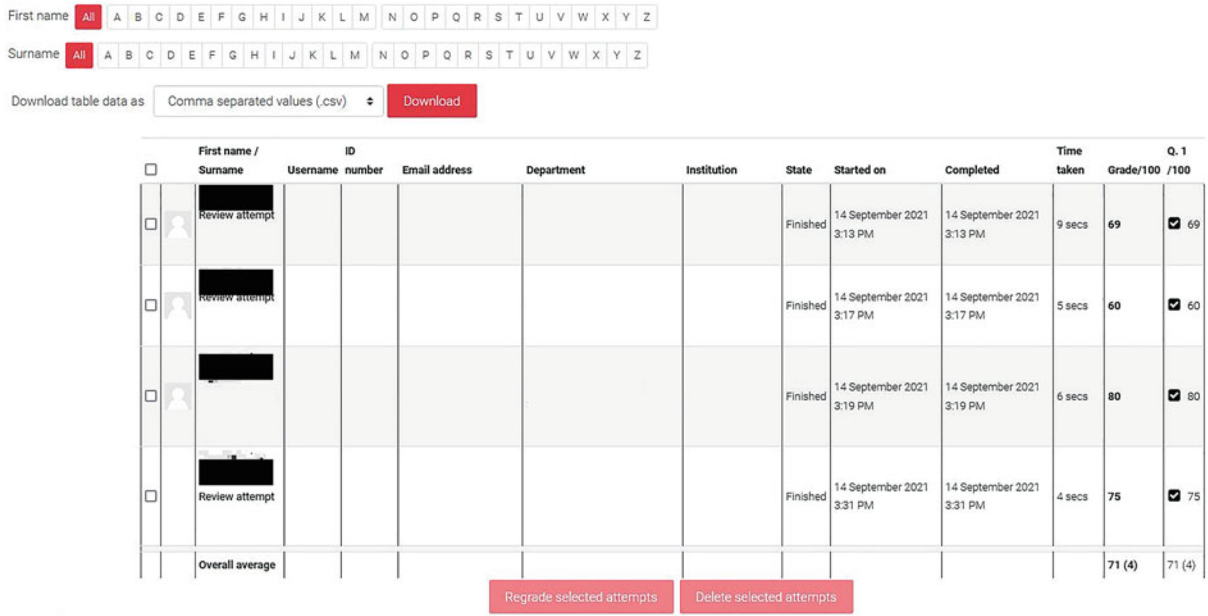


Fig. 12 Essay exam grading (Essay question type—MoodleDocs, 2021; Grading Essay Questions in Moodle Quizzes—UP Moodle Guides, 2020)

preferred. In addition, the preparation of course and assessment designs that prioritize reasoning, decision-making, and effective communication skills will be more productive for this field. In assessment processes, more interactive processes beyond the traditional assessment approach are recommended in the field of legal education (Dutton et al., 2019). Such an assessment design process can be used to create a pool of questions that are formed at the level of knowledge and understanding covering laws and regulations and discussion and collaboration activities to enable learning at the decision-making and reasoning level. Assessing the scores obtained from quiz activities created for each subject, calculating the overall performance score by weighting the quiz scores according to subject density, and case presentations or oral evaluations in the context of designing simultaneous activities in virtual classrooms are other assessment tools that can be used in such a course design (Fig. 13).

A content-oriented design may also be possible in the field of law, as in the field of education. Documents containing laws and regulations in the field of law can be shared on the course page and discussions can be made on some sample cases. For example, a case of conflict between public and legal entities in the field of administrative law can be given as an exemplary case. According to the shared regulations, students can be expected to discuss the dispute by basing their

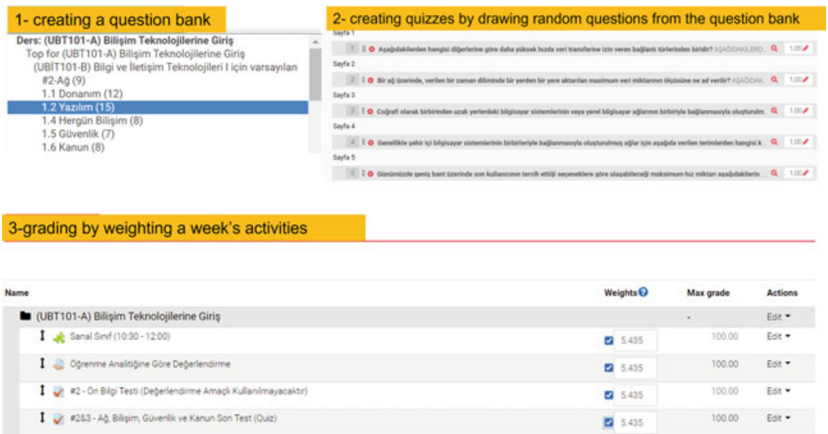


Fig. 13 Using a question bank and weighting the assessment activity (Gradebook—MoodleDocs, 2013; Question bank—MoodleDocs, 2021)

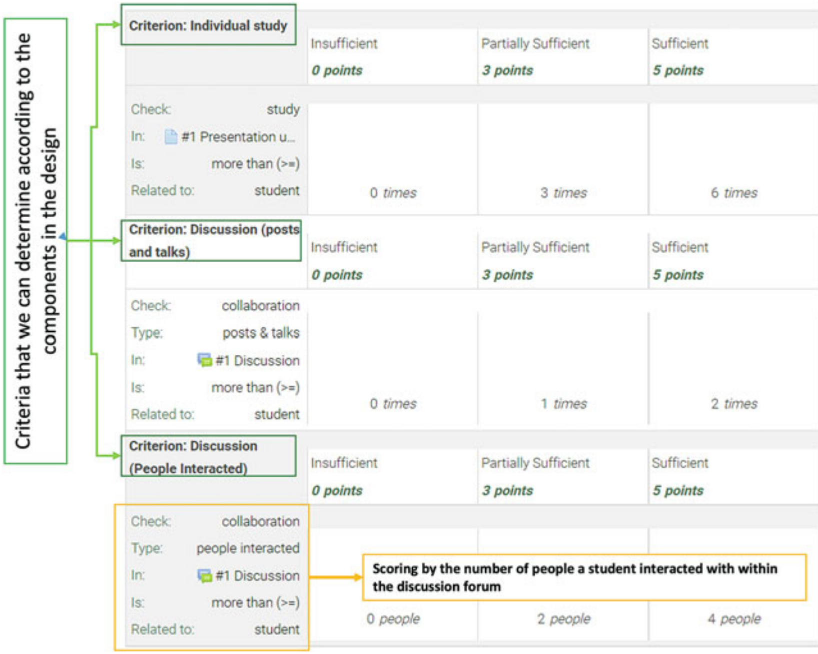


Fig. 14 Learning analytics enriched assignment in moodle (Learning Analytics Enriched Rubric—MoodleDocs, 2021)

claims on the regulations. In such a design, the online assessment process can be carried out through assignments enriched with learning analytics (Fig. 14).

According to Fig. 14, students are expected to review the presentation in the virtual lesson, as well as to participate in discussions and interact with other participants in the discussion, in a one-week period. The assessment of the week can be made by the system by following the participant's performance this week according to predetermined criteria. For example, if a student views weekly content once, comments once on the discussion board, and interacts with only one different person, he or she will have received 9 out of 15 total points. Therefore, an online assessment process designed in this way facilitates the calculation of the overall performance and provides appropriate opportunities for process assessment.

4 Conclusion and Implications

Assessment is an important component that certifies the achievement of competency at the end of the program, motivates the student, and determines the level of learning outcomes. It is considered necessary for a successful assessment that this important process is carried out in an integrated manner with the instructional design (Herron & Wright, 2006; Reeves, 2000). Therefore, an assessment process independent of the content cannot be considered. The American Association of Higher Education (AAHE) described assessment using nine principles that emphasize both the outcome and the experience during the process (Reeves, 2000). This emphasis on the evaluation of the process points to the effectiveness of formative assessment. In online learning environments in which stakeholders are not physically together in particular, the design of the assessment process should be considered as an ongoing process and planned accordingly. For this reason, formative assessment offers a significant advantage to instructors in the online environment, which is to give timely and effective feedback to individuals who are already far away and to prevent possible conceptual mistakes or mislearning. In the absence of such assessment, compensations made as the result of an evaluation at the end of the learning process may be undertaken too late to positively affect learning outcomes (Bhagat & Spector, 2017).

In this book chapter, sample situations as suggestions for different learning areas are specified and instructional design and assessment designs are shared accordingly. Of course, these examples can be diversified, supported by different tools, or integrated into other areas. However, the common point of these shared cases is that they use assessment methods related to the activities carried out in the process, apart from classical assessment methods. In this approach, which provides feedback on learning processes, students' skills and performances can be assessed with feedback produced in the system (Guerrero-Roldán & Noguera, 2018). Although the feedback process is very important in assessment, it remains less commonly used because it requires more time and effort compared to classical evaluation methods (Farrell & Rushby, 2016; Paul & Jefferson, 2019; Peytcheva-Forsyth & Aleksieva, 2021). In online learning processes in which there is no physical and temporal synchronization, the use of feedback increases engagement levels of the students (Martin & Bolliger, 2018) and, accordingly, student satisfaction (Li et al., 2016; Peytcheva-Forsyth & Aleksieva, 2021; Yuliang, 2012), as well as decreases dropout rates (Ivankova & Stick, 2006; Lee & Choi, 2011; Xiong, & Suen, 2018). Considering the sample cases presented here, it can be said that the assessment tools used provide indirect feedback for both the teacher and the student. However, depending on the course design and the teacher's experience, such tools can also be used to provide direct feedback with simultaneous sessions.

It is expected that the effectiveness of the online learning process will increase when the feedback mechanism is presented as real-time data over an LMS using learning analytics. It is very important for the assessment process to utilize data obtained from the interaction of individuals with the system and the content of the learning process (Crisp, 2020). The circular and iterative use of the data to be obtained through feedback and systems in the online learning process will increase efficiency for both students and instructors.

While the positive contribution of an assessment design prepared in harmony with the learning contents of the process cannot be denied, it is quite possible to encounter situations such as reliability or dishonesty that may reduce the effectiveness of this process. Authentic and personalized assessment activities also can prevent cheating or plagiarism (Arnold, 2016; Gikandi et al., 2011). With proctoring or plagiarism software, cheating problems in formative assessment can be partially, if not completely, controlled. Continuous information sharing also can be valuable for learners. Even if they gain in the short term, there are studies showing that informing learners about the learning losses they will experience in the long term is effective in preventing cheating (Arnold, 2016). In addition, if the process is completely online, face-to-face virtual sessions can be considered as oral exams, while if the learning process is blended face-to-face assessment options should be considered.

The weighting of the activities used in the assessment process on the overall grading method is also an important issue in this process. Even if instructors can determine the distribution scores, institutional rules and regulations also play a decisive role in this assessment process. In this context, weighting within the framework of the flexibility presented and providing the necessary information will also make the online assessment process more powerful.

Although the online assessment process is complex and its multidimensional nature requires effort, it can be carried out effectively with the integration of the right tools. The course designs and tools shared here can be diversified or used in different ways, and similar designs can be used in other disciplines not exemplified here.

References

Activity Completion Report—MoodleDocs. (2021). Moodle.org. https://docs.moodle.org/311/en/Activity_completion_report.

Adedoyin, O. B., & Soykan, E. (2020). Covid-19 pandemic and online learning: The challenges and opportunities. *Interactive Learning Environments,* 1–13.https://doi.org/10.1080/10494820.2020.1813180.

Alruwais, N., Wills, G., & Wald, M. (2018). Advantages and challenges of using e-Assessment. *International Journal of Information and Education Technology, 8,* 34–37.

American Association of Higher Education, Nine Principles of Good Practice for Assessing Student Learning, AAHE, Washington, D.C. http://www.aahe.org/.

Arnold, S. D. (2014). Assessing student learning online. In D. G. Sampson, D. Ifenthaler, J. M. Spector, & P. Isaías (Eds.), *Digital systems for open access to formal and informal learning* (pp. 83–100). Springer.

Arnold, I. J. M. (2016). Cheating at online formative tests: Does it pay off? *The Internet and Higher Education, 29,* 98–106. https://doi.org/10.1016/j.iheduc.2016.02.001.

Barton, K., & Maharg, P. (2006). Simulations in the wild: Interdisciplinary research, design and implementation. In C. Aldrich, D. Gibson, & M. Prensky (Eds.), *Games and simulations in online learning* (pp. 115–148). Idea Group Ltd.

Beteille, T., Ding, E., Molina, E., & Pushparatnam, A., T. (2020). *Three principles to support teacher effectiveness during COVID-19*. World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/33775.

Bhagat, K. K., & Spector, J. M. (2017). Formative assessment in complex problem-solving domains: The emerging role of assessment technologies. *Journal of Educational Technology & Society, 20*(4), 312–317.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167. https://doi.org/10.1080/713695728.

Costley, J., Hughes, C., & Lange, C. (2017). The effects of instructional design on student engagement with video lectures at cyber universities. *Journal of Information Technology Education: Research, 16,* 189–207. https://doi.org/10.28945/3728.

Crisp, E. (2020). Leveraging feedback experiences in online learning. *EDUCAUSE*. https://er.educause.edu/articles/2020/6/leveraging-feedback-experiences-in-online-learning.

Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2003). *The student assessment handbook: New directions in traditional and online assessment* (1st ed.). Routledge. https://doi.org/10.4324/9780203416518.

Dutton, Y., Ryznar, M., & Long, K. (2019). Assessing online learning in law schools: Students say online classes deliver. *Denver University Law Review, 96*(3), 493–534.

Edmonds, G. S., Branch, R. C., & Mukherjee, P. (1994). A conceptual framework for comparing instructional design models. *Educational Technology Research and Development, 42*(4), 55–72. https://doi.org/10.1007/BF02298055.

Essay question type—MoodleDocs. (2021). Moodle.org.: https://docs.moodle.org/311/en/Essay_question_type.

Farrell, T., & Rushby, N. (2016). Assessment and learning technologies: An overview. *British Journal of Educational Technology, 47*(1), 106–120. https://doi.org/10.1111/bjet.12348.

Fask, A., Englander, F., & Wang, Z. (2014). Do online exams facilitate cheating? An experiment designed to separate possible cheating from the effect of the online test taking environment. *Journal of Academic Ethics, 12*(2), 101–112. https://doi.org/10.1007/s10805-014-9207-1.

Flores, A. M., & Gago, M. (2020). Teacher education in times of COVID-19 pandemic in Portugal: National, institutional and pedagogical responses. *Journal of Education for Teaching, 46*(4), 507–516. https://doi.org/10.1080/02607476.2020.1799709.

Forum activity – MoodleDocs (2021). Moodle.org. https://docs.moodle.org/402/en/Forum_activity.

Fuller, R., Joynes, V., Cooper, J., Boursicot, K., & Roberts, T. (2020). Could COVID-19 be our 'There is no alternative' (TINA) opportunity to enhance assessment? *Medical Teacher, 42*(7), 781–786. https://doi.org/10.1080/0142159X.2020.1779206.

Gerken, H. K. (2021). Will legal education change post-2020? *Michigan Law Review, 119*(6), 1059–1068. https://doi.org/10.36644/mlr.119.6.will.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004.

Gradebook—MoodleDocs. (2013). Moodle.org. https://docs.moodle.org/27/en/Gradebook.

Grading Essay Questions in Moodle Quizzes—UP Moodle Guides. (2020). https://sites.up.edu/moodle/kbe_knowledgebase/grading-essay-questions-in-moodle-quizzes/.

Guerrero-Roldán, A.-E., & Noguera, I. (2018). A model for aligning assessment with competences and learning activities in online courses. *The Internet and Higher Education, 38,* 36–46. https://doi.org/10.1016/j.iheduc.2018.04.005.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice, 4*(3), 365–379. https://doi.org/10.1080/0969594970040304.

Herron, J. F., & Wright, V. H. (2006). Assessment in online learning: Are students really learning? In H.V. Wright, C. Syzmanski Sunal, & E. K. Wilson (Eds.), *Research on enhancing the interactivity of online learning* (pp. 45–64). Information Age Publishing.

Ilgaz, H. (2019). Adult learners' participation in a blended learning environment: A case study on imposed pace learning. *Malaysian Online Journal of Educational Technology (MOJET)*, *7*(4), 15–29. https://doi.org/10.17220/mojet.2019.04.002.

Ivankova, N. V., & Stick, S. L. (2006). Students' persistence in a distributed doctoral program in educational leadership in higher education: A mixed methods study. *Research in Higher Education, 48*(1), 93. https://doi.org/10.1007/s11162-006-9025-4.

Iivari, N., Sharma, S., & Ventä-Olkkonen, L. (2020). Digital transformation of everyday life – How COVID-19 pandemic transformed the basic education of the young generation and why information management research should care? *International Journal of Information Management, 55,* 102183. https://doi.org/10.1016/j.ijinfomgt.2020.102183.

Lacey, D. (2021). Do all roads lead to Rome? Effectively using asynchronous online reflective practice in an externship course. *UMKC Law Review, 89*(3), 609–644.

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development, 59*(5), 593–618. https://doi.org/10.1007/s11423-010-9177-y.

Learning Analytics Enriched Rubric—MoodleDocs. (2021). Moodle.org. https://docs.moodle.org/311/en/Learning_Analytics_Enriched_Rubric.

Li, N., Marsh, V., & Rienties, B. (2016). Modelling and managing learner satisfaction: Use of learner feedback to enhance blended and online learning experience. *Decision Sciences Journal of Innovative Education, 14*(2), 216–242. https://doi.org/10.1111/dsji.12096.

Lourdes, G., Geoffrey, C., & Ivan, A. (2017). Trends and challenges of E-Assessment to enhance student learning in higher education. In C. Elena & I. Georgeta (Eds.), *Innovative practices for higher education assessment and measurement* (pp. 36–56). IGI Global.

Martin, F., & Bolliger, D. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learning, 22*(1). https://doi.org/10.24059/olj.v22i1.1092.

Moodle plugins directory: Attendance Register. (2017). Moodle.org. https://moodle.org/plugins/mod_attendanceregister.

Myyry, L., & Joutsenvirta, T. (2015). Open-book, open-web online examinations: Developing examination practices to support university students' learning and self-efficacy. *Active Learning in Higher Education, 16*(2), 119–132. https://doi.org/10.1177/1469787415574053.

Pandya, B., Patterson, L., & Cho, B. (2021). Pedagogical transitions experienced by higher education faculty members—"Pre-Covid to Covid". *Journal of Applied Research in Higher Education, ahead-of-print* (ahead-of-print). https://doi.org/10.1108/JARHE-01-2021-0028.

Paul, J., & Jefferson, F. (2019). A comparative analysis of student performance in an online vs. face-to-face environmental science course from 2009 to 2016. *Frontiers in Computer Science, 1*(7). https://doi.org/10.3389/fcomp.2019.00007.

Participation report—MoodleDocs. (2019). Moodle.org https://docs.moodle.org/311/en/Participation_report.

PeerWork Activity—MoodleDocs. (2020) Moodle.org. https://docs.moodle.org/311/en/Peerwork_Activity.

Peytcheva-Forsyth, R., & Aleksieva, L. (2021). Forced introduction of e-assessment during COVID-19 pandemic: How did the students feel about that? (Sofia University case). *AIP Conference Proceedings, 2333*(1), 050013. https://doi.org/10.1063/5.0041862.

Question bank—MoodleDocs. (2021). Moodle.org. https://docs.moodle.org/311/en/Question_bank.

Ragupathi, K. (2020). *Designing effective online assessments resource guide*. National University of Singapore. https://www.nus.edu.sg/cdtl/docs/default-source/professional-development-docs/resources/designing-online-assessments.pdf.

Rapanta, C., Botturi, L., Goodyear, P., Guàrdia, L., & Koole, M. (2020). Online university teaching during and after the Covid-19 crisis: Refocusing teacher presence and learning activity. *Postdigital Science and Education, 2*(3), 923–945. https://doi.org/10.1007/s42438-020-00155-y.

Reeves, T. C. (2000). Alternative assessment approaches for online learning environments in higher education. *Journal of Educational Computing Research, 23*(1), 101–111. https://doi.org/10.2190/gymq-78fa-wmtx-j06c.

Restrict access settings—MoodleDocs. (2021). Moodle.org. https://docs.moodle.org/311/en/Restrict_access_settings.

Rubrics—MoodleDocs. (2019). Moodle.org. https://docs.moodle.org/310/en/Rubrics.

Rubrics and Grading Guides. (2021). Example of Rubric. https://help.lsit.ucsb.edu/hc/en-us/articles/360027568251-Rubrics-and-Grading-Guides.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/BF00117714.

Sheridan, R. (2006). Reducing the online instructor's workload. *Educause Quarterly, 29*(3), 65–67. https://www.learntechlib.org/p/103741/.

Spector, J. M., & Park, S.-W. (2017). *Motivation, learning and technology: Embodied educational motivation.* Routledge.

Stödberg, U. (2012). A research review of e-assessment. *Assessment & Evaluation in Higher Education, 37*(5), 591–604. https://doi.org/10.1080/02602938.2011.557496.

Using Activity completion—MoodleDocs. (2021). Moodle.org. https://docs.moodle.org/311/en/Using_Activity_completion.

Workshop activity - MoodleDocs. (2021). Moodle.org. https://docs.moodle.org/311/en/Workshop_activity.

Yuliang, L. (2012). Instructor feedback, learner satisfaction, and online learning. In Y. Harrison Hao & Y. Steve Chi-Yin (Eds.), *Handbook of research on practices and outcomes in virtual worlds and environments* (pp. 536–550). IGI Global.

Xiong, Y., & Suen, H. K. (2018). Assessment approaches in massive open online courses: Possibilities, challenges and future directions. *International Review of Education, 64*(2), 241–263. https://doi.org/10.1007/s11159-018-9710-5.

Hale Ilgaz is an associate professor at the Faculty of Open and Distance Education, Ankara University. Her research interests include distance education, e-learning, instructional design, cognitive processes in e-learning environments, and human-computer interaction. She is designing and coordinating the online diploma and certificate programs in the university.

Denizer Yildirim is working as a lecturer and education technologist at the Faculty of Open and Distance Education at Ankara. His work focuses on educational data mining, learning analytics, learning design, e-assessment, and AI-based learning management systems, robotics integration in education. He is the researcher of EU and national projects that aim at MOOC design and e-content creation. He is coordinating Learning Analytics and Learning Design related projects.

Support Measures for Students Before and During Written Online Distance Exams: The Case of Vienna University of Economics and Business (WU)

Florian Mosböck, Julia Dohr, and Andrea Ghoneim

Abstract

This article presents a case study of a successful online-assessment practice in university teaching. It focuses on the support measures that have been implemented by the Vienna University of Economics and Business (WU) to meet the needs of all identified stakeholders in written online distance exams. Besides the organizational and technical set-up of the written online distance exams, the article describes support measures for students implemented between April 2020 and June 2021, covering a period of three semesters. Student support includes guidance for general autonomous self-preparation (with special respect to technical requirements and communication settings), a Browser Multimedia Test to simulate automated online supervision, and structured information about the specific exam via a pre-designed text template in the online exam environment. In addition, communication connected with the exam reassures students in the stressful examination situation. The article describes compensatory actions for students with special needs. The case study is a snapshot of a work in progress, because the written online distance exams had to be implemented quickly after the Covid-19 pandemic developed. WU is still evaluating and adapting its formats for written online distance exams.

F. Mosböck · J. Dohr · A. Ghoneim (✉)
Digital Teaching Services, WU Vienna, Vienna, Austria
e-mail: Andrea.ghoneim@wu.ac.at

F. Mosböck
e-mail: Florian.mosboeck@wu.ac.at

J. Dohr
e-mail: Julia.dohr@wu.ac.at

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_9

Keywords

Case study • Online distance exams • Support measures

1 Introduction

In university life, exams play an important role for both students and teachers. Since the turn of the millennium and partly in connection with the Bologna process, online exams have been implemented at an increasing number of European universities (Bauer et al., 2008). In Austria, for example, the University of Klagenfurt has been conducting online exams for the past 20 years. In 2010, they developed a secure web browser to conduct online exams that allowed students to participate in exams on campus, either with their own devices or with a device from the university (Frankl & Schratt-Bittner, 2020).

The stakeholders of online exams can be identified when looking closely at the process of designing, carrying out, and assessing an exam. As early as 2008, Bauer et al. published a standardized process for administering exams. These standards identified the steps for IT-based processes around computer-based exams, which were the precursors of online exams (Bauer et al., 2008). The standardized examination process mentions main stakeholders – instructors, students, student office, IT-administration – but does not refer to the possible support needed by those stakeholders (ibid.).

Online exams, as a complex activity with various processes, participants, responsibilities, and competencies, require a university's administrative support to develop a clear picture of the tailored services needed by specific target groups. It is important to know the stakeholders' needs and to become familiar with the organizational, legal, and technical prerequisites and requirements that affect these stakeholders (Bandtel et al., 2021).

Kuikka et al. (2014) emphasize the importance of staff training for a successful implementation of online exams. Cramp et al. carried out literature research on online exams in 2019 and summarized that "course facilitators" (frequently, the course instructors) usually need technical support to set up the exams (Cramp et al., 2019, p. 7).

Students are also mentioned as stakeholders, as in Bauer et al. (2008) or in Halbherr et al. (2014). Undoubtedly, students need assistance to get acquainted with new exam formats, such as online exams (Cramp et al., 2019; Ilgaz & Adanır, 2020). However, publications before 2020 did not focus on the students' needs for support with online exams. The most likely reason was that, before

COVID-19, online exams were still administered on campus, and students used to be supported by the examiners in the classroom.

The COVID-19 pandemic created a new situation because online exams on campus were not possible. Exams had to be cancelled or carried out remotely. In Austria, some universities changed to remote, take-home or open-book examination formats. However, to simultaneously examine a large group of students, the only possible format was to set up online exams administered at a distance (Pausits et al., 2021).

The sudden shift to distance learning and online distance exams in spring 2020 because of COVID-19 resulted in both new support needs and a shortage of technical equipment that could be borrowed or bought. For example, even university teachers did not always have access to webcams. For the written online distance exams, students now had to use their own computers primarily equipped for their personal studying. Many students faced the new problem of needing a stable internet connection; some students lacked a quiet space for studying and for writing exams (Pausits et al., 2021; EUA, 2020). An internal evaluation of WU students (Ledermüller & Spörk, 2020) revealed that their experiences and needs reflected those issues identified in scholarly literature.

Distance learning and, more specifically, the use of written online distance exams have increased the overall demand for appropriate measures to support examiners and students (Bouchey et al., 2021; Garrett et al., 2020; Johnson et al., 2020). As quickly as possible, the WU's unit for digital teaching support, Digital Teaching Services, set up guidance materials for instructors on distance learning. Guidance on exams included recommendations for oral distance exams and remote take-home exams. Since many courses at WU have exams for a very large number of students, the format of the written online exam had to be defined and its implementation at a distance made possible. In addition, the following support units helped enable written online distance exams: support units of Program Management and Teaching & Learning Support (especially Digital Teaching Services), MyLEARN developers and technical administrators, IT-Services, and Student Services.

In the following case study, the authors will focus on the support measures that enabled students to participate in written online distance exams.

2 Written Online Distance Exams at WU

On February 25, 2020, the first cases of COVID-19 were reported in Austria. As of March 11, the Austrian authorities proclaimed a lockdown for Austria. As a consequence, universities in Austria were closed from mid-March. At the WU campus, all face-to-face teaching was suspended or switched to online formats. On March 26, instructors were informed that all exams would have to be switched to online variants.

At WU, large courses of the type LVP (*Lehrveranstaltungsprüfung*) have exams three times per semester during centrally organized large-scale exam weeks (*Großprüfungswochen*) that include the exams from the introductory phase and the main phase as well as smaller intermediate and final exams Of course, smaller exams also take place outside these official exam weeks and throughout the year. Until spring 2020, these exams were written exams on paper. With the end of April (April 27) and the first large-scale exam week during the lockdown, these exams had to be written online at a distance. Beginning with the exam week in April 2020 until the exam week in June 2021, WU conducted 433 online exams with a total of 93,229 examination attempts by students (cf. Table 1).

The exam week in January 2021 was the biggest so far, with 92 exams and 15,018 examination attempts. Written online distance exams have been conducted with up to 1900 candidates per exam.

Table 1 Number of exams and examination attempts in exam weeks

Exam weeks since the start of the COVID-19 pandemic	Exams within the exam week	Examination attempts per exam week
April 2020	38	12,328
June 2020*	51	11,046
October 2020	17	8984
November 2020*	73	14,566
January 2021*	92	15,018
March 2021	30	8585
April 2021	59	11,512
June 2021*	73	11,190

Note. This table gives an overview of the number of written online distance exams and the total number of examination attempts per exam week from April 2020 to June 2021 and only includes exams during the official exam weeks. The large-scale exam weeks are marked with * in the left column. Own table

There was only a short time between the announcement of the first lockdown in Austria and the first large-scale exams. After April 2020, due to the uncertain situation, administrators repeatedly postponed the return of exams to the WU campus. In the end, most of the major exams remained in online distance mode until the end of the summer semester 2021.

At the beginning of the pandemic, the legal basis for written online distance exams was the examination regulation of WU and the "COVID-19-University and Higher Education Regulation" (*COVID-19-Universitäts- und Hochschulverordnung*), which was published by the Austrian Federal Ministry of Education, Science and Research. Section 11. (1) of the COVID-19-University and Higher Education Regulation sets out the minimum requirements for the proper conduct of electronic exams. Among other things, it states that technical measures must ensure that students work independently. Furthermore, it stipulates that in the event of technical problems not caused by the student, the exam will not be credited (BGBI II No. 2020/171). On the basis of the above-mentioned federal regulation, the WU issued a by-law for distance learning and for the uniform conduct of online exams. When the federal regulation expired after the winter semester 2020/21, so did the regulation of WU. Therefore, a WU directive replaced the expired WU regulation in the summer semester 2021. Beginning with the winter semester 2021/22, Section 76a of the Austrian Universities Act 2002 (BGBI. I No. 120/2002) applies to the conduct of online exams in addition to the examination regulations of WU.

The process for creating and supporting online exams as well as the exam environments had to be set up quickly. The technical system and the administrative process have been repeatedly evaluated and revised based on previous experiences. A first major revision was made during the summer break in September 2020. The changes were retained for the following winter and summer semester. Minor improvements and adjustments were made on an ongoing basis. The second major revision of the online exams was planned for September 2021. The contents of the revisions will be discussed in more detail later.

For examiners, a standardized examination process was designed and implemented. This ranged from creating and preparing the exams, to offering the exam and then assessing the students' submissions. This examination process is supported by the staff of WU's Digital Teaching Services. Among other things, they prepare information material for instructors in text and video formats, support instructors in creating exams, and answer questions on technical issues and exam didactics. Examiners can ask Digital Teaching Services to evaluate the exam format. This process has been continuously evaluated and improved. The following section describes the design of the online examination environment, the design of the exam, and the adaptation of the administrative process.

2.1 Online Exam Environments

The in-house, learning management system MyLEARN serves WU examiners as the exam platform. Written online distance exams are conducted either in the regular online course environments or in especially created, so-called "separate online exam environments" (SOEE). Both are on MyLEARN. The SOEEs are pre-structured environments created specifically for conducting the exams and are not part of the regular online course environments. Between April 2020 and July 2021, 840 SOEEs have been created for processing the exams.

The SOEE allows for online supervision (online proctoring). This in-house supervision tool helps prevent cheating and was developed especially for the written online distance exams in MyLEARN. Examiners can request the use of this online supervision only when using a SOEE. This is not a technical limitation. Rather, it was important to WU that course instructors only use online supervision for interim and final exams that are critical points in the curriculum. Students must read a privacy policy for the online supervision in the SOEE. Recordings of the online supervision are stored on WU servers during the four-week grading period and the two-week appeal period. Afterwards, the data is deleted. Examiners are free to choose if they want to use the online supervision feature or not. If used, the online supervision is usually activated half an hour before the exam starts and is deactivated 30 min after the end of the last exam. Students are asked to access the SOEE in the 30 min before the start of the exam (used as preparation time) to set up the online supervision. Once online supervision begins, students are recorded via the camera and microphone on their own device for the entire duration of the exam. Also, their screen is recorded. While online supervision is active, students can check that the recording is working via preview images. The recordings of the online supervision are checked by staff of the support units or by examiners themselves during or immediately after the exam. From the 433 exams within the exam weeks, 233 exams used online supervision (cf. Table 2).

Since regular online course environments do not support online supervisions, examiners must apply for a SOEE in order to use online supervision. Outside of the exam weeks, however, most written online exams that do not use online supervision are conducted in regular online course environments.

Table 2 Ratio of exams with online supervision in exam weeks

Exam weeks at WU since the start of the COVID-19 pandemic	Exams without online supervision	Exams with online supervision
April 2020	14	24
June 2020*	23	28
October 2020	5	12
November 2020*	38	35
January 2021*	54	38
March 2021	7	23
April 2021	24	35
June 2021*	35	38

Note. This table compares the number of written online distance exams with online supervision to written online distance exams without online supervision per exam week from April 2020 to June 2021. Only exams that took place during the official exam weeks are considered. The large-scale exam weeks are marked with * in the left column. Own table

With a regular online course environment, all registered students in the course have access to the exam and do not need to be added individually. The regular online course environment will, however, contain all learning materials (e.g., texts, videos, exercises). Examiners must activate the examination statement in the regular online course environments for each exam. The examiners independently choose the communication channel(s) during the exam. They can communicate with students via email, a Microsoft Teams program, or a chat program directly integrated into the learning management system.

In spring 2020, the SOEE was first developed and used for exams. The SOEE contained various structured, pre-built elements and offered the possibility to use the online supervision tool (cf. Fig. 1). The start page provided students with information about Microsoft Teams, online supervision, and the examination process. They were informed on what tasks to complete when preparing for the exam. Before the exam began, students had to verify their identity (by uploading a picture showing their faces and their ID card next to each other), confirm access to the exam, consent to online supervision, and accept the examination statement. The side menu offered information for examiners regarding the creation, implementation, and assessment of the written online distance exams. Among other things, it described options for customizing the SOEE and how to proceed in case of technical problems. The SOEE also had text modules that instructors could customize to inform their students about the distance exam, the consent statement for online supervision, the confirmation of the access to the exam, the verification of the identity of the students, the examination statement, and the folder "Exam" containing the exam that only became visible to the students when the exam started (cf. Fig. 1).

In the first exam week in April 2020, exams were only valid if all components were submitted correctly. The unfamiliar situation as well as examination stress led to approximately 7% of the students not giving all consents in time. In most cases, however, examiners set individual grace periods in order to have a valid and complete exam to grade. Some students also had problems setting up the online supervision. The examiners and support units offered very high levels of support, but unfortunately some technical problems could not be solved before the exam started. Some students were therefore unable to take part in the exam.

To improve the clarity of the exam environment and to reduce the number of invalid exams, the first major revision changed the design of the SOEE. A newly developed Pre-Check Wizard for online supervision merged separate tasks, such as the confirmation of access to the exam and the consent to the online supervision. The start page displayed all important information such as the time of the exam if online supervision is used, and the start and end time of online supervision. All unnecessary subheadings in the side menu were removed. Instead, a technical checklist and a link to the appropriate team in MS Teams were added (cf. Fig. 2).

After those revisions, no further changes were made to the structure and contents of the SOEE. Instead, the focus shifted to the activation of online supervision and students' access to the written online distance exams. A newly developed Pre-Check Wizard (cf. Fig. 3) guides students through setting up the online supervision in advance of the exam: enabling the microphone, camera, and screen in the browser.

The examination statement is displayed to the students at the end of the Pre-Check Wizard. If the online supervision is not used, the examination statement will appear at a set date/time (as done in regular online course environments). After having accepted the examination statement, students can access the exam environment and the examination attempt will be accepted and graded.

The second major revision in September 2021 changed the SOEE's technical aspects to better adapt the content and information to the needs of an exam environment and to better distinguish it from regular online course environments in terms of color, structure, and features. The environment's contents are no longer displayed in the side menu, but directly on the start page. The language of the content automatically adapts to the student's language setting.

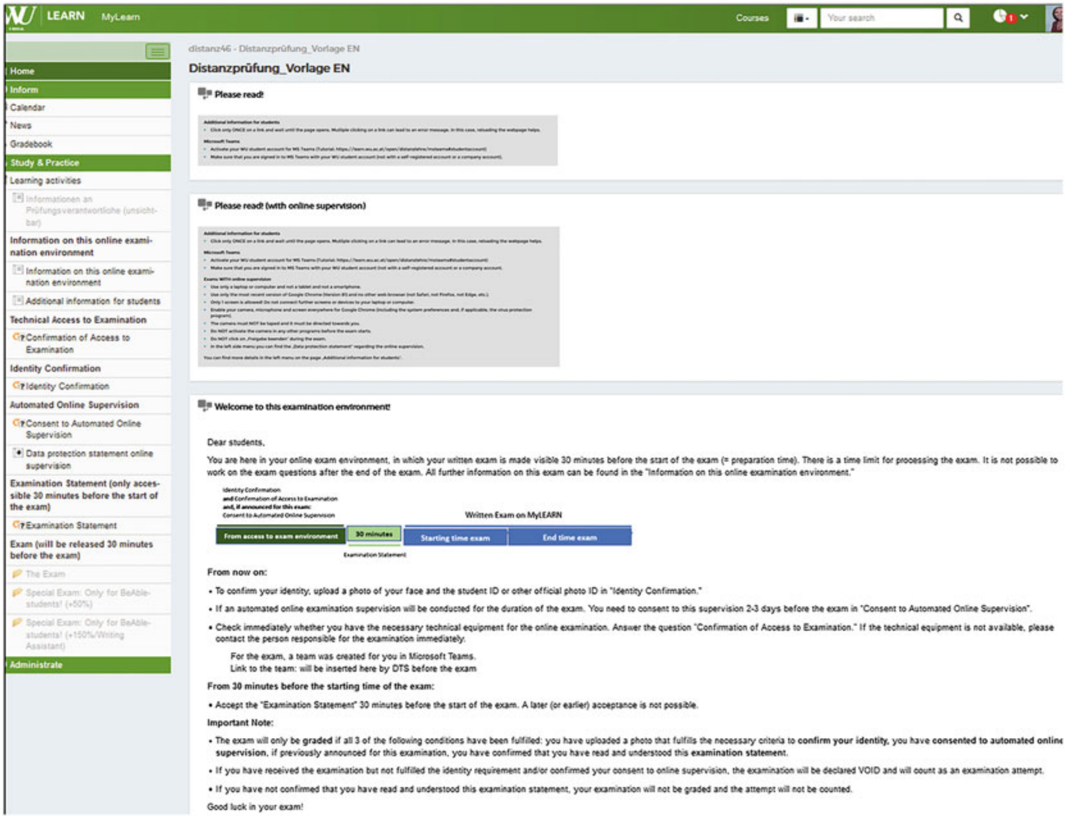


Fig. 1 First SOEE-design. (*Note.* This figure demonstrates what the SOEE looked like as of July 2020. This design was in use from April 2020 until the first major revision in September 2020. Own illustration)

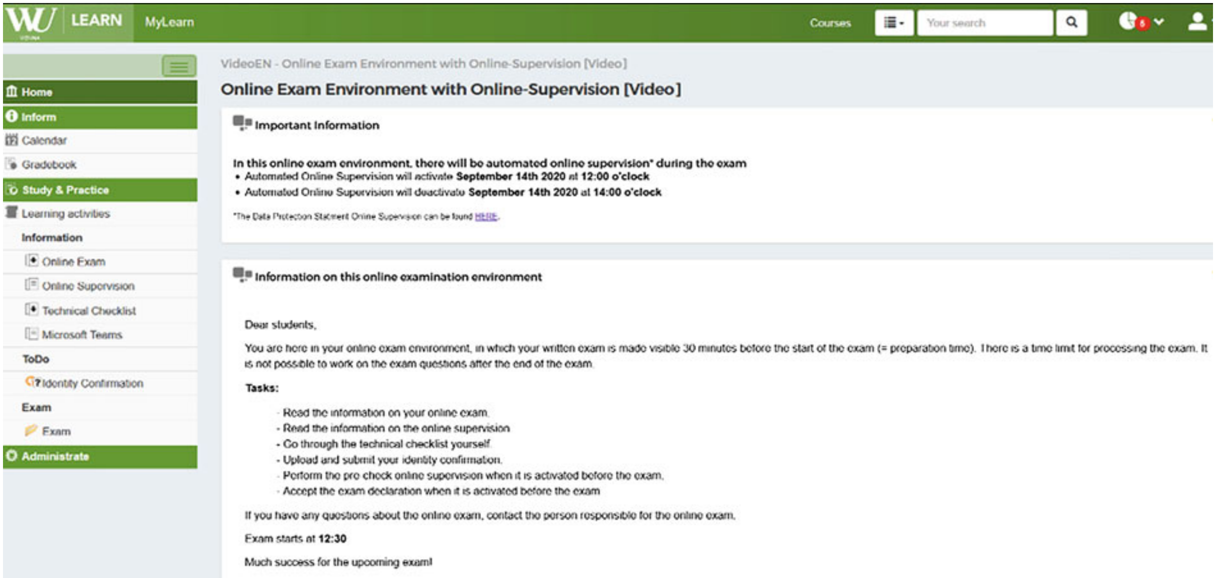


Fig. 2 SOEE-design after the first major revision in September 2020. (*Note.* This figure demonstrates what the SOEE looked like as of July 2021. This design was in use from winter semester 2020/21 until winter semester 2021/22. Own illustration)

Fig. 3 Pre-check wizard for online supervision. (*Note.* This figure demonstrates how students see the Pre-Check Wizard for online supervision. In the picture, the audio wave (red line) indicates an exam participant has successfully granted access to the microphone. Only if this step is successful, the next step for setting up online supervision can be started by clicking on "Next". Own illustration)

For exams within regular course environments, no changes are planned. Medium-term plans are for all exams to be held in the SOEEs.

2.2 Online Exams

The learning management system MyLEARN was designed for teaching and learning settings and not to administer written online distance exams. Accordingly, some adaptations and new features had to be implemented to meet the requirements of formal exams.

In spring 2020, examiners relied almost exclusively on software apps for exercises and exam practice to create written online distance exams. There had been a choice of closed question types, such as multiple-choice and single-choice questions, and cloze, open questions with a text or file submission, as well as some container-elements like learning modules and a sample exam (*Musterklausur*). Closed question types are graded automatically according to a pre-defined scoring scheme. Depending on the exam design, the question types can be combined with each other and, if required, be displayed in a strict sequencing. In order to limit cheating in exams, examiners can create question pools from which different questions are randomly drawn in different orders. In this way, students do not work on the same questions at the same time and this makes it more difficult to compare possible answers. To further limit cheating in exams, the learning management system has implemented some changes such as allowing examiners to determine whether exam questions can be accessed and processed several times during the exam or only once.

Until the first major revision in September 2020, only the existing features of the learning management system could be used to create the written online distance exams. The first major revision introduced in selected pilot exams a software specifically created for examinations. This exam software was originally developed in-house for use in online exams at the WU-campus. However, it can also be used for written online distance exams. In addition to the most common types of questions, the new exam software also offers automatic essay scoring (e.g., automatic highlighting of correct and incorrect terms) and the option to send messages to individual or groups of students during the exam. This exam software offers several advantages for students, especially with regard to usability, such as a better display of exam questions.

After the second revision in September 2021, this exam software has become available to all examiners at WU. Examinations with this new exam software will probably increase until all written online exams will be based on it.

2.3 Administration of Written Online Distance Exams

In the beginning, exam administration was still cumbersome. In spring 2020, examiners were asked to register their exam via the room-booking service and additionally announce all written online distance exams via WU's central ticket system. During this registration, instructors were asked if they wanted to conduct the exam in the regular online course environment or in a SOEE. Using the SOEE required further information such as the language of its content and the name for the online supervision. After the registration was completed, the staff of the Digital Teaching Services manually created the SOEE in MyLEARN. This process was later automatized (Chen et al. 2020). The registration via the ticket system turned out to be too confusing and arduous for examiners and administrative staff. Therefore, a new registration process was implemented.

After the first major revision in September 2020, examiners had to register for a written online distance exam via the WU room-booking system. Part of the registration noted whether the exam should take place in the regular online course environment or in a SOEE. Every day, support staff checked and created the list of newly requested SOEEs. Each SOEE created an accompanying team in the MS Teams program, first manually and later semi-automatically. Examiners received an automated email notification as soon as a SOEE had been created. It contained the link to the SOEE and further instructions such as how to add students as members.

The second major revision in September 2021 changed how examiners had access to the SOEEs. Examiners could now take the initiative to independently create SOEEs. A set-up wizard guides examiners through the development of a SOEE. Within this framework, examiners can now create the contents of the SOEE in a very clear and structured manner.

The following section presents specific support measures that have accompanied the adaptations of the online exams.

3 Support Measures for Online Exams with a Focus on Students

To create specific support measures for target groups using the written online distance exams, WU adopted a three-step-approach.

Firstly, WU's support units identified stakeholders. For the university's support units, the target groups in the area of written online distance exams included instructors, students, and staff who support examiners during exams and who organize and administer exams (e.g., e-tutors, research assistants, and other support staff).

Secondly, WU identified the stable and variable characteristics of the groups and described these groups' specific wishes, problems, and needs. It was obvious that staff of the service unit for teaching coordination and examination offices must be able to administer digital exams. Instructors (as examiners) are usually able to use the technical infrastructure offered by the university. For them it is important to be able to create and assess written online distance exams, and to oversee and design the necessary processes in their responsibility. The students' primary challenges are to take part in written online distance exams and to familiarize themselves with the examination conditions beforehand. Depending on the availability of personal, spatial, and technical resources, the students taking written online distance exams range between:

1. Students with good technical infrastructure and a quiet location to write the exam;
2. Students with poor technical infrastructure and/or without a quiet location to write the exam;
3. Students without a decent technical infrastructure; and
4. Students with the right for support in compensating for a disadvantage.

Thirdly, WU's support units developed an array of needs-oriented, target-group-specific, support measures accompanying the implementation and furthering of written online distance exams.

Because online exams at WU were still in their infancy in spring 2020, students' participation in all written online distance exams was closely examined after each exam week. Among other things, the exam results and type and occurrence of technical problems became the starting point for identifying the challenges for students taking written online distance exams and for implementing improvements. In the first exam week in April 2020, about 3.41% of the students reported technical problems. About half of the problems could be solved before the exam started. The improvements reduced (as of July 2021) the proportion of students with technical problems to a point where it is not surveyed. These improvements can be conceptualized along the following three axes: autonomous self-preparation, calming measures, and compensation for disadvantages.

Students could choose to adopt measures related to their self-preparation. However, calming measures and measures to compensate for disadvantages are embedded in the fundamental structures of administrating written online distance exams.

The following description focuses on the target-group "students". Support measures for examiners are presented in passing because examiners must ensure that exams are set up well for students. Moreover, in many cases, the examiners are also the first point of contact for students encountering problems.

3.1 Autonomous Self-preparation

One of the first implemented and somewhat easiest support measures involves empowering students. It encompasses four measures for an autonomous self-preparation for the online exam and equally addresses all student target groups.

The first measure to empower students was to provide instructions on the specially created website "Online Exams" (URL: https://www.wu.ac.at/en/students/distance-learning-and-online-exams/online-exams). They include texts, downloadable checklists, and videos. The content is adapted every semester and students can refer to it at any time. Examiners can use this website to find out what WU has officially communicated about online exams. This creates a uniform understanding of processes and rules among all stakeholders of online exams and, as a part of it, for written online distance exams.

Examiners benefit from the wide range of information on the subject of "online exams" on the "Distance Learning" website. As of April 2020, this website published information on the exam formats, technical design of online exams, administrative processes, didactic recommendations, and many other topics (cf. Fig. 4).

The website's content was continuously supplemented and completely revised during the second major revision in September 2021.

The second measure to empower students is based on the individual exams. Each SOEE has a text template to describe the specific exam. Examiners use this to write the descriptions of the necessary information for students. This is intended to standardize the level of detail of the information and to ensure coverage of all important topics relating to the written online distance exam. The following topics should be addressed: means of communication, use and rules of the automated online supervision, permissible or necessary aids, optional and obligatory tasks in the SOEE before the exam, details about the exam itself (start and end time, time limits, components, type and number of questions, measures to

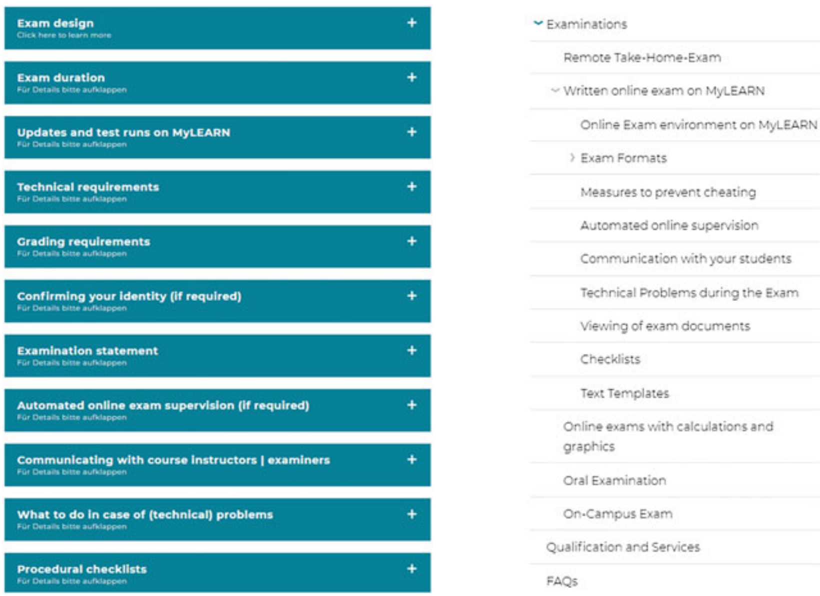


Fig. 4 Target group specific information on online exams at WU website. (*Note.* This figure demonstrates the topics that are covered with regard to online exams on the WU website as of July 2021. Left side: information for students (Website "Online Exams"). Right side: information for examiners (Website "Distance Learning"). Own illustration)

prevent cheating, grading schemes, points total to be achieved), processes related to technical problems, date of publishing the examination results, and organization of viewing exam documents. As a rule of thumb, this information should be available to students as soon as they are added as members in the SOEE. They should be able to view this information at least four days before the exam date. The text template was continuously adapted based on feedback (especially comprehension questions from examiners and students).

The third measure to empower students is an option for self-testing of students' individual setup for camera, microphone, and screen recording independently of any exam. The so-called "Browser Multimedia Test" (URL: https://learn.wu.ac.at/browser-multimedia-test) is a simulation of the Pre-Check Wizard for the online supervision. It allows students to test the settings of their devices at their leisure when preparing for a written online distance exam with online supervision. It automatically lists possible causes of potential error messages. The first major revision in September 2020 added suggested solutions in case of errors.

The fourth measure to empower students offered guided test runs for written online distance exams with online supervision. This allows students to become familiar with the technical requirements and the whole exam setting. These test runs are offered just prior to each exam week. Students can add themselves to the test courses independently as members and then – at the set time – find a simulated exam with online supervision. Here, they can get acquainted with all the components of a SOEE including an appropriate team within MS Teams for the test run. This improves upon the Browser Multimedia Test, because the students can also navigate in a SOEE, learn how to deal with various question types, become familiar with MS Teams as a communication channel during the exam, and remain in online supervision for a longer period. Some technical problems (e.g., with the power supply or the internet connection) only become apparent when students try to maintain the connection to the online supervision over a longer period. The SOEE in which those test runs are organized has been continuously adapted to all changes. This allows students to get to know the latest version of the SOEE, which is also relevant for them.

Students with poor technical infrastructure especially need the opportunity to find out before the exam whether their technical equipment is sufficient for the examination setting. If students notice problems, they have time to adjust and, for example, organize a replacement device from their private context, retrofit a webcam, and change the necessary settings. Students without technical infrastructure can contact the examiners and describe their situation. In individual cases, students can write the online exam at a PC workstation at WU or organize individual online supervision, such as by using a video conferencing system. However, no completely satisfactory solution has yet been found for students without a technical infrastructure. Students are expected to be responsible for personally purchasing the equipment necessary for (distance) study and exams.

3.2 Calming Measures

Many students feel stressed, tense, and anxious during the exams period. Measures were sought to help students feel more relaxed during exam time. Three calming measures were implemented, subsequently evaluated, and continually fine-tuned.

The first calming measure refers to the transparency and level of detail of communication. A general finding of the statistical evaluation of the exam weeks revealed that in an online distance exam setting, students have an increased need for information. As with any (even written-on-paper) exam, they want to be informed about the exam time, the achievable score, the type of questions (closed or open), the permitted aids, and the learning contents. Written online distance exams bring up additional questions regarding technology. Students want to know exactly which question types will be used in the exam, how they will have to enter the answers on the computer, how they will need to navigate within the SOEE, which electronic aids are allowed, whether online supervision is planned, and how to prepare for it. They are most interested in the technical requirements regarding hardware, software, and internet connection so they can participate without interruption in a written online distance exam. These concerns shaped all support measures with regard to the autonomous self-preparation. The communication guidelines were continuously adapted according to needs.

The second calming measure refers to students' fear of being left completely alone with all the questions during the written online distance exam. Here, the previously mentioned MS Teams program has been implemented as a communication tool during the written online distance exams. Examiners who conduct their exam in a SOEE can access a specially designed MS Teams environment: the so-called "exam team". Via MS Teams, students can ask an array of questions – depending on the examiners' choices and preferences – with regard to various aspects of the exam. Templates and a script are used to create the teams in MS Teams. Until November 2020, however, it was all manual input und thus cumbersome work for the support units. The contents for the teams in MS Teams are designed in English or German and are always structured in the same way. The channel "General" is moderated and can be used by examiners to send announcements to all students. The channel "Content" is bidirectional and can be used by students to ask comprehension questions about exam content. The "Technical Problems" channel can be used by students either as a unidirectional reporting tool only or for a bidirectional communication to obtain technical support. Technical support is provided by support units if the number of participants in a written online distance exam exceeds 150 and if the WU's online supervision feature is used. Examiners are free to add other MS Teams channels, such as for internal consultation of supervisors during the exam. During the exam, they can ask questions about the exam environment and technology to the support units' staff via many channels (ticket system/help desk, MS Teams call, e-mail). The description of the team channels has been optimized over time. With the first major revision in September 2020, the exam teams in MS Teams were given a semester-specific logo to clearly distinguish them from unrelated teams (cf. Fig. 5).

The third calming measure focuses on the possibility that students may experience technical problems during the exam. When technical difficulties arise, students easily panic during an exam. Technical difficulties especially involve problems that students did not anticipate in advance, mostly a sudden technical failure of hardware, software, or a sudden loss of internet connection. Experience indicates that technical difficulties become a larger problem when working with

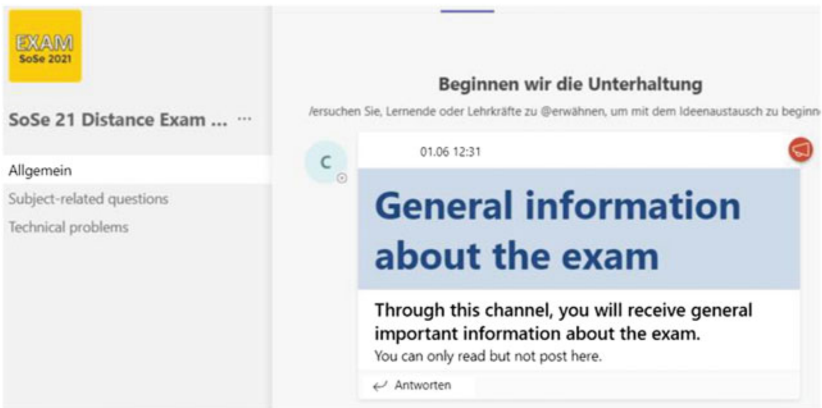


Fig. 5 MS teams team for a written online distance exam. (*Note.* This figure shows the design of a team within MS Teams that accompanies a written online distance exam. Here is the distance exam "Foundations in Socioeconomics" from June 10, 2021. Own illustration)

online supervision. If the online supervision does not work properly, it causes students to be denied access to the exam. If technical difficulties occur more often or for a longer period, this hinders students considerably in the concentrated writing of the online distance exam. After the first major revision in September 2020, a Pre-Check Wizard was developed that provides step-by-step support in setting up the online supervision. This Pre-Check Wizard also provides specific feedback on why the online supervision may have been suddenly terminated. The wizard does not only list the possible reasons for the termination, but it also lists the actions for solving underlying problems (cf. Fig. 6).

Based on the experience gained during technical support, the support units can supplement the catalogue of measures. However, the presence of this Pre-Check Wizard does not eliminate the comfort resulting from a personal technical support. It, however, offers sufficient help for most students to solve their own problems. For the support of other students, support staff can ask about the specific error messages of the Pre-Check Wizard and to point at the already suggested solutions. Only a small percentage of students have technical problems that exceed the resources of technical support, either because none of the proposed solutions solve the problem, or because there is simply no solution in the time available, such as a hardware defect without access to an alternative device. In a similar manner, examiners and support staff have online access to

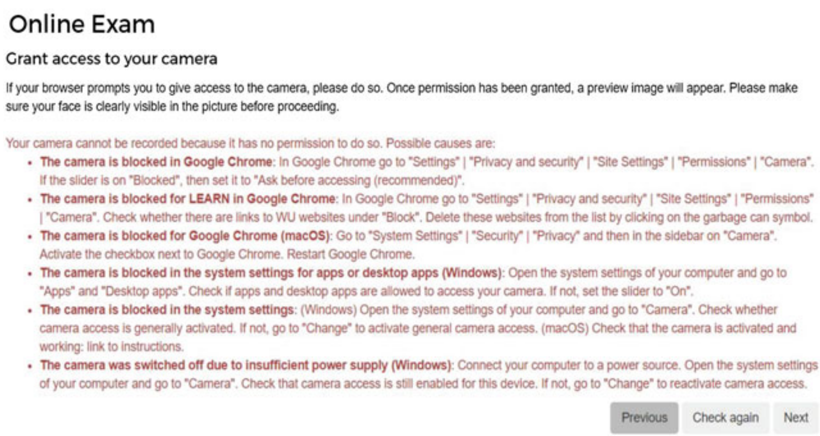


Fig. 6 Suggested solutions for technical problems (pre-check wizard). (*Note.* This figure demonstrates which solutions are offered by the Pre-Check Wizard for online supervision when encountering a problem regarding the access to the camera. Own illustration)

information on possible technical problems and their solutions. Interested parties can access the course content and copy and paste self-help instructions into the MS Teams channel “Technical Support”. This enables examiners without technical expertise to assist students with technical issues should they wish to do so and if no further support by any of the support units could be provided.

3.3 Compensation for Disadvantages

The BeAble program at WU aims to improve the quality of students’ lives in order to compensate for disadvantages that students may have. Students are eligible to participate in the BeAble program if they have special needs and/or chronic illnesses (e.g., physical or psychological impairments, chronic illnesses, or difficulties in reading or writing). The WU measures to compensate for disadvantages only target BeAble students.

With written exams on paper, support has been provided with extended exam time, individual exam rooms, adapted/larger exam sheets, and support via tutors. The transition to online distance exams resulted in the challenge of developing measures appropriate for BeAble students. In order to reach this group of students, two measures were developed and implemented.

The first measure to compensate for disadvantages involved time extensions. In spring 2020, an administrative process began to allow BeAble students to write online distance exams with a time extension of up to 150% greater than ordinary students. While most students only have the standard time available for the exam, some students, for example, can take the exam with a 25% time extension, some students with a 50% time extension and others with a 100% time extension. A feature has been added to the learning management system that allows examiners to assign an exam to one or more people so that only they can see and edit this item (“Individual Assignment”). Thus, an exam can be duplicated and subsequently modified and assigned exclusively to BeAble students (cf. Fig. 7). This procedure has the advantage that most exam takers only see their own (regular) exam and cannot accidentally “click” and enter the wrong online distance exam with the time extension. BeAble students can see two exams: the general exam and their own exam with the time extension. Choosing the correct exam has so far never been a problem.

The second measure to compensate for disadvantages involves, when necessary, a writing assistant to help with or take over the typing of exam answers on the computer. BeAble students usually read the exam information on their own and dictate their answers to the writing assistant. In a few cases, BeAble

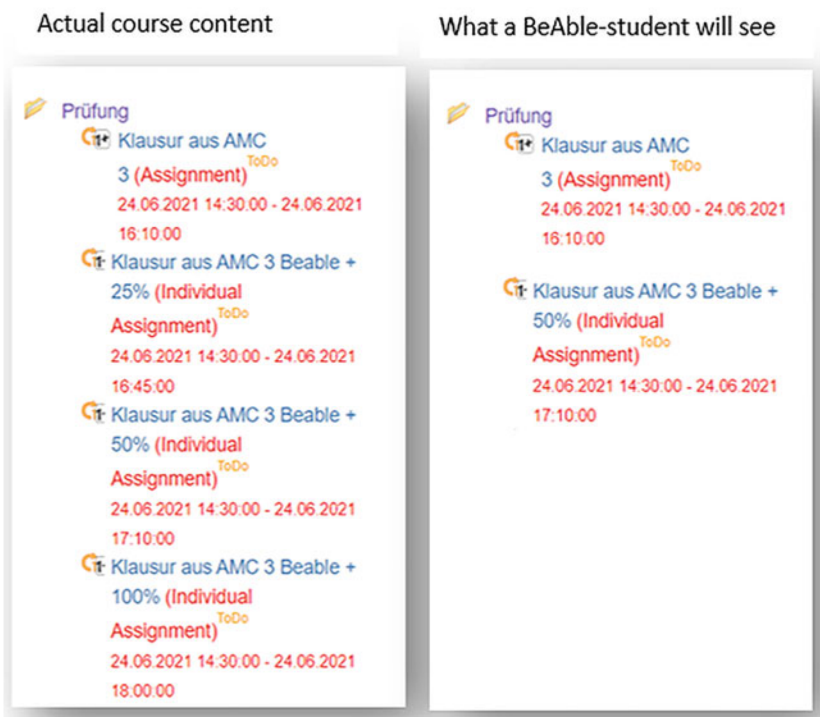


Fig. 7 BeAble exams. (*Note.* The figure demonstrates how course administrators and examiners see the exams within the exam folder (column at the left side) and how a BeAble student with 50% time extension would see the same course (column at the right side). Example taken from the exam “Accounting & Management Control III” from July 24, 2021. Own illustration)

students and writing assistants have met at WU to write the exam there. There was an exemption for these cases even during the strict lockdowns. However, most BeAble students and writing assistants have worked together in a distance setting. The following process was defined to support the collaboration between BeAble students and writing assistants during written online distance exams.

Step 1: BeAble students and writing assistants are entered together as members in the SOEE. Both are assigned the BeAble exam with the appropriate time extension. Both are also entered as members in the general exam team in MS Teams in a private meeting channel.

Step 2: For an exam with online supervision, the writing assistant initiates the online supervision and gains access to the exam. In the appropriate team on MS Teams, the writing assistant shares the screen with the BeAble student in a meeting. The BeAble student activates the webcam in the meeting. During the exam, the BeAble student can see on the screen, what the writing assistant sees in the SOEE. The BeAble students can also read what the writing assistant enters in the answer fields of the exam. In the documentation of the online supervision, the examiners see the writing assistant and, in parallel, the webcam image of the BeAble student superimposed on the screen. The audio recording provides information about what was discussed (cf. Fig. 8).

Step 3: After the exam, the writing assistant's exam submission is graded. This score is then adopted as the grade of the BeAble student.

Although WU has specially equipped rooms for BeAble students, these were not used during the distance examination period. It can be assumed that students with special needs tend to have all the equipment at home.

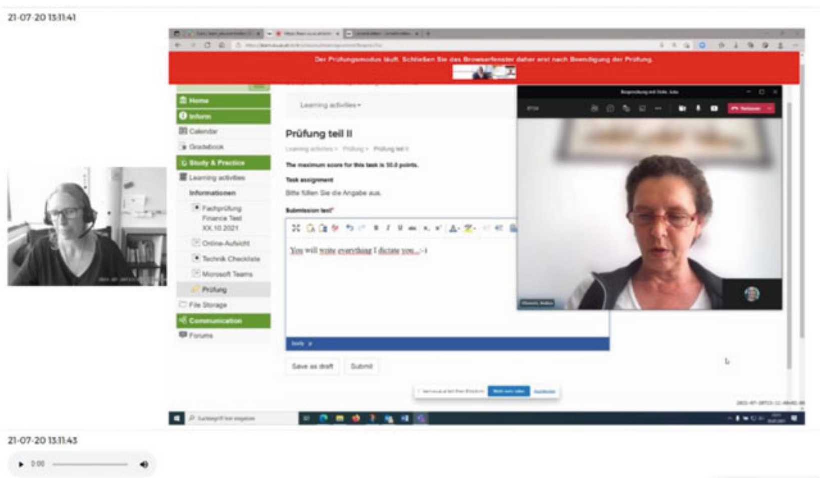


Fig. 8 Documentation of the online supervision in a BeAble exam setting. (*Note.* This figure demonstrates the data collected by the automated online supervision when used in a BeAble Setting. Left side: webcam picture of the writing assistant. Right side: picture of the screen recording including the webcam picture of the BeAble student via an MS Teams team call. Below: Audio recording of the conversation between writing assistant and BeAble-student. The situation here was re-enacted. Own illustration)

4 Conclusion

COVID-19 developments made it necessary for all support units of WU concerned with teaching and learning support to quickly set up a functional and scalable system for written online distance exams. Because the system could not be based on pilot studies, the examination process has been repeatedly evaluated and adapted since the summer semester of 2020 – notably within the two major revision phases in September 2020 and September 2021. From April 2020 until June 2021, 433 online exams have been conducted at WU in the exam weeks.

The SOEEs (separate online exam environments) were revised several times. The SOEE has been designed in a clearer way and the implemented wizard has simplified the activation of the online supervision. The exam application has also been further developed and tailored for exams (not just using applications for learning in an exam mode). An exam software was developed to further professionalize the administration of written online distance exams. The changes in the administrative and technical framework went hand in hand with other support measures that shared the common goal of improving the examination process for all stakeholder groups, i.e., examiners, their supporting staff, and students. The support measures for students encompassed help for an autonomous self-preparation, calming measures, and compensation for disadvantages. The support measures for written online distance exams were well received by instructors and students at WU. Certainly, there is always room for improvement. However, given the short preparation time and the enormous pressure on the support units, the result was satisfactory.

These experiences showed that for online exams to work in a distance setting and thus outside the WU infrastructure, two conditions must be met: the integrity of the exam must not be compromised by the distance setting and students need the appropriate technical equipment and a quiet place to be able to take online distance exams on their own computers. If these conditions are met, then the distance variant can supplement extremely well the campus exams (both paper and digital). Like many universities, WU has ultimately seen the challenge of the COVID-19 pandemic as an opportunity. Online exams or digitally supported exams should always be used when they offer improvements and not only in an emergency. Examples of improvements include greater flexibility in time and place of the exam, more multimedia information, and testing procedures that provide a closer simulation of professional practices. Finally, online exams also offer examiners various benefits such as automated grading and the support in assessing answers in open-ended task formats.

As previously described, guidance and support for all stakeholders in online exams is a key to success. Support measures for students, including those with special needs, are particularly important for written online distance exams.

References

Bandtel, M., Baume, M., Bendenlier, S., Brinkmann, E., Budde, J., Eugster, B., Ghoneim, A., Halbherr, T., Persike, M., Rampelt, F., Reinmann, G., Sari, Z., & Schulz, A. (Eds.). (2021). *Digitale Prüfungen in der Hochschule*. Whitepaper einer Community Working Group aus Deutschland, Österreich und der Schweiz. Version 1.1. Hochschulforum Digitalisierung https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_Whitepaper_Digitale_Pruefungen_Hochschule.pdf.

Bauer, Y., Degenhardt, L., Gilch, H., Kleimann, B., & Wannemacher, K. (2008). *Online Exams as Part of the IT-supported Examination Process Chain*. In D. Sidelmann Jørgensen, M. Hvid Stenalt, & EUNIS VISION IT (Eds.), *EUNIS 2008 VISION IT: Visions for IT in Higher Education: 24–27 June 2008 Aarhus, Denmark*. [no page numbers] University of Århus. https://www.researchgate.net/publication/228744745_Online_Exams_as_Part_of_the_IT-supported_Examination_Process_Chain.

BGBI. I 2002/120. *Bundesgesetz über die Organisation der Universitäten und ihre Studien*. https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20002128.

BGBI. II 2020/171. *Bundesgesetzblatt für die Republik Österreich*. COVID-19-Universitäts- und Hochschulverordnung – C-UHV. ELI: https://www.ris.bka.gv.at/eli/bgbl/II/2020/171/20200422.

Bouchey, B., Gratz, E., & Kurland, S. (2021). Remote student support during COVID-19: Perspectives of chief online officers in higher education. *Online Learning*, *25(1)*, 28–40. https://doi.org/10.24059/olj.v25i1.2481.

Chen, F., Dohr, J., & Mosböck, F. (2020). Schriftliche Distanzprüfungen an der Wirtschaftsuniversität Wien. *fnma Magazin* 02/2020—E-Assessment und E-Examinations, 20–22. https://www.fnma.at/content/download/2087/10778.

Cramp, J., Medlin, J. F., Lake, P., & Sharp, C. (2019). Lessons learned from implementing remotely invigilated online exams. *Journal of University Teaching & Learning Practice, 16*(1), 10.

EUA. (2020). *Survey on "Digitally enhanced learning"—preliminary results*. European Universities Association. http://www.ehea.info/Upload/BFUG_HR_UA_71_8_2_Survey_results.pdf.

Frankl, G., & Schratt-Bittner, S. (2020). Online-Prüfungen an der Universität Klagenfurt. *fnma Magazin* 02/2020—E-Assessment und E-Examinations. https://www.fnma.at/content/download/2087/10778.

Garrett, R., Legon, R., Fredericksen, E. E., & Simunich, B. (2020). CHLOE 5: The pivot to remote teaching in spring 2020 and its impact. *Quality Matters*. https://www.qualitymatters.org/qaresources/resource-center/articles-resources/CHLOE-5-report-2020.

Halbherr, T., Reuter, K., Schneider, D., Schlienger, C., & Piendl, T. (2014). Making examinations more valid, meaningful and motivating: The online exams service at ETH Zurich.

EUNIS Journal of Higher Education IT - Issue 2014/1; Supporting teaching and learning; EUNIS 2014 Congress Issue. http://hdl.handle.net/11366/478.

Ilgaz, H., & Adanır, G. A. (2020). Providing online exams for online learners: Does it really matter for them? *Education and Information Technologies, 25*(2), 1255–1269.

Johnson, N., Veletsianos, G., & Seaman, J. (2020). US faculty and administrators' experiences and approaches in the early weeks of the COVID-19 pandemic. *Online Learning, 24*(2), 6–21.

Kuikka, M., Kitola, M., & Laakso, M.-J. (2014). Challenges when introducing online exam. *Research in Learning Technology, 22,* 22817. https://doi.org/10.3402/rlt.v22.22817.

Ledermüller, K., & Spörk, J. (2020). Lessons Learned aus dem Distanzbetrieb im SS 2020. Ergebnisse aus Studierendeninterviews. Unpublished presentation. Presented at Vienna University of Economics and Business on April 14th, 2021.

Pausits, A., Oppl, S., Schön, S., Fellner, M., Campbell, D. F., & Dobiasch, M. (2021). *Distance Learning an österreichischen Hochschulen im Sommersemester 2020 und Wintersemester 2020/21.* Bundesministerium für Bildung, Wissenschaft und Forschung. https://www.bmbwf.gv.at/dam/jcr:3db6ff5e-68f7-43d0-a31f-0e667d258d69/210701_WF048_21.

Florian Mosböck works at the "Digital Teaching Services" at the Vienna University of Economics and Business. He has been involved in the implementation of the digital exam environment of the university. Additionally he advises and supports the conception and planning of digital exams and offers training courses for teachers. Due to the changes in teaching during the Covid-19 measures, his work has increasingly focused on the realisation and conducting of remote exams

Julia Dohr has been part of the Digital Teaching Services team at WU (*Vienna University of Economics and Business*) since 2018. She supports teachers in the design of the LMS-based teaching and learning space and provides advice, among other things, on video-based teaching and digital exams

Andrea Ghoneim is working with media and learning – as a lecturer, researcher and enabler. To date, she is eDeveloper at Digital Teaching Services, a unit at WU (*Vienna University of Economics and Business*) for support of WU's teaching staff. In the focus of her work are media for collaborative learning, digital support for the assessment of learning achievements and modes of learning and teaching.

E-Assessment of Mathematics and Statistics in Biomedical Sciences

André Heck, Marthe Schut, Michiel van Wijk, Tomas Meijer, and Nataša Brouwer

Abstract

In this case study, we report on the creation and use of cloud-based e-assessment modules in the mathematics and statistics courses for first-year students in the bachelor programme Biomedical Sciences at the University of Amsterdam. These modules include formative and summative parts. They have been developed in a cloud-based environment that enables interactivity through integration of multimedia, randomised examples, and randomised exercises with automated feedback, and that supports randomised formative and summative e-assessment. We present two scenarios for e-learning of mathematics and statistics in the biomedical context: in one scenario, an e-assessment module

A. Heck · M. Schut
Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, Netherlands
e-mail: a.j.p.heck@uva.nl

M. Schut
e-mail: m.f.l.schut@uva.nl

M. van Wijk
Institute of Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, Netherlands
e-mail: m.vanwijk@uva.nl

T. Meijer
College of Life Science, University of Amsterdam, Amsterdam, Netherlands
e-mail: t.t.meijer@uva.nl

N. Brouwer (✉)
Faculty of Science, University of Amsterdam, Amsterdam, Netherlands
e-mail: Natasa.brouwer@uva.nl

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_10

is used alongside a textbook, and in the other scenario all instructional materials are digital. We discuss the use of ICT for support of learning and for empowering the learning experience of students in a design of the continuous e-assessment of mathematical or statistical abilities. This includes the design of differentiated instruction, the creation of e-assessment for learning and of learning, and the monitoring of learning processes and progress of students during the courses. We use the TPACK framework to describe and discuss the experiences of the lecturers about working with online modules and e-assessments, and the specific characteristics of designing these instructional materials.

Keywords

Biomedical sciences • e-assessment • Mathematics • Statistics • TPACK

1 Introduction

Quantitative analysis, modelling, and prediction play increasingly significant day-to-day roles in today's biomedical research. Recommendation #1.5 of the BIO2010 report (National Research Council, 2003) opens with this statement to underpin its advice to educate future biomedical scientists in a more quantitative manner. It recommends that students in life sciences master fundamental mathematical and statistical concepts, and become sufficiently familiar with the elements of programming to carry out simulations of physiological, ecological, and evolutionary processes. They should be adept at using computers to acquire and process data, carry out statistical characterisation of the data and perform statistical tests, and graphically display data in a variety of representations. In addition (recommendation #2), more emphasis needs to be placed on motivating mathematics and statistics and showing how they are used in research.

The updated Guidelines for Assessment and Instruction in Statistics Education College Report (Carver et al., 2016) lists commensurable recommendations for teaching and learning of statistics. Their main advice is: (1) teach statistical thinking, (2) focus on conceptual understanding, (3) integrate real data with a context and purpose, (4) foster active learning, (5) use technology to explore concepts and analyse data, and (6) use assessments to improve and evaluate student learning. In summary, the GAISE report promotes the learning of statistics as scientific inquiry consisting of problem-solving and decision-making processes, rather than a collection of unrelated formulas, methods, and techniques. The goal and desired result of all introductory statistics courses is statistically educated students who have the ability to think statistically, including understanding that variables have distributions. Use of statistical software replaces computations by hand in student inquiry.

The Biosciences Committee of the Royal Netherlands Academy of Arts and Sciences (2011) points at the importance in higher education of 'New Biology'. This can be characterised as a cross-disciplinary science in which omics-based techniques of analysis ('omics' like genomics, proteomics, or metabolomics) and system biology enable researchers to quantify biological processes in and around cells, organs, and organisms. Biomedical students need to get acquainted with systems biology and systems medicine. In mathematical terms, this means an introduction to dynamical systems in the context of biomedical processes (Segel & Edelstein-Keshet, 2013).

The bachelor curriculum of Biomedical Sciences at the University of Amsterdam was reformed in 2017 in line with the recommendations listed in the above reports. Two new courses were created for first-year students: Basic Mathematics and Statistics for Biomedical Sciences (September-December) and Advanced Statistics for Biomedical Sciences (January-March). About 150 first-year students attend these courses each study year and they have diverse mathematical abilities due to different study programmes at pre-university level. In this article we focus on the first course, which starts with a statistics module (September–October) and finishes with a mathematics module (November–December). Each module is structured in course weeks with one 2-h lecture and two 2-h tutorials, and ends with an exam. The course is only successfully completed when a pass mark has been obtained for each module.

The large size and heterogeneity of the student population motivated us to design instruction in which students and lecturers are given useful information on the mastery of course topics during a course, so that they can adjust their learning and teaching strategy in time. Traditional summative assessment, in which the feedback is hardly more than a mark that comes available only after finishing learning, has long been recognised as the most influential factor in shaping what and how students in higher education choose to learn (e.g., Laurillard, 2002; Ramsden, 2003). But its information, even when it is a midterm, comes too late to shape teaching and support learning because insufficient time is left for repairs. Formative assessment of key concepts and skills, which informs students and lecturers about the current competency level and the necessary steps to make progress, serves this purpose much better. Practice testing and diagnostic testing are common forms of formative assessment. They are especially useful during or prior to the learning because the feedback on a test constitutes input for further learning. It is crucial that feedback is promptly available, preferably instantly, and that it is presented in an intelligent way, even if it is just a worked-out solution.

At this point, e-assessment comes on the scene: Without using a computer, it is impossible to render good service to a large heterogeneous group of students through provision of a variety of exercises and instant feedback from formative assessment.

The main characteristic of assessment-driven, examples-based digital instruction is that each step in the learning path begins with a problem that a student is supposed to solve. When a student does not master the step, the e-learning environment gives a hint, offers a worked-out solution, or may split the problem into sub problems. Hereafter the parameter-based implementation of the problem allows the loading of a new version of conceptually the same problem for the student to practise with and to demonstrate the newly acquired mastery. Repeated use of formative e-assessments allows students to independently practise and check their understanding. In this way they can raise their competency level to what is required and otherwise find out where they need more support. For tracking and analysis of formative assessment performance it is necessary to use a data-lossless e-learning environment, i.e., a system that stores all data about every assessment attempt on an assessment database. Then it can also be used to inform lecturers about the progress that students make during their course and about the subjects that need more attention during lectures and/or tutorials than originally planned.

At the Faculty of Science, University of Amsterdam, the cloud-based environment SOWISO (Heck, 2017) is used for e-learning of mathematics and statistics because it meets all requirements for implementing assessment-driven, examples-based instruction in these subjects. Even more, it extends the features for personalised assessment-driven learning with generous possibilities to provide meaningful feedback while a student is completing a task, by not only showing whether the given answer to a question was right or wrong, but also whether it meets requirements regarding the mathematical shape of a mathematical formula or the precision of a numeric answer, and by explaining what went probably wrong in case of an incorrect answer. Instant feedback often allows students to use the online mathematics document as a worksheet in which they progress line by line towards the final solution of a problem, while receiving feedback at each step. An example of a simplification task, taken from the mathematics module, is shown in Fig. 1. The left-hand side of the equations are always given as a prompt and not entered by the student. The student reasons by equivalence, that is, writes line by line equivalent expressions, starting with the given expression and ending with the simplified result in the requested form. Automated, intelligent feedback at each step guides the student's work.

In Sect. 3 we discuss the design and implementation of the task shown in Fig. 1 in more detail. The numbers 2, 4, 5, and 7 in the expression are in fact

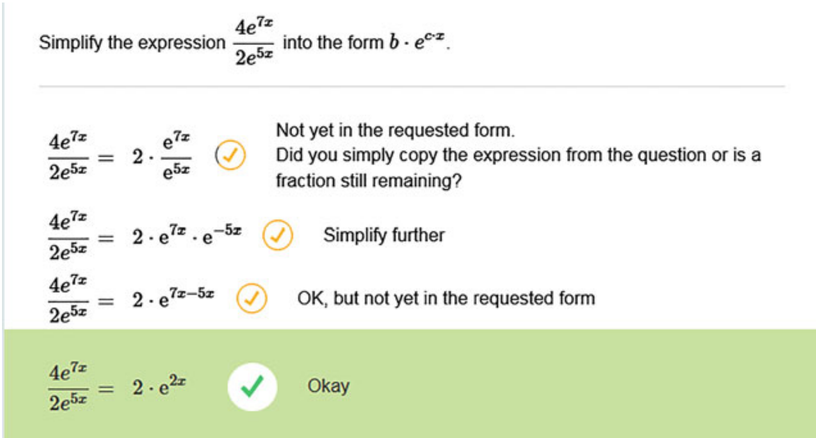


Fig. 1 Use of the online document as a worksheet to solve a simplification task in small steps with intermediate feedback

values of task parameters that are randomly selected when the task is instantiated. This randomisation offers a student an almost endless pool of equivalent exercises to practise with. Almost all exercises in our course are parametrically randomised.

SOWISO uses built-in software engines for creating a task, checking and marking an answer, and generating feedback. Available engines are, amongst others, the computer algebra system Maxima (2020), the dynamic mathematics software suite GeoGebra (International GeoGebra Institute, 2020), and the R programming software (R Core Team, 2020). In addition, SOWISO is a cloud-based multimedia authoring environment for creating interactive mathematical documents, i.e., collections of mathematical pages containing theoretical explanations, randomised examples, randomised exercises, and simulations that can be viewed online in standard web browsers on computers, tablets and smartphones. There is no prescribed hierarchical structure for going through instructional materials: Students may take their own paths through the pages and may study the same page more than once, but they get different examples and exercises each time.

In this case study, we discuss two scenarios for e-learning of mathematics and statistics in the biomedical context. In the first scenario for learning statistics, a statistics textbook about analysis of biological data (Whitlock & Schluter, 2020) is used. In parallel, an e-assessment module in SOWISO is used for tutorials that teach students how to use the R programming language (R Core Team, 2020) and RStudio (RStudio Team, 2020) for doing statistical analysis, and for formative and summative assessment. The ability of SOWISO to use an R engine greatly facilitates the task of generating data with desired statistical properties and its subsequent analysis in RStudio. This approach has allowed us to successfully assess students' ability to perform involved statistical analyses on individual data sets using RStudio. In the second scenario for learning mathematics, a fully online course has been designed on dynamical systems in a biomedical context. In this case, SOWISO is used as an interactive module for learning, practising and assessing mathematics, in which RStudio is used by students to explore dynamical systems in an inquiry-based approach (Heck et al., 2021). Both scenarios rely on availability of good ICT facilities for education: Every student must have good internet access at home and must have a laptop that allows smooth working with discipline-specific software tools. The first requirement is easily met because practically all Dutch homes have broadband internet access. The second requirement is part of the bring-your-own-device policy at the Faculty of Science.

We use the TPACK model (Mishra & Koehler, 2006) as a lens to report on the types of knowledge that the lecturers in both scenarios developed for teaching mathematics and statistics through an assessment-driven, examples-based approach for a large heterogeneous group of biomedical students. We also describe and discuss the experiences of the lecturers about working with online modules and e-assessments, and about the specific characteristics of designing these instructional materials. This includes the design of differentiated instruction, creation of e-assessment of learning and for learning, and monitoring of the learning processes and the progress of students during the modules.

The remainder of this article is organised as follows: In Sects. 2 and 3 we describe the design and implementation of e-assessment in the statistics and mathematics modules, respectively. In Sect. 4 we briefly explain the TPACK model and use it to analyse the development of the lecturers' knowledge about teaching a subject in mathematics and statistics, doing this effectively from a pedagogical perspective, and using technology. We end the article with a discussion of our findings and with concluding remarks.

2 The Scenario for E-Assessment in the Statistics Module

The general goal of all our statistics courses for biomedical students is to let them become statistically educated and to familiarise them with statistical computing through literate programming in the statistics language R. This involves development of statistical literacy, statistical reasoning, and statistical thinking. We follow here the definition of Garfield and Ben-Zvi (2008, p. 34): "statistical literacy involves understanding and using the basic language and tools of statistics," "statistical reasoning is the way people reason with statistical ideas and make sense of statistical information," and "statistical thinking involves a higher order thinking than statistical reasoning and is the way professional statisticians think." The people and professionals we have in mind here are biomedical researchers.

Graph sense is an important component of statistical literacy. It is defined by delMas et al. (2005, p. 2) as "the ability to recognise components of graphs, speak the language of graphs, understanding relationships between tables and graphs, respond to questions about graphs, recognise better graphs, and contextual awareness of graphs." In this definition, graphs sense consists of a wide range of abilities and it not only involves reading and making sense of graphs, but also constructing graphs that best convey information and having a critical attitude toward the use of graphs. This critical attitude is needed because different types of graphs are appropriate for different purposes and one consequently has to make many decisions about how to visualise data. This is promoted in all our statistics courses through learning activities in which students use the R programming language and RStudio to acquire and develop graph comprehension and graph construction. An advantage for the first course is that teaching graphics early makes an attractive entry point to learning the R programming language. In terms of the framework of Shaughnessy et al. (1996), we pay attention to the following graph comprehension abilities:

- *reading data*, i.e., locating presented information and translating it from graphical into another form of communication such as text or a tabular representation;
- *reading between data*, i.e., interpreting and integrating presented information;
- *reading beyond data*, i.e., generating information;
- *reading behind data*, i.e., looking critically at graph use and connecting the graphical information with the context by deep analysis and causal reasoning that is based on subject matter knowledge and experience.

We concur with Nolan and Perrett (2016) that visualisation of data is important not only for final reports, but also as an integral part of data analysis, simulation studies, and thinking with data, and for this reason should play a large role in a statistics course, especially for non-specialists.

Statistical reasoning involves making interpretations based on sets of data, representations of data, or statistical summaries of data, and it combines ideas about data and chance. We pay much attention to this in our statistics courses by providing students with many learning activities to reason about data and come to understand exploratory data analysis (EDA), while learning skills, procedures, and concepts of probability and statistics. EDA, promoted and further developed by Tukey (1977), is the discipline of organising, describing, representing, and analysing data, with a heavy reliance on visual displays and computational tools.[1] In EDA it is all about making sense of data. Data analysis can be viewed in its simplest form as a four-stage process: (a) pose a question and formulate a hypothesis, (b) collect data, (c) analyse data, and (d) interpret the results and communicate conclusions. Statistical software supports EDA by making it possible to quickly manipulate and display data in numerous ways, and to quickly compute statistical summaries of data and statistical tests.

During the statistics module in the first course, each week starts with the lecture in which relevant theory is explained and with studying the associated textbook chapter(s). The formative e-assessment module in SOWISO is used to guide the students in a practical way through the theory, and it teaches the students how the theory can be implemented using the R programming language. To this end each e-assessment exercise contains three elements: (1) a reference to the relevant theory (i.e., a section in the textbook); (2) a link to a chapter in SOWISO that exemplifies how this type of problems can be solved using the R programming language; and (3) feedback through a complete worked-out solution. Each week ends with a small formative test that lacks hints to theory and implementation of R and for which the worked-out answers are only available after the complete test has been submitted.

In this way we give students ample opportunity to acquire knowledge and skills of statistical computing and data analysis, to apply the new abilities to authentic problem situations, and to get acquainted with modern software tools for doing this. In this way we hope and expect to advance students' statistical

[1] EDA actually led to the development of statistical computing packages and the S programming language, which can be considered as the predecessor of R.

thinking, which involves according to Ben-Zvi and Garfield (2004, p. 6) "understanding of why and how statistical investigations are conducted and the 'big ideas' that underlie statistical investigations."

In the assessment-driven, examples-based statistics module we make use of the possibility to embed R programming code within SOWISO for the creation of randomised exercises and randomised examples that take advantage of the statistical functions available within R. This also allows us to create rather complex statistics e-assessments for formative and summative assessment purposes. In exercises, students receive their own unique data set, generated by SOWISO in combination with R, which they are able to download directly into a stand-alone version of RStudio on their own computer, and they perform the necessary operations in RStudio before submitting their answers to SOWISO. In Fig. 2 is shown a screen shot of a statistical task with a computer-generated data set. The creation of the data set with desired statistical properties and presented in comma-separated values (csv) format, the required visual displays, and the correct solution are realised in an R script and not in SOWISO. The SOWISO environment is in this example mainly used for the deployment of the task, the user interaction for providing the answer(s) found, the feedback to the student's solution, and the administration of student results. This approach appeals to lecturers in biomedical sciences who are in general familiar with R.

Technically the question is structured in two parts, namely a SOWISO part and an R part. The SOWISO part contains the question and its solution as presented to the student, together with the definition of the question parameters that will be used to present the question as well as to perform the check and grade of the student's answer and to generate feedback. The R part of the question is an R script that will be executed on every occasion the task is instantiated. An execution of the R script will calculate the question parameters (including the answers) and feeds them back to SOWISO, enabling the environment to present the question and correct solution to the student. The SOWISO-R combination is very powerful for enriching the learning experience of students and for creating meaningful statistics e-assessments.

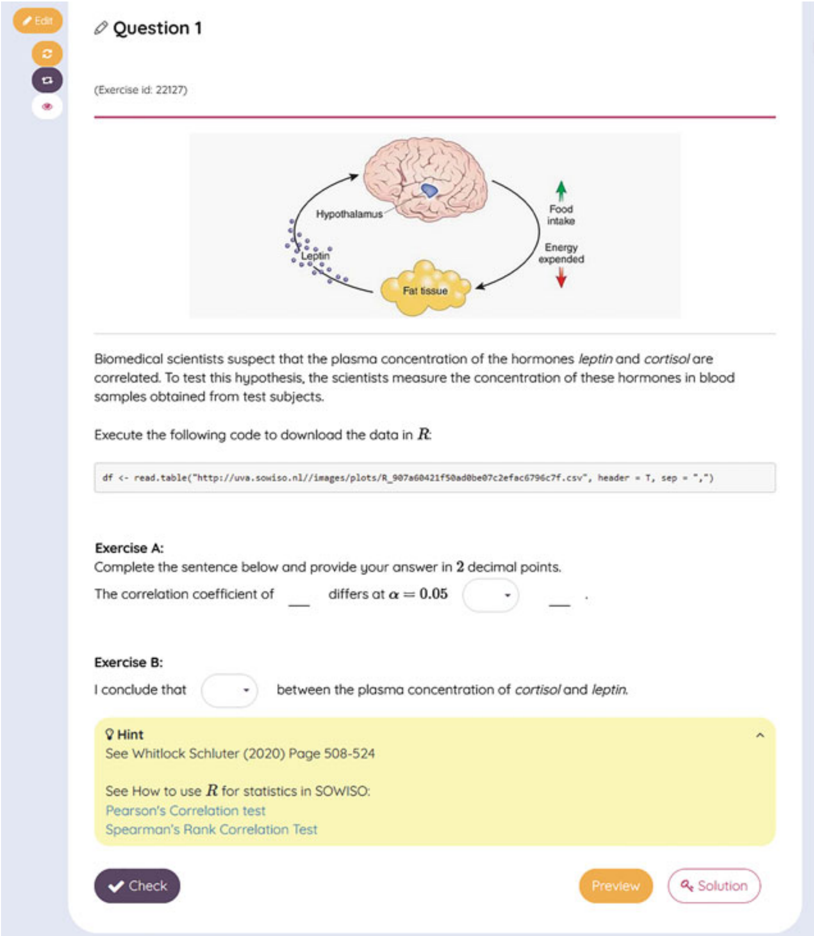


Fig. 2 Exercise taken from the statistics module. Students can directly download a randomised data set in RStudio and after inspecting the data they decide which correlation test is most appropriate. The hint that links the exercise to theory and R examples is only available upon request

3 The Scenario for E-Assessment in the Mathematics Module

In line with the call for teaching 'New Biology' in higher education (Biosciences Committee of the Royal Netherlands Academy of Arts and Sciences, 2011), the new first-year Biomedical Sciences curriculum includes now the study of mathematical models of growth, chemical kinetics, and quantitative pharmacology with the purpose that students see where and how mathematics is applied in biomedical sciences. It prepares students who need advanced mathematical methods and techniques for quantitative modelling of processes of illness and health.

The basic mathematics module lasts seven lesson weeks and one week for self-study before the exam. The focus is on quantitative mathematical modelling, i.e., students explore mathematical models with digital tools using real data and models from research studies. The module gives students in this way an orientation on system biology and systems medicine, where mathematics is a powerful means to explore processes of change in biomedical contexts. The focus is on dynamical systems in biomedical context, with main mathematical concepts like direction field, stability of an equilibrium, bifurcation, and most importantly solving a differential equation algebraically, numerically, and graphically. For the study of dynamical systems, students use the R packages deSolve (Soetaert et al., 2010) and phaseR (Grayling, 2014) so that sophisticated analytical methods beyond the students' mathematical level can be avoided. Also new in this module is the inquiry approach to mathematics teaching and learning (Heck et al., 2021). In this case study, we discuss the use of ICT tools in the mathematics module for formative and summative assessment.

In the mathematics module we use the SOWISO environment for interactive documents containing mathematical text, randomised examples and randomised exercises with automated feedback, and online formative and summative assessments. There is a reference chapter, based on the textbook of Soetaert et al. (2012), in which the basics of R, regression analysis in R, and studying differential equations with R are summarised. It helps teaching assistants prepare R-based tasks and instructions, and students look up short explanations of R use. For learning, we often create and use GeoGebra tools and JavaScript-based simulations in the module. Figures 3 and 4 illustrate tailor-made GeoGebra-based explanations and simulations created with the Easy Java/JavaScript Simulations (EjsS) tools (García Clemente et al., 2017), respectively. These tools can also be used for formative and summative e-assessment.

We use SOWISO to automatically assess students' answers to randomised mathematics questions and give appropriate feedback. To give an impression of

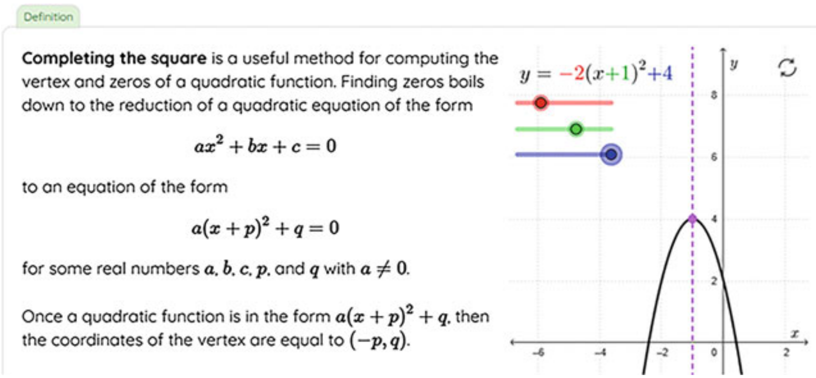


Fig. 3 A GeoGebra tool embedded in a SOWISO theory page

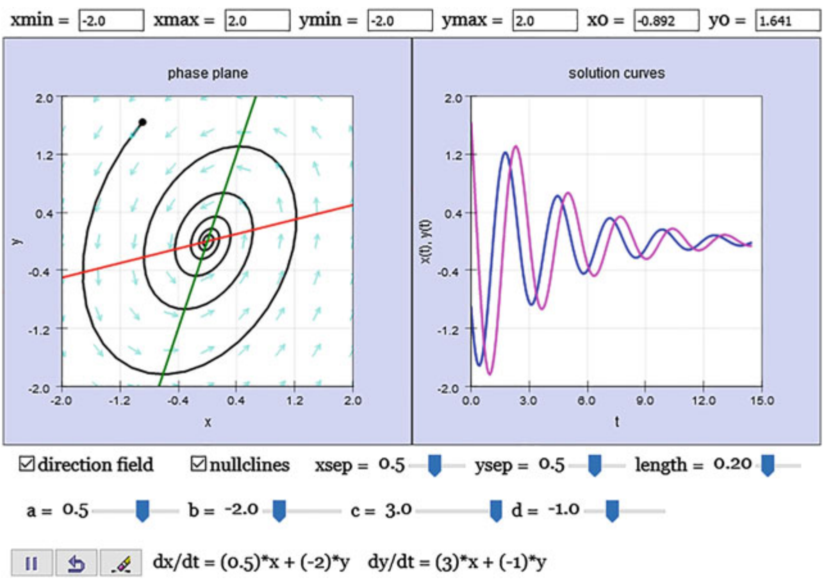


Fig. 4 An EjsS-based tool for exploring solutions of a linear system of two differential equations

how randomisation and instant feedback is realised in our mathematics module, we explain what goes behind the scene in the simplification task shown in Fig. 1. In the shown task, $4e^{7x}/2e^{5x}$ is replaced by the template be^{dx}/ae^{cx}, in which a is a number randomly taken from {2, 3, 4}, b is a multiple of a with a scale factor also randomly take from {2, 3, 4}, c is a number randomly taken from {2, 3, 4, 5}, and d is a random integer greater than c and smaller than 9. Note that the replacements are defined in such manner that no special cases of simplification occur. The requested form can be determined as $f \cdot e^{g}$ with $e = b/a$ and $f = d - c$. After the introduction of variables for randomisation of the problem situation and the definition of auxiliary variables based on the randomisation variables, a complete worked out solution can be written with mathematical formulas formatted in LATEX. Whatever answer a student enters, all that needs to be done is checking whether the answer is algebraically equivalent to the requested final answer, and if so, whether it is in the requested form. The computer algebra engine Maxima is used to check algebraic equivalence. The requested form of the answer has not been entered yet when the student's answer contains a minus sign, a division symbol, or more than one variable x. SOWISO checks the form of the student's answer and returns positive feedback that has been written for each case. When the student enters an incorrect answer because of a common error, such as application of the erroneous rewrite rule $e^{u}/e^{v} \rightarrow e^{u/v}$ the negative feedback informs the student about what apparently went wrong. With the implemented positive and negative feedback, we handle most student answers automatically. It is actually the authors' decision how far they want to go with this feedback, what partial credit is given to a particular wrong answer or to a correct but incomplete answer, and how much effort they actually want to put into detailed assessment of an answer.

Sangwin and Köcher (2016) have shown that the approach of e-assessment outlined above is feasible for a significant proportion of the questions as currently assessed in high-stakes mathematics examinations at the school-university interface. The most significant barrier is the requirement from examiners that students provide evidence that they have used an appropriate method to come to the correct answer, and if not, how far they were on track. We discuss in Sect. 4 how we have tackled this problem in the final exam of the mathematics module using a semi-automatic assessment approach.

The student's workflow and feedback for questions that can be solved by reasoning by equivalence can be generalised in a formative assessment scenario as follows:

1. one of the back engines (Maxima, GeoGebra, or R) of SOWISO generates a question which contains a (possibly randomised) mathematical object such as a data set, a formula, a graph, or a geometrical figure;
2. the student decides what to do next, for instance, use a hint, look at a solution, make an attempt, or go through an example;
3. the student manipulates the generated object, that is, answers the question in an answer field (possibly using an external software tool), drags objects to other positions, or interacts with the figure;
4. a back engine automatically establishes properties of the student's action;
5. outcomes, including feedback, are assigned;
6. the student interprets the generated outcome with feedback and returns to the second step, namely decides what to do next.

An example in which GeoGebra is used for formative e-assessment is shown in Fig. 5. It is a GeoGebra-based exercise for drawing a lineal element at a random point for a randomly generated differential equation. The tool menu of the GeoGebra applet in this exercise was adapted to only contain the necessary tools for drawing a lineal element, selecting and deleting objects. A student can sketch a reasonable approximation of the lineal element to get it marked as correct.

4 Design of E-Assessment Based on TPACK

TPACK (Mishra & Koehler, 2006) stands for Technological Pedagogical Content Knowledge and is a framework that identifies three types of knowledge lecturers need to integrate for successful use of digital technology in teaching and learning, namely technological knowledge (TK), pedagogical knowledge (PK), and content knowledge (CK). In addition, the TPACK framework acknowledges the importance of social and contextual factors and how they impact lecturers' decisions on applying technology in their teaching (Koehler & Mishra, 2009). TPACK comprises seven components (Fig. 6).

The TPACK approach helps lecturers, course designers, and authors of instructional materials design constructively aligned (Biggs & Tang, 2011) blended or online courses. It helps them reach the intended learning outcomes by making relevant choices about the instructional materials, selecting appropriate teaching methods for their courses and choosing those digital tools that best support students' learning and facilitate the content in a pedagogical relevant way, and last but not least by aligning these choices with the institutional context (Brouwer

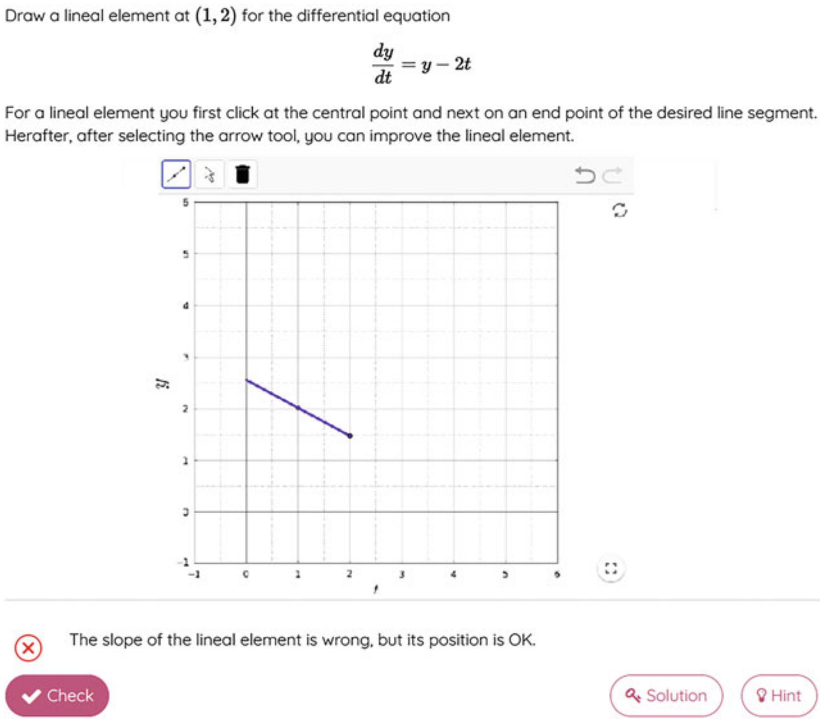


Fig. 5 A randomised GeoGebra-based exercise in SOWISO with automated feedback

et al., 2013). In this case study, we have applied components of the TPACK model in all stages of the development of both modules, namely in the

- use of RStudio in the mathematics and statistics modules (TPK);
- use of SOWISO to create interactive documents for learning, practising, and assessing mathematics and statistics (TCK)
- design of intelligent feedback in SOWISO exercises (PCK, TK);
- adoption of an assessment-driven active learning pedagogy for the modules and the design of online instructional materials and e-assessments, both formative and summative, in line with the adopted pedagogy (PK, TPK, TCK);
- organisation of (online) collaboration between the students (PK, TK);

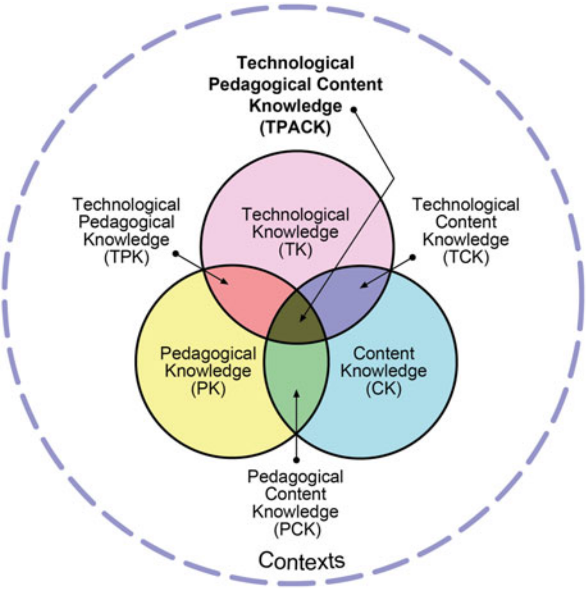


Fig. 6 The TPACK framework, reproduced by permission of the publisher, ©2012 by tpack.org

- alignment in a relevant way with the context of the discipline where researchers use R in their research practice (CK, TK); and
- alignment of the course with the programme curriculum, the educational vision, and the technical facilities (e.g., LMS and digital tools) and support (Contexts) provided at university level.

The modules were meant to have a blended learning course design, having face-to-face lectures and face-to-face tutorials and group work, and asynchronous tailor-made online individual learning and practising. In the situation of a lock-down due to the COVID-19 pandemic, the blended learning modules were transformed in 2020 to online modules while keeping the active learning course design. We explain how this was realised in the mathematics module.

Lectures became synchronous online meetings in Zoom using breakout rooms for small group activities. Explanations were written with a pen tablet during whole classroom discussions. Polls were added to Zoom meetings for engaging students in lectures. To keep structure and prevent chaos in online sessions, at all sessions two lecturers were present. Students stayed muted during the lecture except when asked for a direct reaction, but they could ask questions via chat at any moment. A co-lecturer was continuously monitoring the chat, immediately answered simple questions, and would interrupt the lecturer for answering questions that seemed interesting for the whole class.

Tutorials became MS Teams sessions. Main reasons for using MS Teams instead of Zoom were that students could use this platform also outside the scheduled contact time, have private meetings with peers whom they liked to work with, and could share application screens or digital images with each other (e.g., SOWISO screens) and pass control of applications to others (e.g., giving a TA control over the RStudio environment). The asynchronous tailor-made individual learning part of the module remained unchanged. Each week there were two tutorials. The first tutorial was set up for students to practise (with pen and paper) randomised exercises in SOWISO, entering answers found into the e-learning environment to get feedback. The second tutorial was devoted to learning to work with the R programming language and RStudio in system biology. Each week there were formative e-assessments in which students could check their newly acquired mathematical knowledge and skills by an exercise sequence with minimal feedback before submission, which means that in these formative tests students could only verify whether they had entered a correct or incorrect answer before submission so that they could still improve their work if needed. They could compare their answers with a correct worked-out solution only after submission of the complete assessment.

The summative assessment consisted of a SOWISO home exam with randomized questions under Zoom invigilation. Students worked out open questions with pen and paper, and entered their final answer into SOWISO. Questions in which a student had to find answers by writing R code, always included a final subquestion to copy the created R script into a SOWISO answer field or to upload the file. After submission of the digital exam, a student would scan her/his paperwork into a single PDF file and hand it in via a separate SOWISO test with a single upload question. Teaching assistants and lecturers afterwards checked and marked in SOWISO the answers to questions for each individual student based on the marking scheme that was included in the SOWISO answer model. When an answer had already been automatically marked correct or partial credit had been assigned for a common mistake, skimming through the written student answer was enough; when the answer was not correct, the scanned answer was carefully inspected for partial credit. In our experience, this semi-automatic checking and marking speeds up the process by a factor three compared to traditional processing of exam papers. But what is more important, students can fully inspect their marks after they have been released, and they can complain when they disagree with the marking, or ask questions when in doubt about the correctness of a submitted alternative solution.

We could have opted for asking electronically intermediate questions and avoiding in this way inspection of students' pen-and-paper work. The main reasons for not doing this are: (1) Intermediate questions force students to follow a certain solution path but at the same time reveal part of the solution; (2) Students make mistakes with copying their (possibly correct) final answer or may have trouble with the syntax when entering mathematical formulas online; (3) Students occasionally express a correct answer in an unexpected way. The last two reasons add to the TPACK of lecturers and authors of SOWISO resources. In order to make it clear for the students what is expected from them and to limit student doubts about how to write the answer to a SOWISO question, authors often add pre- and post-input text, and use formula templates when possible. For example, in a question to calculate the Nernst potential of an ion, it is better to start with a prompt "$E_{ion} =$" in front of the answer field and end with a clarification "mV (rounded to 1 decimal)" after the answer field. Such tricks of the trade belong to the author's TPACK and play an important role in the repertoire of methods to assess mathematics in a digital way. For example, matching questions, drag-and-drop questions, and GeoGebra-based questions are more difficult or impossible to arrange in a traditional, pen-and-paper assessment. However, the lecturer must have knowledge about the construction of those questions and be aware of the benefits and the drawbacks of such possibilities in order to fully benefit from the technological affordances of the environment for e-learning and e-assessment. One important reason for inclusion of different question types in a formative assessment is to entice students into practising more with mathematical exercise, not only in terms of a larger number of exercises but also in terms of a greater variety of exercise types. This variety is expected to help students get a better conceptual and procedural understanding of mathematics and become more flexible in applying problem-solving skills. This is part of the lecturer's TPACK as well.

5 Discussion and Conclusion

The TPACK framework allows lecturers to address the challenges and opportunities involved in integrating digital technology in teaching and learning. The SAMR model (Puentendura, 2010) identifies four forms of integration: (1) substitution without changing facilities for learning; (2) augmentation in which digital technology replaces another tool and changes functionality; (3) modification in which digital technology allows for a significant redesign of tasks; and (4) redefinition in which technology allows for creation of new tasks, previously inconceivable. In this case study, we integrated digital technology in our course mainly in the latter two ways.

At university level, one may expect that content knowledge of lecturers is not much an issue. But there are two points to raise here: (1) Statistics courses for non-specialists are commonly taught by practitioners and not by research statisticians, so they may not have deep knowledge of all mathematical details of advanced methods and techniques that they use; (2) When lecturers are out of their comfort zone, like mathematics lecturers who are asked to teach applied mathematics in a field they are not very familiar with, they must learn new content to make fruitful links between mathematics and subject matter. In this case study, for example, the mathematics lecturers had to invest in learning about applications of dynamical systems to neuroscience, medical science, and chemical kinetics.

The incorporation of technology in the course had its own challenges. The lecturers of the mathematics module had ample experience in using SOWISO and mathematical software in mathematics teaching and learning (Heck, 2017; Heck & Brouwer, 2015), but they hardly had experience with working with the R programming language and the RStudio environment. So these lecturers first explored, shortly before the start of the module, how to use R for studying dynamical systems in a uniform and consistent way so that their students would not get lost in the pool of different specialised R packages for studying differential equations. At the same time, they had to design a learning trajectory for their students to use the technology in their confrontation with dynamical systems. Thus, the lecturers were on the one hand developing schemes for use of R in doing mathematics and on the other hand were developing schemes for use in teaching their students how to use R for learning and doing mathematics. The only way the lecturers could proceed was by developing their TPACK on the use of RStudio in mathematics education while they were teaching. The lecturers of the statistics module on the other hand had much practical experience with using R in biomedical data analysis, but they taught statistics for the first time, had no prior experience as SOWISO authors, and were challenged by the fact that gradually more R functionality came available in SOWISO during the development of the module. The latter point means that TPACK development of the statistics lecturers took place under changing circumstances. TPACK is by definition a dynamic framework, but in this case the dynamics of the learning environment amplified this aspect. In addition, there were only a few case studies on e-assessment in statistics education using an R server as a back engine from which they could learn: WebWork (Cubranic et al., 2014) and DEWIS (Gwynllyw et al., 2015; Weir et al., 2021) were the only known comparable e-environments with this functionality.

Although the lecturers in both modules were aware that the use of digital technology in teaching and learning is not easy, they were still a bit surprised by the students' difficulties with programming in R and working with RStudio. The use of the R programming language in the classroom is especially complex and stress-inducing for students with little or no programming experience (Gomes & de Sousa, 2018). As most students are familiar with user-friendly apps with functions easily accessed by one click, the need to type in the statistical commands may, at first, cause some discomfort. An instrumental approach to digital tool use in mathematics education (Trouche, 2020a, 2020b) helps us understand these difficulties. Whereas the lecturers considered the R tasks as opportunities for students to explore mathematical and statistical concepts, the students were in fact still coming to grips with the use of R and RStudio as tools to carry out tasks rather than tools for learning mathematics and statistics. Many students were not yet far enough in the so-called process of instrumental genesis, i.e., in developing suitable utilisation schemes and techniques to transform tools as artefacts into instruments suitable for a task or activity. In order to help students in their instrumental genesis, lecturers discussed during lectures and tutorials literate programming, wrote code live to explore data, and in this way let students observe coding by an expert, the lecturer, and hear an expert's rationale for the choice of a statistical or mathematical technique and computational tool. Nevertheless, the lecturers still have a great task in further developing their TPACK. We have finished the third year of blended instruction and e-assessment in the way we have described before and are ready to reflect on whether our approach is successful or not and can be used as a model for other courses in life sciences as well. The performance, experiences and well-being of our students are important to us and we consider it encouraging that the pass rate for the course is 70% and that students express in evaluations positive opinions about the instructional design of the course, the quality of the instructional materials, and the use of SOWISO, even though they find it is a challenging course because of the use of R and RStudio. We note that the majority of students are more engaged with the subject matter through the formative e-assessments than in a traditional educational setting and that students who practise more, get higher marks. Although this case study has focused only on two slightly different modules in one course for first-year students in biomedical sciences, a similar e-assessment approach has been used successfully in the advanced statistics course for biomedical students and for courses in other disciplines at the Faculty of Science. We are confident that the success of our e-assessment approach is transferable to all disciplines in which mathematics and statistics play an important role. The COVID-19 pandemic has triggered interest in e-assessment at many higher education institutions and has led to an increased usage of e-assessment.

Acknowledgements The authors thank the teaching assistants who helped in the instructional design and implementation of the modules. Special thanks go to Ebo Peerbooms, Jurre van Tiel, Julia Tuinman, and Marouscha de Vries for all their work on the mathematics modules; we truely formed a community of inquiry.

References

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3–15). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2278-6_1.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press-McGraw-Hill Education.

Biosciences Committee of the Royal Netherlands Academy of Arts and Sciences. (2011). *Strategische verkenning Nieuwe Biologie [Strategic foresight on New Biology]*. Koninklijke Nederlandse Akademie van Wetenschappen. https://knaw.nl/nl/actueel/publicaties/strategische-verkenning-nieuwe-biologie.

Brouwer, N., Dekker, P., & van der Pol, J. (2013). *E-learning cookbook. tpack in professional development in higher education*. Amsterdam University Press. https://doi.org/20.500.12657/31178.

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., & Wood, B. (2016). *Guidelines for assessment and instruction in statistics education college report 2016*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf.

Clemente, F., Esquembre, F., & Wee, L. (2017). Deployment of physics simulation apps using Easy JavaScript Simulations. *Proceedings of the 2017 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1093–1096). https://doi.org/10.1109/EDUCON.2017.7942985.

Cubranic, D., Dunham, B., & Kim, D. (2014). On-line homework in probability and statistics: WEBWoRK incorporating R. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9)*. ISI/IASE. http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_C178_DUNHAM.pdf.

delMas, R., Garfield, J., & Ooms, A. (2005). Using assessment items to study students' difficulty reading and interpreting graphical representations of distributions. In K. Makar

(Ed.), *Proceedings of the fourth international research forum on statistical reasoning, literacy, and reasoning (on cd)*. University of Auckland. https://www.causeweb.org/cause/archive/artist/articles/SRTL4_ARTIST.pdf.

Garfield, J., & Ben-Zvi, D. (2008). Developing students' statistical reasoning: Connecting research and teaching practice. *Springer Verlag*. https://doi.org/10.1007/978-1-4020-8383-9.

Gomes, D., & de Sousa, B. (2018). Teaching with R—A curse or a blessing? In M. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the tenth international conference on teaching statistics (icots10)*. ISI/IASE. http://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_9C2.pdf.

Grayling, M. J. (2014). PhaseR: An R package for phase plane analysis of autonomous ODE systems. *The R Journal, 6*(2), 43–51. https://doi.org/10.32614/RJ-2014-023.

Gwynllyw, R., Weir, I., & Henderson, K. (2015). Using DEWIS and R for multi-staged statistics e-assessments. *Teaching Mathematics and its Applications, 35*(1), 16–26. https://doi.org/10.1093/teamat/hrv018

Heck, A. (2017). Using SOWISO to realize interactive mathematical documents for learning, practicing, and assessing mathematics. *MSOR Connections, 15*(2), 6–16. https://doi.org/10.21100/msor.v15i2.412.

Heck, A., & Brouwer, N. (2015). Digital assessment-driven examples-based mathematics for computer science students. In N. Amado & S. Carreira (Eds.), *Proceedings of the 12th International Conference on Technology in Mathematics Teaching* (pp. 403–411). University of Algarve. 10400.1/6081.

Heck, A., Schut, M., & Brouwer, N. (2021). Design and implementation of an inquiry-based mathematics module for first-year students in biomedical sciences. In I. Gómez-Chacón, H. Hochmuth, B. Jaworski, J. Rebenda, J. Ruge, & S. Thomas (Eds.), *Inquiry in university mathematics teaching and learning: The PLATINUM project*. Masaryk University Press. https://doi.org/10.5817/CZ.MUNI.M210-9983-2021-12.

International GeoGebra Institute. (2020). *GeoGebra*. https://www.geogebra.org.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60–70. https://citejournal.org/volume-9/issue-1-09/general/what-is-technological-pedagogicalcontent-knowledge.

Laurillard, D. (2002). Rethinking university teaching: A conversational framework for the effective use of learning technologies. *Routledge/Falmer*. https://doi.org/10.4324/9781315012940.

Maxima. (2020). *Maxima, a computer algebra system. version 5.45*. https://maxima.sourceforge.io.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

National Research Council. (2003). *BIO2010: Transforming undergraduate education for future research biologists*. The National Academies Press. https://doi.org/10.17226/10497.

Nolan, D., & Perrett, J. (2016). Teaching and learning data visualization: Ideas and assignments. *The American Statistician, 70*(3), 260–1054. https://doi.org/10.1080/00031305.2015.1123651

Puentendura, R. (2010). A brief introduction to TPCK and SAMR [web log post]. http://www.hippasus.com/rrpweblog/archives/2011/12/08/BriefIntroTPCKSAMR.pdf.
R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org.
Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). Routledge/Falmer. https://doi.org/10.4324/9780203507711.
RStudio Team. (2020). *RStudio: Integrated development environment for R*. RStudio, PBC. https://www.rstudio.com.
Sangwin, C., & Köcher, N. (2016). Automation of mathematics examinations. *Computers & Education, 94,* 215–227. https://doi.org/10.1016/j.compedu.2015.11.014.
Segel, L. A., & Edelstein-Keshet, L. (2013). A primer on mathematical models in biology. *SIAM, 10*(1137/1), 9781611972504.
Shaughnessy, J., Garfield, J., & Greer, B. (1996). Data handling. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 205–237). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-009-1465-0_7.
Soetaert, K., Cash, J., & Mazzia, F. (2012). Solving differential equations in R. *Springer Verlag*. https://doi.org/10.1007/978-3-642-28070-2.
Soetaert, K., Petzoldt, T., & Woodrow Setzer, R. (2010). Solving differential equations in R: Package deSolve. *Journal of Statistical Software, 33*(9), 1–25. https://doi.org/10.18637/jss.v033.i09.
Trouche, L. (2020a). Instrumentalization in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (2nd ed., pp. 392–403). Springer Verlag. https://doi.org/10.1007/978-3-030-15789-0_100013.
Trouche, L. (2020b). Instrumentation in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (2nd ed., pp. 404–412). Springer Verlag. https://doi.org/10.1007/978-3-030-15789-0_80.
Tukey, J. W. (1977). *Exploratory data analysis.* Addison-Wesley.
Weir, I., Gwynllyw, R., & Henderson, K. (2021). A case study in the e-assessment of statistics for non-specialists. *Journal of University Teaching & Learning Practice, 18*(2), article 05. https://ro.uow.edu.au/jutlp/vol18/iss2/05/.
Whitlock, M. C., & Schluter, D. (2020). *The analysis of biological data* (3rd ed.). Freeman.

André Heck earned MSc degrees in mathematics and chemistry, and a doctoral degree in mathematics and science education. He is senior lecturer at the Faculty of Science of the University of Amsterdam. His research area is the application of ICT in mathematics and science education, including digital assessment-driven examples-based mathematics teaching and learning. He has collaborated in many national and international higher education innovation projects, in particular about inquiry-based STEM education, ICT in STEM education, and learning analytics.

Marthe Schut graduated in physics. She is lecturer at the Faculty of Science of the University of Amsterdam. Her work includes the design and teaching of mathematics deficiency

courses for pre-university students, teaching mathematics to life science students, and teaching laboratory work to physics students. She has participated in a EU project on inquiry based learning of mathematics.

Michiel van Wijk holds a Ph.D. in biology. His research focuses on the perception of, and the behavioural response to chemical signals in biological systems. He is the programme director of the Science and Society BSc programme at the University of Amsterdam. He is a lecturer in statistics, genetics and evolution and behavioural neuroscience.

Tomas Meijer is a lecturer at the faculty of Science of the University of Amsterdam. Tomas has a background in population ecology, modelling brown bear populations in Eastern Europe. Currently he is teaching full time and implementing a coherent statistics track within the Biomedical Sciences track at the University of Amsterdam.

Nataša Brouwer is a senior consultant STEM higher education working at Teaching and Learning Centre Science, University of Amsterdam. She has a PhD in chemistry. Her expertise and research area is in the use of digital technology in teaching and learning in higher STEM education including e-assessment, inquiry-based learning, learning analytics, and the professional development of academic STEM educators. She has collaborated in many national and international higher education innovation projects.

Is Anonymous Aggregated Peer-Evaluation as a Learning Activity Feasible to Identify Differences in Dentistry Students' Clinical Reasoning Performance?

Michaela Wagner-Menghin, Corinna Bruckmann, and Hady Haririan

Abstract

Background: When teachers lack time for evaluation and feedback-giving while developing dentistry students' clinical reasoning skills, peer-evaluation has a double benefit: The evaluat*ees* receive the feedback, which is vital for learning, while the evaluat*ors* have another chance to apply their reasoning skills. Online peer-evaluation tools support anonymous evaluation and increase fairness by aggregating multiple peer-evaluations. The present study addressed the feasibility of anonymous aggregated peer-evaluation by (1) piloting an online-learning scenario and (2) exploring the extent to which this evaluation approach replicated teachers' evaluations.

M. Wagner-Menghin (✉)
Department of Psychiatry and Psychotherapy, Clinical Division for Social Psychiatry, Medical University of Vienna, Vienna, Austria
e-mail: michaela.wagner-menghin@meduniwien.ac.at

C. Bruckmann
Clinical Division of Periodontology, Medical University of Vienna, Vienna, Austria
e-mail: corinna.bruckmann@meduniwien.ac.at

H. Haririan
Department of Periodontology, Dental Clinic, Faculty of Medicine, Sigmund Freud University, Vienna, Austria
e-mail: Hady.haririan@med.sfu.ac.at

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_11

Approach: Undergraduate dentistry students documented the diagnosis and a treatment plan for a virtual patient in a doctor's letter, which was anonymously reviewed by 5-8 peers using two rubrics (from 0 to 3 points). In an experiment, letters with erroneous and correct diagnoses and treatment plans, according to teachers' benchmarks, were peer-evaluated together with letters submitted by students.

Results: Students considered the learning scenario as informative. The peer-evaluations indicated performance differences to be less pronounced than the benchmarks, because of peer-evaluators' bias towards passing scores.

Conclusion: In clinical reasoning training, anonymous aggregated peer-evaluation is a feasible formative assessment activity, but should not be used as a summative assessment, as peers hesitate to indicate weaknesses when this translates to a 'failed' decision.

Keywords

Clinical reasoning skills training • Formative assessment • Moodle™ workshop activity • Peer-evaluation • Undergraduate dentistry students

1 Introduction

To be well prepared for clinical work in dentistry, practicing dentists require medical as well as dentistry knowledge to assess signs and symptoms for proper diagnosis. Further, treatment options should be considered, and practical skills for treating patients, such as assessing a patient's problems, should be developed. Most of all, practicing dentists need to be able to integrate their knowledge with a patient's problem to provide state-of-the-art care for the patient. For dentistry students new in the clinic, the need to integrate theory and practice for clinical reasoning is especially challenging. Students struggle to integrate a patient's clinical findings with diagnostic criteria and arrive at correct treatment recommendations. Furthermore, students may find it difficult to justify their diagnostic and treatment decisions in discussions with peers and clinical supervisors.

Today's educational approaches suggest introducing students to clinical reasoning early in their medical curriculum. Learning scenarios targeting the acquisition of knowledge should be complemented with tasks requiring the application of this knowledge (Krathwohl, 2002) to real-life problems of suitable complexity (van Merriënboer et al., 2003). Only the repeated, deliberate application of knowledge facilitates the development of a detailed mental map of a disease, the so-called 'illness script' that can be navigated quickly and easily or recognised as a pattern when one encounters a patient with matching characteristics (Gavinski et al., 2019). Timely and frequent feedback by means of formative assessment is essential to drive learning and accelerate expertise development (Hattie & Timperley, 2007).

Within medical education, simulated learning scenarios requiring students to diagnose and treat so-called 'virtual patients' have been implemented, to give students a safe opportunity to practice clinical reasoning. The presentation of virtual patients and the interaction with them for learning purposes are routinely supported by electronic communication tools (Cook et al., 2010). Scenarios currently implemented include not only highly guided learning activities requiring the student to complete a series of questions relating to the case materials provided (Fischer, 2000), but also less-guided activities such as moderated discussions mimicking clinical rounds (Grasl et al., 2012; Pokieser et al., 2009). The complexity and authenticity of the learning scenario are determined by the tasks and prompts accompanying the virtual patient (Ellaway & Davies, 2011), as well as by the possibility of providing the learner with timely feedback. Capturing learners' decisions by means of a multiple-choice response format may reduce authenticity but also allows for automated evaluation and the timely provision of feedback. With an open-ended response format, the situation in the clinic is mimicked more closely, but providing feedback for the individual student is difficult. The person providing feedback has to read and comprehend the problem-solving attempt to analyse the quality of the solution provided. This is time-consuming for clinical teachers, especially with problems requiring lengthy written elaborations and large student cohorts. Thus, providing individual feedback is often suspended, which is disadvantageous from a learning sciences perspective, as receiving timely, individual feedback is important for self-regulated learning.

As such, tasking students not only with solving a real-life problem but also with evaluating their peers' solutions comes with a double benefit and has already been shown to promote students' learning (Li et al., 2020): The evaluating students are given another opportunity to apply their knowledge, as analysing the quality of a peer's solution for deriving feedback on the solution is itself a valuable learning activity (Krathwohl, 2002). The evaluated students receive feedback on their problem-solving attempts. Engaging students early in their curriculum in evaluating each other is also favourable from a practical clinical perspective. Introducing junior colleagues to their clinical duties and supervising them constitute a well-established routine in clinical practice communities (Egan & Jaye, 2009; Wenger-Trayner & Wenger-Trayner, 2015).

1.1 Feasibility of Peer-Evaluation in Medical Training

Peer-evaluation has been utilised most frequently to evaluate students' essay-writing skills (Li et al., 2020). It has already been proved as generally feasible to provide frequent and timely feedback in the training of basic clinical practical skills (Basehore et al., 2014; Cushing et al., 2011; Krause et al., 2017; Perera et al., 2010), as well as in dental treatment planning (Teich et al., 2015). Performance ratings provided by same-level peers showed high correspondence with performance ratings provided by teachers (r = 0.70–0.85; Basehore et al., 2014). Similar correspondence between teacher and peer-evaluation has been found related to communication skills (Perera et al., 2010). Dental students (Teich et al., 2015) as well as medical students (Basehore et al., 2014) perceive engaging in providing a solution to a given problem as more effective for learning as compared to evaluating a peer's solution, although both groups of students consider themselves well prepared to evaluate others. Combining problem-solving tasks and evaluation tasks stimulated the highest gain in knowledge and skills (Basehore et al., 2014).

Although same-level peer-evaluations have shown good correspondence with teacher evaluation, students doubted their fairness and effectiveness. In one study (Teich et al., 2015) nearly half of the students (44%) considered the evaluation task as 'not helpful' for learning, and 7% stated the task as disrupting the learning process. Based on a 7-point Likert scale (1 = not at all true, 7 = extremely true), fairness of peer-evaluation was rated as only 'somewhat' (mean = 4.42), although students considered themselves more than somewhat able to evaluate their peers' performance accurately (mean = 5.31) and objectively (mean = 5.67; Basehore et al., 2014). This might be a problem since students are more likely to act upon evaluative feedback to regulate their learning, e.g. reviewing areas with weaknesses, only when they accept and trust the competence of the feedback-provider (Eva et al., 2011). In previous studies (Basehore et al., 2014; Teich et al., 2015), a double-blinded evaluation, collecting one evaluation for each performance, was conducted. Students were aware of performance differences within the peer group, but as they did not know who evaluated their performance, they did not trust the evaluation.

1.2 Overcoming Barriers to Implement Peer-Evaluation

Despite its double benefits, peer-evaluation is currently not routinely implemented in the teaching and learning of clinical reasoning, for three possible reasons: First, administering the process, especially the anonymous evaluation, is complex and time-consuming in a face-to-face educational setting. Second, enabling students to evaluate their peers validly requires the development of instructional materials, such as evaluation criteria and scoring rubrics to better communicate to students how to evaluate validly (Li et al., 2020; Rico-Juan et al., 2021). Third, results on how to best foster and maintain students' positive attitudes towards the peer-evaluation procedure are currently inconclusive. Individual studies focussing on students' subjective experiences of peer-evaluation (Basehore et al., 2014; Teich et al., 2015) indicate that non-anonymous feedback might be important to boost students' trust in the feedback received. But studies focussing on the learning outcome of peer-evaluation found higher effect sizes when peer-evaluation was conducted anonymously (Li et al., 2020).

Online learning management systems, such as Moodle™, provide collaborative tools supporting the administrative process of peer-evaluation and addressing doubts concerning the fairness of anonymous peer-evaluation by providing an algorithm to aggregate the ratings of multiple peer-evaluations. However, as experiences in the use of this algorithm are missing, teachers hesitate to make the effort related to implementing a peer-evaluation procedure in their learning scenarios.

Especially in the context of designing virtual-patient-based clinical reasoning tasks, which are already routinely supported by electronic communication tools (Cook et al., 2010), implementing peer-evaluation as a means of feedback-giving might contribute to the design of more effective online-learning activities. To promote the development of authentic and effective online-learning scenarios in the domain of clinical reasoning skills training, we explored the feasibility of combining such a learning scenario with anonymous, aggregated peer-evaluation and describe our experiences.

2 Aggregated Peer-Evaluation: Online Tools and Procedures

The Moodle™ Workshop activity (moodle.org, 2020) supports the administration of anonymous peer-evaluation and provides an algorithm for the aggregation of the ratings of multiple peer-evaluations, collected with user-defined rubrics (AcademicMoodleCooperation, 2017; moodle.org, 2021). Depending on configuration, the tool is capable of handling other forms of peer-evaluations, e.g. non-anonymous evaluations and reviews with free comments; however, these applications are not covered here.

Written assignments (e.g. summarising medical findings, or developing a treatment plan) submitted by a student (e.g. student A) can be evaluated by means of a provided rubric by a specified number of other students (e.g. 6 students: B, C, D, V, W, Y; Fig. 1). In the present example, student A's assignment was evaluated independently by six peers (B, C, D, V, W, Y) using two predefined scoring rubrics (from 0 to 3 points). By the creation of the aggregated peer-evaluation measure (mean = 3.5), the influence of extreme individual ratings, e.g. too lenient or too strict ratings, was reduced, resulting in a more valid and fairer peer-evaluation for student A's assignment. Additionally, this procedure allowed for the determination of how well the evaluator performed by creating a distance measure (see here for details: moodle.org, 2020). As this measure is not covered within this study, it is not further explained here.

		... is rated by ...							
		A	B	C	D	V	W	Y	Aggregated performance rating
Assignment submitted by	A	--	4	2	5	5	4	1	3,50
	B	5	--	6	5	5	5	4	5,00
	C	X	X	--	X	X	X		
	D	X	X	X	--	X	X		
	V	X	X	X	X	--	X		
	W	X	X	X	X	X	--		
	Y	X	X	X	X	X	X	--	

Fig. 1 Example for a peer-evaluation procedure implemented via the Moodle™ workshop activity, with rubrics. (*Note.* Using two scoring rubrics, each covering from 0 to 3 points, the evaluator can assign a total of 0 to 6 points for an assignment)

3 Feasibility of Online Peer-Evaluation in Clinical Reasoning Tasks

We are not aware of studies providing empirical results in the domain of case-based learning in medicine concerning peer-evaluation with the procedure presented above. To further promote the implementation of anonymous, aggregated peer-evaluation in the domain of clinical reasoning in medicine and dental medicine, the following research aims were pursued in this study:

1. Piloting of a learning scenario with a structured, aggregated peer-evaluation learning activity accompanying a virtual-patient-based clinical reasoning task for students in dentistry.
2. Exploring how the aggregated peer-evaluation measure replicates performance differences as established by teacher performance evaluation.

3.1 Piloting the Learning Scenario

The case-based learning scenario designed for this project comprises three sequential learning activities simulating typical clinical activities, such as diagnosing and planning treatments, summarising considerations in writing, and reviewing others' considerations. Two debriefing activities completed the scenario (Fig. 2). All activities are described in the subsequent section. These case-based activities (CBA) complemented the lectures and seminars of a 4-week module (6.4 ECTS). On average, the module included 79 h of scheduled instruction in optional lectures and compulsory seminars and approximately 81 h of independent study time. The three online CBAs were scheduled as one 90 min compulsory seminar-unit (without face-to-face time), with an anticipated amount of ca. 90 min of independent study time. The anticipated average workload for completing the three online CBAs was thus 180 min. Completion of each CBA was possible within the defined run time of each phase (9 days, and 5 days, Fig. 2).

The debriefing activities were scheduled as a face-to-face lecture (45 min) with an anticipated overall workload of 90 min. The CBA's independent seminar work phase started after the 8th day of the module, when (part of) the medical knowledge required for completing the activities and an introduction of how to work on them had already been covered during lectures and face-to-face seminars. As the regulations for the dentistry curriculum require teachers to give a pass/fail

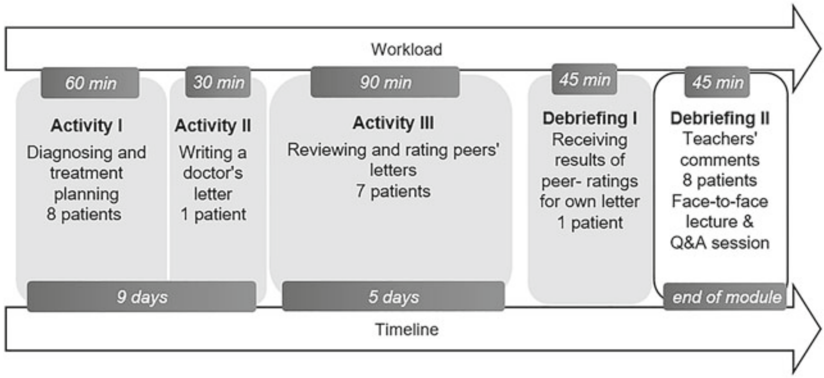


Fig. 2 Timeline, description, and anticipated workload of case-based activities (CBA) within the learning scenario

mark for seminar performance, it was decided by the responsible clinical teachers (HH, CB) to assign a 'pass' mark for overall ratings of 6, 5, and 4. Thus, overall ratings of 3 and lower indicated a 'fail' mark. Upon failing the CBA, the student was required to take an oral exam to remediate the 'fail' and to complete the module.

3.1.1 Description of Learning Activities

3.1.1.1 Activity 1 – Treatment Planning

Students planned the treatment for eight virtual patients (Haririan, 2015). This involved deriving a diagnosis, using the electronically provided patients' files, clinical findings, pictures, and radiological images, as well as coming up with a suitable treatment plan for this diagnosis (see Table 1 for an example).

Patients' files were accessed via Moodle™, and all available patient information for all patients was presented at once, allowing the students to compare and contrast the material during self-study as they wished.

3.1.1.2 Activity 2 – Writing a Doctor's Letter

Students summarised their findings in a doctor's letter for *one* of the cases. To support the writing and to facilitate a fair rating in the subsequent peer-evaluation activity, students were given two prompts resembling the two evaluation criteria (see Fig. 3). The task in principle was designed as a formative assessment

Table 1 Structure of information provided for each patient case

Case	Pain in lower jaw/front
Chief complaint	Male patient, 25 yrs. Pain in lower jaw
Periodontal history	Gingival atrophy/recession […] Dental care […] (how, how often) Professional dental hygiene […]
Personal history	
• Medication	None
• Alcohol/Nicotine/Drugs	Smoker (the last eight years, […]
• Social background/Work	Works as […]; […]
Examination	
• General/nutrition	Inconspicuous; 83 kg/1.83 m
• Radiological examination	Panoramic radiograph Detailed radiograph lower front
• Images	Mouth front/left side/right side Upper jaw/lower jaw Portrait, smiling (partly pixelated for anonymity)
• Basic periodontal examination	Status sheet (scan)

task; however, teachers communicated what they expected students to be able to achieve.

3.1.1.3 Activity 3 – Evaluation Task: Evaluating the Quality of Peers' Letters

Each student anonymously evaluated seven doctor's letters. Two 4-point rubrics with verbal anchors (Fig. 4) were provided to evaluate the letters' quality. Based on the experience-based opinion of HH and CB, it was decided to evaluate two global criteria, 'diagnosing' and 'treatment planning', to keep the procedure simple.

To construct the verbal anchors, HH and CB provided their experience-based opinion on what constitutes an excellent performance (3 points), a good performance (2 points), and a weak performance (1 point or 0 points). The problem of dependencies between the two evaluation criteria – treatment planning requires a diagnosis – was handled by deciding that credits for criterion treatment planning can be earned even when students suggested appropriate treatment for the 'wrong diagnosis' (0 points or 1 point for criterion diagnosing). Teachers' expectations about what students should be able to achieve translated to an overall score of 4.

Summarise your consideration regarding diagnosing and treatment planning in writing. Imagine you are writing a comprehensive doctor's letter. Your text should cover the following aspects:

Based on the findings in the patient's file, which differential diagnosis do you suggest? Justify your suggestion (criterion a).

> *We expect you to identify* ***and justify*** *the main diagnosis as indicated by the patient's findings and diagnostic criteria. Considering differential diagnoses, e.g. discussing why one might or might not consider other main diagnoses or differential diagnoses, improves the quality of your answer.*

Based on your diagnosis and patient's findings, which treatment plan do you recommend for the patient? Justify your recommendation (criterion b).

> *We expect you to recommend, based on your main diagnosis, (a) suitable therapeutic options, including (b) a realistic number of treatment sessions and appropriate therapeutic options. If necessary, adjuvant therapy (no overtreatment!), correctly timed, improves the quality of your answer.*

Note: Credits will be assigned for criterion b, even when the main diagnosis is incorrect.

Fig. 3 Instructions for writing a doctor's letter

To support students' positive attitudes towards peer-evaluation and their accountability for their evaluation, a free comment field labelled as 'overall feedback' was provided together with the scoring rubrics. Additionally, students were informed explicitly that their evaluation would be aggregated with other evaluations to form an aggregated measure for each letter, and they were reminded to justify their evaluation to evaluate fairly.

3.1.1.4 Activities 4 and 5 – Debriefing

To prepare for the debriefing session with their teachers, students received the aggregated, anonymised peer-evaluation, together with the verbal overall feedback. Teachers prepared for the debriefing by reviewing samples of submitted letters together with the submitted peer-evaluations and verbal comments.

Evaluate your peers' letters using the following rubrics:

Criterion a – differential diagnosis

- No main diagnosis has been indicated, or a diagnosis without justification for the main diagnostic criteria has been indicated (= 0 points).
- A diagnosis matching only some of the patient's findings has been indicated and justified; however, this was not the correct main diagnosis (= 1 point).
- The correct main diagnosis has been indicated, and a justification for diagnostic criteria has been included (= 2 points).
- The correct main diagnosis together with the justification criteria has been indicated; other likely diagnoses have been evaluated and correctly excluded (= 3 points).

Criterion b – treatment plan/adjuvant treatment
Note: Assign credits for criterion b, even when the main diagnosis is incorrect.

- A treatment plan not matching the diagnosis was proposed (= 0 points).
- A treatment plan matching the diagnosis in general was proposed, but the number of treatment sessions and/or their timing was unrealistic (= 1 point).
- A treatment plan matching the diagnosis was proposed, including a realistic number of treatment sessions and/or appropriate temporal arrangement. However, either adjuvant therapy that would not be primarily necessary or relevant adjuvant therapy that would be used unfavourably in the temporal course was proposed (= 2 points).
- A treatment plan matching the diagnosis was proposed, including a realistic number of treatment sessions and/or appropriate temporal arrangement. If necessary, appropriate adjuvant therapy was proposed and used correctly in the time course. No overtreatment (= 3 points).

Fig. 4 Verbally anchored rubrics for two evaluation criteria

3.1.2 Qualitative Evaluation of the Learning Scenario

3.1.2.1 Samples and Procedures

Students in dentistry, attending the module of periodontology in the study-year 2015/16 ($n = 92$) at the Medical University Vienna worked on the first version of the activities. Students were welcomed by teachers and informed about the activities in a face-to-face lecture and via e-mail by the 'e-learning event and content management' team. Together with the e-mail information about the start of Debriefing I, students were informed that all tasks had been newly developed for them and were invited to give anonymous feedback to the developers to *"help with improving the case-based-activities for next year's cohort and better address students' needs"*.

3.1.2.2 Materials, Parameters, and Analysis

Feedback was collected via the Moodle™ Feedback activity (moodle.org, 2019) in an open response format as well as via rating scales. Questions addressed the overall quality of the activities as learning material for students (6-point Likert scale, 0 = strongly disagree, 5 = strongly agree), the individual workload for the completion of activities 2 and 3, and suggestions for improvement (text answer). Students' answers were analysed qualitatively, by describing and categorising their content into themes.

3.1.2.3 Results

Thirteen students (14%) volunteered feedback for improvement. All of them '*strongly agreed*' that the activities were suitable as learning material for year 5 students in dentistry. They indicated requiring, on average, 162 min (SD = 63 min) to complete the activity II (write doctor's letter) and 170 min (SD = 217 min) to complete activity III (rate seven peer letters). The following topics emerged in the suggestions for improvement.

Appropriate workload. Workload was addressed by five students, two of whom considered the workload too high. Both stated to have worked 180 min on activity II and 100 min (first student) as well as 420 min (second student) on activity III.

Constructive alignment. Eight students addressed issues of the tasks' alignment within the whole module. Four of them expressed dissatisfaction with the alignment of lectures, which were seen as a prerequisite for the diagnostic tasks. However, in their statements, especially critical students elaborated on the problem of having too little inherent competence to learn clinical reasoning as novices: As novices have not yet organised their knowledge in the form of illness-scripts, they cannot utilise their knowledge effortlessly (Gavinski et al., 2019). They must work through a problem analytically step by step, which is characterised as 'effortful' or as 'not knowing enough'. "*Another point of critique is how we were prepared for doing the tasks. It is difficult to do realistic treatment planning now* [=halfway through the module], *even if I read the materials provided. I think it would be more productive first to complete all lectures and seminars and then do the virtual-patient tasks* [at the end of the module]. *Maybe a guide or checklist can be provided as orientation, to not forget important steps of treatment.*"

Guidance and Instruction. Three students emphasised the need for more guidance for writing the doctor's letter as the current instruction did not result in a uniform strategy. Peers perceived reading of lengthy texts as cumbersome: "… *give an approx. page count (e.g. min–max). I have heard from other students that they rated a work lower because the text was perceived as too long and they did not* *want to read that much*". Especially when the quality of writing was low, workload related to evaluating those letters was considered to be unreasonable. In addition, this problem was aggravated when peers could not decide which entry of the verbally anchored rubric they should assign: "*There needs to be more guidance for writing. It is possible to write very detailed – many colleagues did – many didn't. On the contrary, their elaboration was not well written or structured – however, as they included the correct main diagnosis and relevant differential diagnosis one had to assign full points for the scale, although the text did not deserve this evaluation.*"

None of the students commented explicitly on issues of fairness related to the peer-evaluation procedure.

3.1.2.4 Discussion

A learning scenario for undergraduate dentistry students learning periodontal diagnosing and treatment planning was developed by combining virtual-patient activities with an anonymous peer-evaluation activity. Students were invited to give anonymous feedback to the developers to inform further improvement. Those students (14%) who accepted that invitation were unified in acknowledging the scenario's suitability as a learning procedure for year 5 dentistry students and provided constructive remarks for further improvement. No feedback that included only complaints was submitted. It was thus concluded that the content as well as the procedures of the newly developed learning scenario posed appropriate challenges for students at this level of education.

The subgroup of students participating in the evaluation needed, on average, 332 min to complete the activities, which was considerably more than the anticipated workload of 180 min. At least some of these students considered this effort as too great, which might be problematic with respect to motivation. From a teacher's point of view, however, this result might be welcome. It can be interpreted as having succeeded in creating a learning environment where students feel the need to learn and to perform well. Students likely had to repeat or self-study content covered in the optional lectures to work efficiently on the case-based task. As the module scheduled 28 h of lectures, including another 28 h of independent study time, the excess 152 min of study time needed for completing the three CBA tasks still seemed appropriate from a teacher's point of view.

Given the complexity of the diagnostic and treatment planning tasks, it is not surprising that some students reflected on how to break down the tasks' difficulty further and increase their feeling of competency. For further alleviation of such feelings, teachers might reassure students that the virtual-patient tasks have been designed to allow them to experience safe practice *before* they enter the clinical workplace. It is acceptable to feel insufficiently competent, as diagnosing and treatment planning are difficult, but competence can be learned only by engaging in diagnosing and treatment planning; thus, errors in this phase of training are likely and necessary for development of those skills.

To address issues with guidance and instruction, it was decided to improve the instructions for working on the letter-writing activity. The new instructions give an explicit example of how to structure the letter, provide prompts on the best way to start a sentence, and give a maximum word count/page count (Fig. 5).

3.2 Does Aggregated Peer-Evaluation Replicate Performance Differences?

To study how the aggregated peer performance-evaluation measure replicates performance differences as established by teacher performance evaluation, a field experiment was developed. Specifically, the following questions were researched: Is the aggregated peer-evaluation measure able to replicate performance differences as established by teacher evaluation? Is the aggregated peer-evaluation measure able to replicate the pass/fail decision as intended by teachers?

3.2.1 Method

3.2.1.1 Sample and Procedure

Students in dentistry, attending the module of periodontology in the study-year 2016/17 ($n = 94$) at the Medical University of Vienna completed the CBA piloted in the above chapter as part of their curriculum. Subsequently, students were invited to give feedback to the developers if they saw need for improvement of the CBA.

3.2.1.2 Experimental Manipulation

It was decided to enter letter pairs into the peer-evaluation procedure for better control of case-specificity while manipulating word count and performance differences. For six cases, we constructed benchmark letter pairs with defined performance differences to observe the peer-evaluation procedures' capability of replication. To blend the letters validly with letters submitted by the students, suitable texts illustrating excellent or good performance (6 or 5 points, for cases 1, 2, 4, 8, 6, and 7) were selected from assignments submitted in the year 2015/16. Those texts were structurally modified to comply with the new letter template, and the benchmark was confirmed (HH discussed the benchmark with other clinicians). Subsequently, those excellent (or good) letters were modified to express weak performance (fail with 2 or 3 points, except for case 7, which was evaluated

Summarise your consideration regarding diagnosing and treatment planning in writing. Imagine you are writing a comprehensive doctor's letter to the referring dentist.

Writing your letter, stick to the structure and cover the topics as described here:

Dear colleague,
Thanks for referring patient x.y. to us .... .*Give a brief summary of the patient's anamnestic details in this section.*

The basic **periodontal examination/inspection** (Criterion 0) resulted in …*Give a brief summary of basic findings in this section.* Further diagnostic measures (*insert measures*) showed …*Summarise additional findings (e.g. radiological findings).*

Main diagnosis/differential diagnosis (Criterion 1)

In this section we expect you to (a) name a main diagnosis, (b) justify the main diagnosis as indicated by the patient's findings and diagnostic criteria, and (c) consider differential diagnoses, if necessary discussing why one might or might not consider other main diagnoses or differential diagnoses to further improve the quality of your answer. (--> During Peer Review your answer is rated based on these aspects (Criterion 1).)

The following treatment plan (timing and type of treatment) is proposed (Criterion 2)

In this section we expect you to (a) list suitable therapeutic options, including (b) a realistic number of treatment sessions including suitable therapeutic options. (c) If necessary, adjuvant therapy (no overtreatment!), correctly timed, improves the quality of your answer. (Note: Credits will be assigned for criterion b, even when the main diagnosis is incorrect.)

Because of the **side finding** (insert appropriately) … I request the following treatment prior to periodontal therapy: (*d) If necessary for the case, report side findings (=dental findings, not related to periodontal therapy). (--> During Peer Review your answer is rated based on these aspects (Criterion 2).)*

Kind regards,

Withhold name to enable peer review to be performed anonymously!

How to write:

Copy this template in a new document and save on your PC/Drive.
Write in keywords or full sentences – others need to be able to follow your argumentation.
Delete the instructions before submitting your assignment.
Write a max. of 800 words – your text should not exceed two A4 pages.

Fig. 5 Modified instructions for writing the doctor's letter

Table 2 Characteristics of benchmark letter pairs

Word count of letter pairs	Benchmark performance of letter pairs (performance difference)	Decision	Case*
Benchmark letter pairs controlling for word count while manipulating performance differences			
~220 words	6 vs. 2 points (4)	Pass/fail	1
~330 words	6 vs. 2 points (4)	Pass/fail	2
~440 words	5 vs. 2 points (3)	Pass/fail	4
~490 words	6 vs. 3 points (3)	Pass/fail	8
~800 words	6 vs. 2 points (3)	Pass/fail	6
~800 words	6 vs. 4 points (2)	Pass/pass	7
Benchmark letter pairs controlling for weak performance while manipulating word count			
780 vs. 500 words	3 (0)	Fail/fail	3
540 vs. 360 words	vs. 0 (0)	Fail/fail	5

* *Note.* The variable case could not be manipulated systemically during the field experiment

as 4 points) but without altering the overall structure of the text, thus creating 6 benchmark letter pairs with defined performance differences (4, 3, and 2 points). HH and MWM). Two cases were used to create two letter pairs differing in word count but not in performance, for study of the influence of word count on peer-evaluation. Given the importance of identifying weak performance for steering learning, we decided to use lengthy text modules evaluated as weak performance (0 points, for case 5; 3 points, for case 3, indicating a fail) to construct the letters for this part of the experiment (Table 2).

The benchmark letters were entered into the peer-evaluation procedure together with the students' letters. The peer-evaluation was performed anonymously; letters were evaluated across different study groups. Students were blinded towards the experiment.

3.2.1.3 Peer-Evaluation Measures

The CBA required each student to submit one letter and review seven letters so that all students could work through all the learning objectives relevant for the module. This resulted in the collection of 5 or 6 peer-evaluations per letter for calculation of the aggregated peer-evaluation measure.

3.2.1.4 Evaluation Questionnaire

Feedback was collected by the same procedures as described in Sect. 3.1.2.1.

3.2.1.5 Analysis

To visualise the extent to which the peer-evaluation replicated performance differences within the eight benchmark letter pairs, a bar chart was produced. A scatterplot was created to visualise the replication of the benchmark evaluations as well as the pass/fail decision by the aggregated peer-evaluation measure for each of the 16 benchmark letters. Fisher's exact test was calculated for statistical evaluation of the replication of the pass/fail decision.

3.2.2 Results

3.2.2.1 Replication of Performance Differences

Performance differences within each benchmark letter pair (6 pairs, mean = 3.33; *SD* = 0.75) were partly replicated by the peer-evaluation measure (6 pairs mean = 1.61; *SD* = 0.30). The weaker letter in each benchmark letter pair was identified by the peer-evaluation procedure (cases 1, 2, 4, 8, 7, and 6; Fig. 7, left side). However, the performance difference indicated by the aggregated peer-evaluation was less pronounced, as established by teachers' benchmark evaluation. Within the weak benchmark letter pairs (cases 3 and 5, Fig. 6, right side), the peer-evaluation indicated no or only small performance differences.

3.2.2.2 Replication of the Pass/Fail Decision

To visualise how well the aggregated peer-evaluation procedure replicated the pass/fail decision as established by the teachers, the benchmarks and the aggregated peer-evaluations were plotted for each of the 16 letters entered into the experiment (Fig. 7). Out of seven letters evaluated as 'pass' by the teacher benchmark, six were also evaluated as 'pass' by the aggregated peer-evaluation procedure. Only one letter was falsely evaluated as 'fail'; interestingly, this letter was the shortest letter entered into the procedure. Of nine letters evaluated as 'fail' by the teachers, five were also rated as failed by the aggregated peer-evaluation, but four were not. Short letters' peer-evaluation seemed to replicate teachers' benchmarks more closely as compared to long letters, whose evaluation had a bias towards the 'pass' decision. This bias was most pronounced with the letter pair for case 5, which was evaluated with 0 points by the teachers, but received close to 4 points (= fail) in the short version and close to 5 points (= pass) in its longer version. Fisher's exact test was 0.145 and was not significant, indicating that for the 16 letters overall, the pass/fail decision could not be replicated by the peer-evaluation procedure.

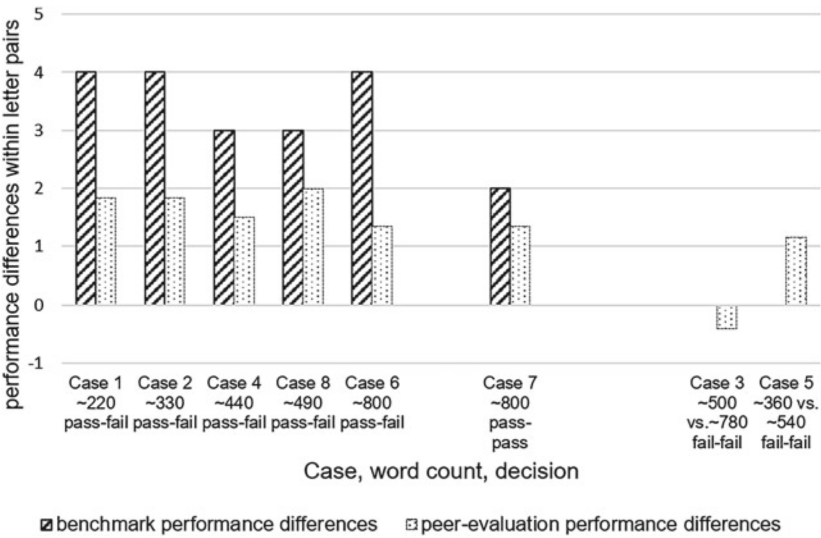


Fig. 6 Replication of the performance differences within benchmark letter-pairs by the peer-evaluation procedure

3.2.2.3 Qualitative Evaluation

Similar to the evaluation during the scenario's pilot phase, 13 students out of $n = 94$ (14%) provided feedback, and all but one '*strongly agreed*' about the activities' suitability for learning. Students indicated requiring, on average, 190 min ($SD = 183$ min) to complete activity II (writing the doctor's letter) and 123 min ($SD = 157$ min) to complete activity III (rate seven peer letters), thus replicating the findings from the pilot study, wherein students needed time to re-study lecture content to work effectively on the case-based activities. Within this cohort, two students commented that the *true workload* was too high, and five criticised the *alignment* of lectures again, similar to the previous cohort wherein the problem of feeling incompetent as a novice was expressed: "*All in all, the [tasks] were a good idea; however, the alignment with lectures and the hand-in date was not optimal. We had the lecture'topic' the same day we had to hand in the assignment. Above that, some of the cases required knowledge we do not have yet. Thus, diagnosing them and evaluating others' performance was really difficult and caused a lot of discussion amongst the students. All in all, this was a good thing, because we have to deal with complex and difficult cases and elaborate our thoughts, but maybe this*

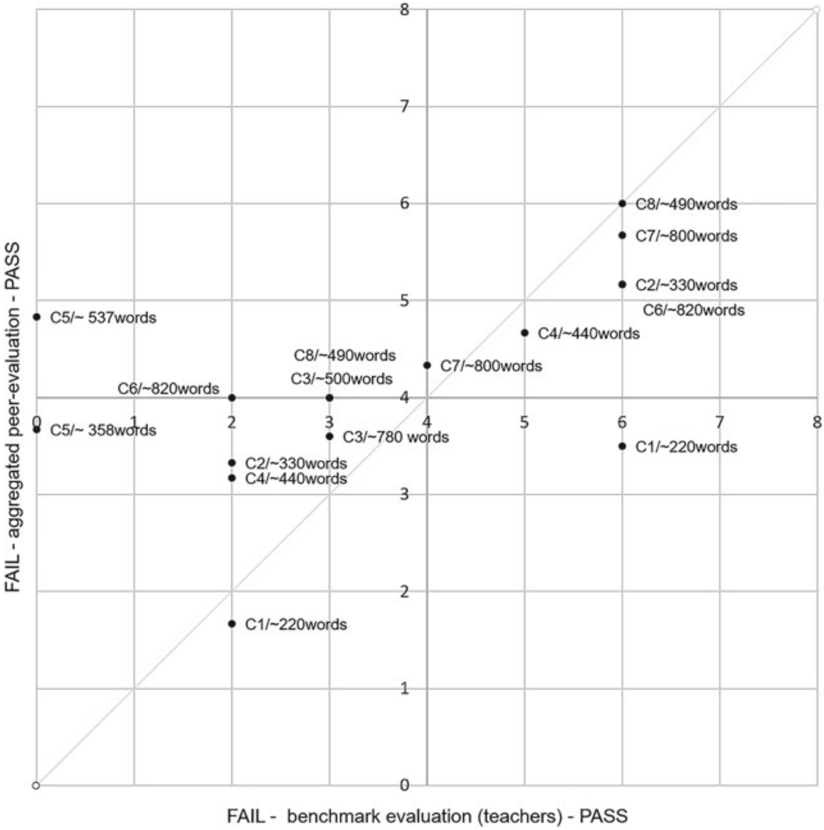


Fig. 7 Replication of the benchmark pass/fail decision by the aggregated peer-evaluation procedure. (*Note.* Dots lying on or close to the diagonal indicate perfect or good replication)

would be more appropriate in a face-to-face seminar." Two students expressed their satisfaction with practical challenges:"… *beyond that, I consider this module as being very well organised! One has to keep learning and I could benefit a lot out of this module. Having to engage in practical clinical tasks within this preclinical module was interesting and informative.*" Nobody commented on the *Guidance and Instruction,* but two students mentioned the *Quality of rubrics*: They suggested working with more evaluation criteria (e.g. include the quality of interpretation of findings), to allow for a more detailed evaluation.

3.2.3 Discussion

Material to be implemented as a field experiment was developed to investigate whether the aggregated peer-evaluation procedure is capable of replicating performance differences and pass/fail decisions in students' performance. Performance differences were replicated, albeit to a lesser extent as defined by the teachers' benchmark evaluation. Pass/fail decisions were only partially replicated, and for weak performances, there was a bias towards a 'pass' decision. Even letters rated as '0' by teachers were given points by peers, especially when the text was generally well written. Students seemed to be generally capable of identifying performance differences but hesitated to indicate them when this would lead to a 'fail' decision for a fellow student. We received no evidence that the experimental manipulation was identified during the run time of the task, which we take as an indicator that our experiment was blinded as anticipated.

As a limitation in this experiment, we should mention that the goal of constructing benchmark letter-pairs mimicking real student letters as closely as possible interfered with the goal of systematic manipulation of text length and performance. To further ensure the integrity of the field experiment, the procedure had to be designed in a way that allowed students to cover all learning goals. Students had to be exposed to all eight cases, but to keep the workload manageable, we were not able to manipulate the cases as factors in the experiment by including more benchmark letter-pairs per case. Future studies with similar tasks may thus skip text length as an experimental factor and may instead put more focus on the types of errors (omission of relevant information in letters vs. wrong information in letters) and the quality of writing in general when refining the benchmark letters.

4 Conclusions and Implications

Providing students with challenging tasks for which they can apply newly acquired knowledge and skills, assessing how they perform, and giving them feedback are vital to help students direct their learning constructively. When time constraints preclude teachers from engaging in formative assessment activities, engaging students in well-designed online peer-evaluation activities is an instructional design option with a double benefit. Peer-evaluators are challenged once more to apply and expand their skills, and evaluatees receive feedback which in turn also promotes learning. This effect has previously been shown to be most pronounced when students have the possibility to act as evaluators during an anonymous, computer-mediated peer-evaluation activity (Li et al., 2020). The use of software to organise the anonymous peer-evaluations adds the possibility for the collection of multiple evaluations per submission for automatic aggregation. Our results contribute to highlighting the feasibility of the peer-evaluation approach (which has previously been used mainly for evaluating writing skills) for evaluating dentistry students' clinical reasoning performance. Further, we provide empirical evidence regarding the feasibility of aggregated peer-evaluation scores.

Using an open-source online tool (Moodle™/Workshop activity), we were able to design and implement a learning scenario for clinical reasoning skills utilising anonymous, online peer-evaluation, with aggregated peer-evaluation scores for students in dentistry. We found the students to be motivated to learn and revise lecture content, as they perceived the task of diagnosing and treating a virtual patient as challenging. Similar to studies of one-on-one peer-evaluation in face-to-face settings (Basehore et al., 2014; Perera et al., 2010) whose authors found peer-evaluation to correspond quite well with teacher evaluation, we found students capable of indicating performance differences as indicated by teachers' benchmark ratings. However, despite using an anonymous peer-evaluation procedure as suggested by previous literature to reduce peer pressure related to peer-evaluation (e.g., Vanderhoven et al., 2015), we found students to be hesitant to indicate weaknesses when this would result in their peers failing. This is especially surprising as the activity was implemented as a formative assessment. It did not contribute to the overall grade for the 4 week module; the consequences for students failing the activity required them to participate in a face-to-face remediation activity, including being quizzed by the teacher. We can only speculate as to why the students in our sample felt the need to spare their peers a weak evaluation. It became increasingly obvious during the conduct of both studies that students considered the entire learning scenario interesting but rather challenging. Our scoring rubric comprising 0 to 6 points for performance evaluation, combined with the teachers' expectation that students should reach 4 points (66% of total points) to receive a pass mark for the learning scenario, might thus have contributed to students' hesitation to indicate weak performance. It might be beneficial if the rubrics were designed to represent lower stakes, allowing for more leeway in indicating student weaknesses while allowing them to stay within socially acceptable boundaries. This might be achieved, for example, by using 10-point rating scales instead of 4-point scales and by communicating a less challenging cut-off for a 'pass' score.

The scoring rubrics and the instructions on how to approach the CBA can serve as templates to construct refined materials for further studies in similar settings. The newly proposed experimental setup, comprising the identification of evaluation bias in the peer group by engaging the group in evaluating benchmark materials, can be further developed to allow for estimating the accuracy of peer-evaluation results.

Implementing anonymous, aggregated peer-evaluation by means of an electronic communication tool is feasible as long as these peer-evaluations have a predominantly formative assessment for feedback function. Still, when students approach the peer-evaluation with a summative assessment function in mind, they are hesitant to indicate weaknesses, especially when they are difficult to justify. Often, formative and summative assessment functions are not clearly distinguished and not explicitly reflected in studies on peer-evaluation (Li et al., 2020). In our example, teachers' expectations of how well students should perform translated into a score that might have been perceived by some students as a summative, 'high stakes' assessment. Together with the notion of having to take an oral exam upon failing to receive this score might have biased students' peer-evaluation.

What should also be kept in mind when implementing peer-evaluation is that students value expert feedback. This includes being examined by teachers. For best results, peer-evaluation should precede teacher evaluation. Frequent online anonymous peer-evaluation activities are feasible to provide students with the training and the formative feedback they need to prepare for a high-stakes face-to-face teacher-evaluated summative exam.

Acknowledgements The authors owe a thank-you to Medical University of Viennas' Moodle-team and the team of *ComputerLernStudio* for helping us to set up the workshop activity and providing the extra technical support necessary to run the experiment. Above all, the studies would not have been possible without our experienced Event and Content Manager, Evelin Gschosmann, who managed the complex procedure in the Moodle Learning Management System. We also would like to express our thanks to all the students who made the effort to give us feedback on how they perceived the activities. Their contributions helped us to keep working on improving case-based learning.

References

AcademicMoodleCooperation. (2017). AMC-Cheat Sheet. Peer-Review mit dem Workshop Modul abbilden. [Organizing Peer-Review with the Moodle Workshop Module]. https://www.academic-moodle-cooperation.org/fileadmin/user_upload/p_aec/Cheat_Sheets/Peer-Review_mit_dem_Workshop_Modul_abbilden-DE.pdf.

Basehore, P. M., Pomerantz, S. C., & Gentile, M. (2014). Reliability and benefits of medical student peers in rating complex clinical skills. *Medical Teacher, 36*(5), 409–414. https://doi.org/10.3109/0142159x.2014.889287

Cook, D. A., Erwin, P. J., & Triola, M. M. (2010). Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Academic Medicine, 85*(10), 1589–1602. https://doi.org/10.1097/ACM.0b013e3181edfe13

Cushing, A., Abbott, S., Lothian, D., Hall, A., & Westwood, O. M. R. (2011). Peer feedback as an aid to learning—What do we want? Feedback. When do we want it? Now! *Medical Teacher, 33*(2), e105–e112. https://doi.org/10.3109/0142159x.2011.542522.

Egan, T., & Jaye, C. (2009). Communities of clinical practice: The social organization of clinical learning. *Health, 13*(1), 107–125. https://doi.org/10.1177/1363459308097363

Ellaway, R. H., & Davies, D. (2011). Design for learning: Deconstructing virtual patient activities. *Medical Teacher, 33*(4), 303–310. https://doi.org/10.3109/0142159x.2011.550969

Eva, K. W., Armson, H., Holmboe, E., Lockyer, J., Loney, E., Mann, K., & Sargeant, J. (2011). Factors influencing responsiveness to feedback: On the interplay between fear, confidence, and reasoning processes. *Advances in health sciences education: Theory and practice, 17*(1), 15–26. https://doi.org/10.1007/s10459-011-9290-7

Fischer, M. R. (2000). CASUS—An authoring and learning tool supporting diagnostic reasoning. In *Use of computers in medical education (Part II)* (Vol. 1, pp. 87–98): Ch. Daetwyler.

Gavinski, K., Covin, Y. N., & Longo, P. J. (2019). Learning how to build illness scripts. *Academic Medicine, 94*(2), 293. https://doi.org/10.1097/acm.0000000000002493

Grasl, M. C., Pokieser, P., Gleiss, A., Brandstaetter, J., Sigmund, T., Erovic, B. M., & Fischer, M. R. (2012). A new blended learning concept for medical students in otolaryngology. *Archives of Otolaryngology-Head and Neck Surgery, 138*(4), 358–366. https://doi.org/10.1001/archoto.2012.145

Haririan, H. (2015). Ein 25 jähriger Patient mit Schmerzen in der Unterkierferfront... Acht virtuelle PatientInnen Fälle zur Illustration parodontologischer Differentialdiagnosen. [A 25 year old patient with pain in the lower jaw... Eight virtual patient cases illustrating periodontal differential diagnoses]. In P. Pokieser, E. Gschosmann, & M. Wagner-Menghin (Eds.), *Fallsammlungen für die Medizinische Lehre. [Clinical Case-Collections for Medical Education]*. Medical University Vienna/Teaching Center/Unified Patient Program.

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

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Krause, F., Schmalz, G., Haak, R., & Rockenbauch, K. (2017). The impact of expert- and peer feedback on communication skills of undergraduate dental students—A single-blinded, randomized, controlled clinical trial. *Patient Education and Counseling, 100*(12), 2275–2282. https://doi.org/10.1016/j.pec.2017.06.025

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193–211. https://doi.org/10.1080/02602938.2019.1620679

moodle.org. (2019, 29th March). Feedback https://docs.moodle.org/39/de/Feedback. Accessed 19 Sept 2021.

moodle.org. (2020, 13th May). Gegenseitige Beurteilung. https://docs.moodle.org/311/de/Gegenseitige_Beurteilung [Workshop activitiy. https://docs.moodle.org/311/en/Workshop_activity]. Accessed 14 Sept 2021.

moodle.org. (2021, 10th August). Bewertungsstrategie bei gegenseitigen Beurteilungen. https://docs.moodle.org/311/de/Bewertungsstrategien_bei_gegenseitigen_Beurteilungen. [Workshop grading strategies. https://docs.moodle.org/311/en/Workshop_grading_strategies]. Accessed 14 Sept 2021.

Perera, J., Mohamadou, G., & Kaur, S. (2010). The use of objective structured self-assessment and peer-feedback (OSSP) for learning communication skills: Evaluation using a controlled trial. *Advances in Health Sciences Education: Theory and Practice, 15*(2), 185–193. https://doi.org/10.1007/s10459-009-9191-1

Pokieser, P., Brandstaetter, J., Moritz, T., Domanovits, H., Sodeck, G., & Fischer, M. R. (2009). Unified Patient Project. Erfahrungsbericht zu web-basierten Falldiskussionen im klinischen Unterricht. *zeitschrift für e-learning, lernkultur und bildungstechnologie, 4,* 49–59.

Rico-Juan, J. R., Cachero, C., & Macià, H. (2021). Influence of individual versus collaborative peer assessment on score accuracy and learning outcomes in higher education: An empirical study. *Assessment & Evaluation in Higher Education*, 1–18.https://doi.org/10.1080/02602938.2021.1955090

Teich, S. T., Demko, C. A., & Lang, L. A. (2015). Students' perception of peer-assessment in the context of a treatment planning course. *European Journal of Dental Education, 19*(1), 8–15. https://doi.org/10.1111/eje.12095

van Merriënboer, J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*(1), 5–13.

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education, 81*, 123–132. https://doi.org/10.1016/j.compedu.2014.10.001

Wenger-Trayner, E., & Wenger-Trayner, B. (2015). Communities of practice a brief introduction. Retrieved from https://wenger-trayner.com/wp-content/uploads/2015/04/07-Brief-introduction-to-communities-of-practice.pdf.

Michaela Wagner-Menghin is an associate professor at the Medical University of Vienna, Austria, Department of Psychiatry and Psychotherapy, Clinical Division for Social Psychiatry. As an educational researcher and (clinical) psychologist, she contributes to research projects on learning clinical reasoning skills, the integrated teaching of clinical diagnostic sciences, and self-directed learning, including self-judgment and feedback.

Corinna Bruckmann is an academic teacher from the University Clinic of Dentistry, Medical University of Vienna, Austria. She has contributed to research in conservative dentistry and periodontology and to teaching in periodontology in the dentistry degree programme as well as in the periodontology postgraduate degree programme.

Hady Haririan is a professor and head of the Department of Periodontology at the Dental Clinic of the Sigmund Freud University Vienna, Austria, after having worked for ten

years as an assistant at the Medical University of Vienna. He is responsible for education in periodontology in the bachelor's and master's programme.

Part II
Exam Formats

Peer Assessment in MOOC of Students Performance with Paragogy Framework: Evidence: Higher Education Institutions in Indonesia

Dessy Seri Wahyuni and Gede Ariadi

Abstract

This study scrutinizes the effect of peer assessment in enhancing students' writing, especially scientific manuscript, in MOOCs (Massive Open Online Courses) in Indonesia. Peer learning theory through peer assessment was chosen to evaluate the student's skill in writing the scientific manuscript. Thirty undergraduate students majoring in technology and vocational education were involved in this study. All projects in this course were performed by ten groups of three students. Their scientific proposals were reviewed based on the following aspects: normativity, objectivity, and logicality. The result showed that the students' writing was enhanced after the peer evaluation, especially on the logicality. The peer assessment has indicated improvement in the undergraduate students' writing skills for scientific proposals.

Keywords

Logicality • Massive open online courses • Normativity • Objectivity • Peer assessment • Peer learning

D. S. Wahyuni (✉)
Informatics and Education Department, Graduate Program, Ganesha University of Education, Buleleng, Bali, Indonesia
e-mail: dsy.wahyuni@gmail.com

G. Ariadi
Management Department, Faculty of Economic and Business, Universitas Kristen Satya Wacana, Salatiga, Indonesia
e-mail: gede.ariadi@uksw.edu

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_12

1 Introduction

The MOOCs (massive open online course) phenomenon has received mixed responses from the global academic society. Many questions were raised about their methods of assessment. Designing and running a MOOC requires an integration of the assessments throughout the Design, Development, and Delivery (Rofe, 2011). More particularly, it requires varied MOOC routines and modules as well as clear and specific targets to be achieved. The goal is to ask the students to evaluate their peers' writing and work collaboratively. MOOCs are known for their potential to facilitate more comprehensive learning with pedagogic trialing and an official brand boost. In contrast, some critics, especially those familiar with online distance learning, regard MOOCs as publicity which has urged content providers to level up before adequately having pedagogical measures and structures in place, especially proper assessment methods. The latter part of the dispute argues for the distressingly low achievement degrees monitored through MOOCs (Halawa et al., 2014). Despite these critical voices, MOOC establishment keeps on increasing. At the same time, course schemes have spread further since the initial generation of MOOCs, specifically again concerning the development of assessment (Bayne & Ross, 2014).

The MOOC proposes an open course; however, more essentially, it recognizes the link between studying and assessment that enables a high degree of student-to-student communication and incorporates various peer learning chances. The content comprises a sequence of online assessments, defined by Salmon (2002) as e-activities. This enables a dynamic study in an online circumstance. E-activities consist of self-reflection and peer collaboration. These serve as evaluations for learning. However, the evaluation tool of MOOC only allows participants to check the number of correct or incorrect questions, which is how the scores are obtained. This reduces the process to a task that focuses solely on knowing the outcome of a particular teaching–learning process (Fernández et al., 2014). Therefore, they do not offer in-depth information about the learning and appear to be designed to exercise control over a crowded teaching–learning environment rather than to use evaluation as a process of reflection (Fernández et al., 2014). The debate in the literature has focused almost exclusively on evaluation concerning certification, forgetting other essential functions such as teaching information about participant expectations, the learning process, and improvement-oriented evaluation.

In line with the pedagogic emphasis, lecturers need to propose a marking guideline for each lesson, such as a numeric method that describes the calculations of the final scores. The main challenge we have found is the rigidity of the assessment method. The assessment, which is centered on multiple-choice and coursework, is inapplicable in peer learning which involves discussion. Implementing a way out in the theory of peer learning is labeled paragogic, which embraces a secondary position in a pedagogical context (Corneli & Danoff, 2011). Thus, there is pressure on the conventional expectancy of vertical conveyance of instructing from lecturer to learner and the parallel stream of peer learning that considers all contributors as learners (Lee, 2015). Bowles (2014) stated the contrary argument, contending that MOOCs have no impact on signifying the vast networked performance of collaborative learning. In the top-down model, MOOCs have degenerated into one-to-many interactions rather than discussions. Learners obtain knowledge via videos of lectures instead of allowing the chance to ask for the report containing learning modules. The multiple-choice test as a primary way of MOOC evaluation is an instance of this top-down method.

The essay is another approach that can be applied. The use of essay questions enables the students to have various answers, making it difficult for teachers to streamline them into one conclusion. Here, the difference between multiple choice and essay models lies in the firmness, straightforwardness, and clarity of students' answers. However, the essay model has the advantage of conveying the level of critical thinking and students' creativity in communicating their opinion.

Planning is essential to prepare students for peer assessment. The teacher's roles are as follows: to communicate the intent and purpose of peer assessment to all participants, to determine the assessment criteria, to train students in carrying out the assessment, and monitor the results of the assessment. Knowing the defiance of the scale marking, we decided to flip the assessment practice. Learners were instructed to have free-style discussions related to the assignment with their classmates before properly proposing their evaluation task. The implementation of the peer assessment procedure was as follows: student A would assess student B and student C; student B would assess student C and student D; student C would assess student D and student A, and student D would assess student B. This was done in such a way that the assessors remain unknown. Students were permitted to attempt as many periods as desired in the discussions while collecting enough responses from their classmates to develop the part they preferred to propose. Students' performance observation sheets were provided in the media forum of MOOCs. The forum provided an opportunity for the learners to learn how others extend their study and better involvement and participation in the learning. As a part of the peer assessment, the group proposed responses on different parts of their performance, such as cooperation and interaction competencies. Various outlines and rubrics to support the peer assessment were provided, which needed to be customized later. Expert appraisal of the result and the progression of the student's performance need to be conducted. To minimize the ineffectiveness of the peer assessment, the teacher needs to provide an introduction to ensure an understanding of the evaluation and the rubrics being used. Lack of understanding of the procedure of the peer assessment results in inaccuracies in the evaluation results.

Attention to involving learners in task exercises has mainly required writing to construct transforming knowledge in a virtual environment. Bazerman (2006) states that to recognize the proper techniques for scrutinizing the learners' transcripts, their composing capabilities and study progress need to be analyzed. MOOC should facilitate learners and teachers to address conflict resolution issues such as the assessment process and evaluation results. The student performance observation sheets (SPOP) are integrated into MOOC. Students evaluate their peers using SPOP, which can be seen by other groups. For example, students of group A will assess the work of students of group B by using SPOP as an assessment sheet. Students of Group B will receive the assessment results from group A, who then will make necessary adjustments based on the feedback. The content of the SPOP that is integrated into MOOC is an inter-group assessment based on 3 rubrics, namely objectivity, normativity, and logicality. It is necessary to improve the evaluation system tool based on automatism that is combined with peer-to-peer assessment, that also offers a space for evaluation exchanges between teachers and learners and between learners themselves (for example, making student work visible to other learners) (Braga Blanco et al., 2020). It is about embodying a basic pedagogical principle that is necessary to support the learning process.

These evaluation elements provide a scaffold to sustain the learner's interactional performance to accomplish the task. In the current study, essay assignment is regarded as a communal endeavor, with collaboration between scientific composing and other performances, such as peer assessment and discussion. This assignment was designed as a paragogy concept that integrated peer assessment and discussion. Paragogy is learning between peers starting from the process of identifying problems, the learning process, to the evaluation stage of learning (Corneli & Danoff, 2011). The supposition was created in which the incorporation of peer assessment and discussion would significantly influence learners' essay writing improvement. This research describes the improvement of students' abilities in evaluating the performance of peers and will answer the following questions: How did captivating in peer assessment and arguing writing experiences affect learners' performance in accomplishing the task in the background of a Project Management course?

2 Literature Review

2.1 Theory Peer of Learning

Peer learning is an instructional praxis where learners act together with other learners to accomplish didactic objectives (Corneli & Danoff, 2011). We apply the term paragogy to illustrate the vital education and praxis of peer learning (accurately, "para-" together with others, "-gogy" guiding, modifying the conventional theory of pedagogy toward the peer learning framework (Corneli, 2012). Paragogical values are: (a) grown-up learners are self-guided learning (b) which generate a variety of experiences in the learning background (c) which join learning backgrounds prepared to study (d) which are problem-focused on learning in studying. (e) which are greatly encouraged by inner aspects. Our central beliefs are understood as a criticism of andragogy, which is principally a concern of viewpoint. The andragogy (acquiring the comprehending of the mature person teaching approach), pedagogy (the art of teaching method), and contrasting heutagogy (attentions to self-directed studies) are based on the explanation. In contrast, paragogy meets in situations where students are dynamically involved in co-generating their studying atmospheres (Hase & Kenyon, 2001). The method within a constrained situation which encourages various kinds of communications and concurrently allow students to be self-directed is only consequential within an interactive environment. Therefore, students are expected to collaborate with other students in carrying out scientific writing in the context of peer learning.

Scientific writing is a mandatory learning activity that comes from deepening problems and students' critical studies of questions that arise from themselves to find answers and find solutions (Keys, 1999). Although students are able to write scientifically, they must also be able to provide strong data and evidence for accountability or scientific evidence from the descriptions made in scientific writing (Chen et al., 2013). This is the value contained when implementing a peer assessment. As a result, students have a level of accuracy and sharpness of thinking in analyzing. A previous study by Deng et al. (2019) stated that normativity, objectivity, and logicality are preserved as the necessities of a comprehensive scientific script. These three aspects must be implemented conventionally (integrated to meet the terms with standard language vocabularies and academic norms) and distributed by associates of the academic society (normativity). In this approach, authors and readers can generate the perception of specific outcomes and create agreement when writing and capturing the actual evidence and the factual development of scientific performance (objectivity). Like social activities and epistemological exercises, the authors of scientific writing can produce their arguments and describe their conclusions in more consistent and credible methods by concerning structured and rational argument (logicality). These three elements of the scientific script are crucial to improving the learners' ability to compose a script. To understand the application of these elements in a science script, learners are required to take part in accurate, detailed activities. In the present research, the incorporation of peer assessment toward scientific writing is offered as a strategy for improving learners' scientific writing practice.

2.2 Applying the Integrated Peer Assessment to MOOC

In the present study, the undergraduates were requested to write an essay for their MOOC course. Each student was given a username and password by the college so that they could access the MOOC and select the course. Writing assignments were given with the approach of encouraging students to become experts in providing assessments and feedback that could help other students to improve their writing. This assignment was called 'journal writing duty, whereas the matter of the writing in our case concerned investigating in empirical contexts'. This method can support learners with an enthusiastic contribution to learning (Bangert-Drowns et al., 2004).

To increase the quality of learners' scientific scripting, they must interpret some scientific articles earlier in their first writing endeavor in the usual scientific type. We assumed that by understanding good-writing scientific articles, learners could be offered an ancillary circumstance to collaborate with a prospective researcher and familiarize themselves with his(her) research. Through this 'single-way interaction', learners would be able to not only attain a more empathetic way of denoting the written expressions but also have a possibility to indirectly acquire the scientific experts related to the fundamental aspects and construction of a scientific manuscript, the guidelines and norms of the inscribed language, and techniques of simplifying ideas and expression.

Upon their achievement of each writing assignment, an opportunity was given for learners to commence peer assessments using the guidelines of normativity, objectivity, and logicality. Students had the opportunity to revise their writings after receiving feedback from their peers. Peer assessment is a representative social activity in knowledge. Students carried out social interactions through scientific arguments that are poured into SPOP. So, SPOP is a place to implement peer social interaction. It provided access for learners to comment on others' assignments and evaluate and confirm the suggested arguments. Furthermore, it is also a valuable method of aiding learners in enhancing their writing by admitting their peers' detailed recommendations (Lu & Law, 2012). The learners contributed to peer assessments by evaluating each other's written manuscripts and providing feedback on the strengths and weaknesses of each other's writing.

Additionally, the process of assessing the articles of other authors may provide learners with the learning opportunity and self-reflection on their writing (Luo et al., 2014). Before conducting a peer review of related student scientific writing, students must first understand the components or rubrics used as references when conducting assessments. They must use the components of the rubrics, such as logic, objectivity, and normativity, as a reference in conducting peer assessments. This was considered a guideline rather than a strict procedure for assessing the manuscript. This option was applied with the hope that such a methodology would allow flexible manuscript evaluations.

In this study, peer assessment was considered as a social movement that takes place among two learners or groups. It is intensive on interconnected opinions, responses, and criticisms in virtual circumstances. When learners were requested to commence a peer assessment, they might be concerned about faults in the sentences and words themselves, devoid of deliberating the progression of notion structure or discussing the conceptions inscribed for acquiring interpretation and recommending changes (Storch, 2005). As peer assessments are known to lack an element of encouraging applicable comments to authors, online discussions were appended to the peer assessment. These dialogues were aimed at enriching the learners' social relations. Learners acquire their manuscripts well while there are normal speaking and writing relations when all learners and the lecturers involved can argue the scientific manuscripts following each research and the proficiencies of taking on the peer assessments in the online environment (Chinn & Hilgers, 2000). For instance, shared work can support the learners to advance questions and contribute to their thinking about the scientific manuscript anytime and anywhere. Furthermore, this facilitates all of the associates to deliver beneficial recommendations to improve the writing by recommending diverse viewpoints. These recommendations will be practical if a process supports a learning atmosphere that encourages learners' thinking.

3 Method

The total number of participants was thirty which consisted of Indonesian undergraduate learners (nineteen-eight men and nineteen-two women). They were all learners at the Ganesha University of Education in Bali, Indonesia, and were majoring in the education of technology and vocation. In regards to their prospective occupations, some intended to achieve information technology (IT), some intended to join the IT industry, and others intended to work in high schools as IT teachers. Before the internship period, they were required to write a scientific proposal for their project. The sample of learners was all in the sixth semester. That they had accomplished the "Object Oriented Programming" course, "Web Programming" course, "Database Programming" course, and "Entrepreneurship" course. In order to further improve their skills in developing an application based on web or mobile, they took higher-level coursework termed "Project Management" course. All students selected for the research were registered in the course and keen to contribute to the study. From this group of learners, a sample was chosen by an online course automation system. After the aims of the present study were described to them, the nominated learners stated eagerness to contribute to this research.

The goals of the Project Management course are to develop undergraduate learners' competence in creating IT applications and simultaneously in formulating scientific writing for proposals related to their IT projects. All projects in this course were performed by ten groups (each group consisting of three members). This course is focused on preparing the proposal for IT projects. In order to prepare the manuscript of the proposal, the students need to possess the knowledge, capabilities, equipment, and techniques required for the project to fulfill the standard prerequisites before developing the IT application.

In this coursework, the lecturer requested groups to write proposals that generally meet the aims, backgrounds, literature, phenomena, implication, work-breakdown structure, and conclusion related to their IT projects. In doing such works, group members need to collaborate and discuss their proposal manuscripts. The lecturer marked learners' performance with regard to the substances of the final proposal, giving close notice to the qualitative and quantitative narratives related to their IT projects. This study was performed from the beginning of January 2021 to the end of June 2021. Before developing the IT application, each group was given a month to write one proposal manuscript to describe the project. They were required to explain all processes of the scientific performance and their expectation of results after finishing each project.

The group's task was to compile a digital-based business feasibility analysis proposal based on several studies or scientific analyses. In this context, "proposal" is defined as a planning document consisting of project initiation, project scope definition, and project planning of the proposed IT-based application development. For example, a group proposed the title "developing learning media based on discovery learning." The title submitted by students must convey the background of the application being offered and the features of the application. Then, the following month was the stage of peer assessment and improvement of peer assessment results. This exercise aimed to equip students with the skills to write scientifically. Feasibility analysis is a scientific paper because when students prepare a proposal, they must be equipped to analyze consumer needs and technological needs in developing an application submitted to the proposal.

The proposals that must be submitted were assignments from the Project Management course. The expected learning achievement in this project management course is that students are expected to make digital-based business designs and plans, which are shown in the form of a business feasibility analysis proposal. Students were asked to compile a proposal containing technology-based business analysis content under consumer needs analysis in this lecture. The proposal format was given before they started working. Some of the analytical studies contained in this proposal are business feasibility analysis, work breakdown structure analysis, business timeline to be carried out, and human resource management structure that will be required of the work.

After the inscribed manuscripts were submitted, the groups were designated into pairs to perform peer assessments for one month. Each group would conduct peer assessment on specific days, namely Tuesdays and Wednesdays. Each group would be given the task of reviewing the other groups anonymously. Every group obtained a scientific writing manuscript from their appointed peers (anonymous) and a guidelines rubric (Table 1) related to the assessment. At this stage, each group reviewed the clarity of the content analysis in the proposal (according to the provisions of the components that must exist in the feasibility analysis proposal). At this phase, the groups saw the brief explanations of the normativity, objectivity, and logicality of scientific writing and marked the peer's scientific manuscript in each dimension by grading 1-5.

During the group's review process, the lecturer served as a facilitator and guide in reviewing the three components, namely normativity, objectivity, and logicality. The roles of the lecturer were to convey the intent and purpose of peer assessment to all participants involved; to determine the assessment criteria that must be developed and delivered to students; to train students in conducting the assessment, and to monitor the results of the assessment. After peer assessment, every group received the evaluation outcomes from their colleague. During the peer-evaluation stage, the lecturer continuously assisted learners in handling the problems of perception of objectivity, logicality, and normativity. The results of peer reviews from other groups will be returned to that group which will then be corrected and re-evaluated by different groups. Every week during the evaluation process, one month before the re-evaluation, discussions are sustained by

Table 1 Peer assessment Rubric (Deng et al., 2019)

Dimension	Explanations	Apply '✓' to mark (high quality-low quality)				
		5	4	3	2	1
Normativity	The proposal must comprise all critical aspects of the paper, precisely the entire procedure of the project management course. The manuscript, words, sentences, paragraphs, graphs, tables, and literature are correctly composed following the scientific manuscript standard					
Objectivity	The substance of the manuscript must be in line with the existing established scientific ideas with no mistakes and expose the customer needs analysis, phenomena, product features, work breakdown structure (WBS), risk breakdown structure (RBS), timeline schedule, and conclusion properly					
Logicality	The manuscript should be prepared logically so one can recognize the correlations amid various segments. The logical argument is also correct, and for one to consider the conclusions confirmed by the writer					

the online platform. The online platform utilized in this research was the group discussion room, a software developed by the Ganesha University of Education. Every student in the group would mention the name of their group when communicating in the group discussion room to facilitate more flexibility in reading and giving feedback. The topics debated in the group discussion room principally converged on learners' understanding of scientific manuscripts and peer assessment. Their notions of enhancing scientific manuscripts and contemplations were an essential aspect of the discussion. In online discussions, they could share ideas to improve the quality of analysis of proposals and reflections.

These two examples demonstrated some of the potential approaches to the enhancement of learners' scientific manuscripts (Figs. 1 und 2). By performing peer assessment, learners critically reviewed others' writing content and recognized how others posed this content. By comparing their proposals with others, they could establish the benefits and drawbacks of each proposal. For example, in the project initiation section of the proposal, students will observe how other groups describe the proposed project to be made such as the background of the proposal, and application features. When exchanging their capabilities in online circumstances, they could paste other manuscript sentences or review them openly to improve their scientific proposal. In other situations, the researchers continually connected their discussion to deliver some drawbacks and displayed some clarifications when needed. At some point in the problem–solution sharing phase, the students and researchers were required to propose their views to attempt to get consent, step by step, by continually inquiring, conferring, and arguing.

It should be explained that during the discussion stage, all the learners would have entered to consolidate the discussion. The discussion among the learners was done to obtain the response of their problems in peer evaluation. Because of drawbacks and problems in completing the proposal manuscript, each group would receive feedback proportionately from the groupmates and lecturers. The reflection of the peer assessment is that some students were not experienced in carrying out peer assessment, and there were misconceptions about the rubric

Yoga (Group A): How would you describe the business constraints put forward in the proposal? Has the constraint been adjusted to the internal situation and condition? Another proposal explains that the constraints are described based on cost, time, and human resources analysis.

Agus (group B): Business constraints have been adjusted to the company's internal situation and conditions, as seen in the cost description in the budget plan section, time in the timeline schedule section, and the human resources involved in the project team's organizational structure. However, this proposal has not detailed the financing of each activity at each stage, so it has not shown a precise percentage of each stage of the business feasibility analysis proposal.

Maria (group C): Constraint has been adjusted to the situation and condition as well as the phenomenon. The explanation of the constraints is strengthened by providing SWOT analysis and a SWOT image that produces the constraints. We will provide improvements by adding cost, time, and human resources study to the SWOT analysis.

Fig. 1 Peer assessment data 1 (original scenario transcript classroom activity)

Ayu (group A): The business planning steps have not been clearly described; there is no time division and the team that will handle each description of the business plan; this can be seen from the work breakdown structure that has not comprehensively accommodated the timeline and the team in charge.

Doni (group B): The work breakdown structure has not been described systematically and adapted to the theory of WBS formulation.

Dona (group C): WBS has not been systematically structured and logically adapted to the time and costs proposed in the budget plan and project timeline. There are no details of the costs charged to each activity.

Researcher: You can describe WBS by going back to the following lecture. WBS must be described clearly and be adapted to the needs of application development. WBS is equipped with activities in each stage of application development. A detailed explanation regarding each activity is required to determine the number of costs incurred. Cost planning is crucial to know the stages that will use the most significant budget.

Fig. 2 Peer assessment data 2 (original scenario transcript classroom activity)

listed in the SPOP. The next thing to consider when applying peer assessment is (1) students' understanding related to the assessment rubric (2) honesty and objective factors when assessing peers. Peer assessment can only be successful if there is mutual trust between students and teachers or between students and students. To create this, teachers must monitor the implementation of peer assessment so that it is effective and the validity of the assessment can be accounted for.

3.1 Data Resources and Analysis

Analysing the data from a case study entailed four approaches that depended on theoretical propositions; performing your data from the ground up, extending a case narrative, and assessing reasonable rival clarifications (Yin, 2014).

The procedure of data gathering and analysis of this research was performed by considering these four aspects. Theoretical attention to the fit scientific proposal was utilized in assessing learners' scientific manuscript performances. The elements and performances were all selected directly by the data collection, and the three dimensions of objectivity, logicality, and normativity were applied to recognize the performances. After analyzing, all states attempted to be summarized to explain learners' diverse conditions of improvement. The reasonable opposite groups' justifications were also regarded.

This research data mainly came from students' scientific proposals, such as IT application development project designs, and was analyzed in certain methods to identify and determine things about the quality of students' scientific proposals by forming attention to their objectivity, logic, and normativity. For the written analysis (Fairclough, 2003), each proposal was assessed distinctly at the beginning. First, the components in each proposal were identified. After recapitulating the recognized components in all the proposals and discussing them with the lecturers, the phenomena, aims, problems, literature, method, budget, estimating results, conclusions, and references were considered the fundamental and essential components of script a scientific proposal of an IT project. Second, every proposal's deficient performances linked to normativity, objectivity, and logicality were identified. The result of identification in the normativity rubric was that learners did not pay attention to writing the background and insert references in the finish. Deficient in objectivity was that learners were not constantly appropriate in writing the analysis aspect. Conversely, in the logicality aspect, the students have shown an understanding of substantiation and applied it in a more adequate, realistic, and appropriate way. Thus, all the related deficient performances were distinguished in each proposal; they were reviewed to improve the quality of the proposal. While employing the reviews in each dimension, every group's proposal was reanalyzed across textual coding. In regards to normativity, the lack or inadequate performances in each component and the deficient performances linked to sentences and academic standards were identified. Within the viewpoint of normativity, these inadequacies were observed when the learners could not deliver rational explanations for the unsatisfactory evidence, and unpredicted phenomena also attempted to obtain by outlining very minimal arguments such as "improper system' and "coding debug" in the discussion aspect at the opening. However, such explanations could not satisfy the scientific community norms because they did not provide clear ideas of the reason for the unsatisfactory results and unexpected phenomena. Thus readers lose interest in reading the article. For each of the recognized segments, the groups' proposals were marked on each script component. For instance, there were missing components such as the background, aims, problems, phenomena, literature, methods, and conclusions for assessing normativity performance. As presented in the background example, the learners started to set into view some explanations by contemplating the displays of the problems and aims entirely. The students stated all problem information in a tabulated sheet. Furthermore, the codes of academic standards specified that there were a few learners who had problems in creating tables, such as ignoring the order number and the caption.

In order to fulfill the scientific category, the components were graded as well. The SPOP created comprised performance markers that were simple to take and remark by the peer-assessed. The SPOP was provided with a rubric with a scale from 1 to 5 (low quality-high quality). The outcomes of the assessment were transferred into percentages. In occurrences where the groups could not follow the norms of the scientific proposal, their appraised performances were appointed to each component of the writing proposal. When assessing objectivity, the explanation was not proper with the actual condition during the estimated results. Process methods, which did not suit the existing settled scientific notions, were identified. In this case, the researchers' evaluation of the writing involved evaluating the scientific ability of synthesis, embracing aspects such as the instrument used for the processing and consistency between the description of phenomena and actual condition (objectivity). The final evaluation of the groups' writing concerned logicality, indicating the complete logical coherence of the groups' writing. For this aspect, all the deficient performances, which did not follow the logic of writing and lacked evidence, were classified. The logical flow of the scientific proposal was evaluated using the orders of the writing components, and the validity of the groups' argumentation comprised key dimensions of evidence such as adequacy, applicability, reliability, the strength of the necessitates, and the clarification of the estimated outcomes.

4 Results

Peer assessment provides opportunities for students to share related scientific writing rules on proposals, think critically in assessing business feasibility seen in the phenomenon of current consumer needs, compare writing presentation techniques with other groups, conduct self-reflection by group on input and assessment from other groups, and make continuous improvements to the evaluation results. Students freely put forward arguments for the assessment during the discussion process and applied logic and judgment in assessing. Students were required to fill in the rubric and understand the components of the assessment first. Here lies the structural thinking of students' thinking in assessing. Students must first understand the components (normativity, logicality and objectivity), pay close attention to the proposals they value, and link them to the components of the supporting data in the proposal. Students would indirectly compare their proposals with other groups' proposals, and they know the weaknesses and strengths of the proposals. The advantages of other group proposals will be the opportunity for these students to make writing improvements to their own proposals. By using assessment rubrics (normativity, objectivity, and logicality), students would systematically use this rubric as a reference in preparing proposals that are, of course, adapted to the proposal format. For instance, learners would mention or refer to others' composing exactly to initiate the discussion; as an outcome, these linked sections were selected. The most constructive sections were those which displayed the impacts of peer assessment on learners' improvement in scientific writing; for illustration, where the subject of the discussion linked to those parts of the proposal in question requiring improvement or adjustment. The limitations of the peer assessment process in scientific writing are: first, students had no experience in conducting peer assessments; second, there were differences in the interpretation of criteria and indicators from the SPOP rubric with those delivered by lecturers; third, the objectivity factor in the assessment where students are more likely to give a neutral assessment score even though the answer is irrelevant.

The results showed that peer assessment facilitated students' continuous improvement as a form of positive change. Students were expected to focus on the content presented and adapt to the format and reference rubric in the proposal. The proof was needed for the consequences they described in the proposal (as shown in Fig. 1). In the excerpt below, arguments are given in the "constraint project planning" section. A project proposed in the proposal must have a constraint review and be accompanied by supporting evidence. According to the theory of constraint, constraints are translated into three elements: cost, time, and human resources (Blackstone et al., 2009). Each group must assess the constraints of its project. This constraint can be illustrated in a SWOT analysis. The feedback of peer evaluation was delivered to assist learners in accomplishing the suggestion by sharing and evaluating proposals from different groups. While this suggestion can be considered self-reflection, it was affected by interpersonal actions. Conversely, the discussion part pointed to supporting the learners in accomplishing their joint reflection via problem-solving in a social circumstance.

The snippet above illustrated an assessment related to "objectivity." In writing a business feasibility proposal, an objective analysis must be carried out. Students objectively assessed constraints in conducting business feasibility analysis studies. Sometimes in proposing a business, students were always optimistic and did not pay attention to the objective; in this case, they faced the constraint. For example, in the excerpt above is, Yoga from group A commented related to the constraints. Yoga has also made comparisons with the proposals of other groups. Yoga gave advice, and if you describe constraints, you should focus on cost, time, and human resources. Constraints are adjusted to the analysis of existing conditions and situations.

Likewise, Agus in group B gave a reaffirmation related to constraints. The constraints were clearly described in the work breakdown structure (WBS) table according to group B. If it is associated with an assessment rubric, then this is a focus on "normativity." Details of the constraints must be accommodated and clearly described in the WBS table. Agus explained this constraint in a sentence, but the quantity is determined in the WBS table. The presentation of the constraints has not been normatively provided and has systematically reduced the amount of each constraint quantitatively. This was followed with a snippet showing the assessment on the normativity rubric but in another format.

The excerpt above (Fig. 2) emphasizes the details of the WBS. The presentation of the WBS is the result of integration between analysis of phenomena, product features, stages of work, details of activities carried out in each stage, costs and time, and human resources of each activity, accumulated costs, time, and people in each stage, accumulated costs, time and total people involved. The human resources involved must match the expertise in each activity. The burden of financing expertise or salary is also adjusted to the length of time they work and their workload. Constraint analysis is a holistic integration that must be in the proposal. Integration analysis is called normativity and logicality under the assessment rubric.

Furthermore, Ayu (group A) argued that the proposed WBS has not comprehensively integrated the constraints. The argument was also confirmed by Doni's statement from group B, who stated the same thing, that the WBS had not been presented in a systematic and structured manner. Dona understood the arguments expressed by Ayu and Doni from group C. Dona agreed that the WBS they made in the proposal did not comprehensively include constraints with supporting data for quantitative analysis. In this case, the researcher's role is to guide and facilitate by providing clarification, reinforcement, and explanations related to the systematic presentation of the WBS. Researchers referred to a literature review that must be re-examined by group C to improve and perfect the WBS.

Finally, peer assessment provides honest feedback and includes more flexible discussions. Each group was able to assess the proposals and make comparisons with proposals in other groups. The arguments given are also accompanied by evidence. Weaknesses and strengths would be the material for reflection in each group. The opportunity to improve gradually was given to each student. This will produce a business feasibility analysis proposal that fits the existing praxis in the field. The point of view of assessment that still refers to the rubric was also the main point in peer assessment. Each group had a different point of view regarding the meaning of the rubric that was adapted to the contextualization of the proposal.

5 Conclusions

In the current study, peer assessment assisted learners in accomplishing the reflection by exchanging and comparing manuscript proposals by various groups. Even though this reflection might be considered self-reflection, it was triggered by relational conduct. Conversely, the discussion phase was intended to aid the learners in finishing their joint reflection through problem resolution from a social perspective. Those two accomplishments have brought positive contributions to the learners' scientific manuscripts.

The reflection, which was enabled by the peer assessment, was continually spontaneous. It identified the learners who were able to comprehend the problems in their own scientific writing and improved by exchanging and comparing with others. Conversely, the reflection which occurred during the discussions was not natural. As exhibited in the results section, some of the discussions were stimulated by the researchers. Occasionally, the researchers wanted to mention the main questions or even uploaded and summarized resolutions to specific problems in some situations. The results showed positive changes linked to the peer assessment, especially concerning normality, objectivity, and logicality. This was indicated by the group of students who were able to demonstrate the essential components sufficiently and were more empirical and scientific in comparing their manuscripts with others. The aspect linked to logicality focuses on showing the evidence where students recognized how to be adequate, realistic, and appropriate in displaying their data before developing the projects.

In conclusion, the task of peer assessment has indicated the improvement of Indonesian undergraduate students' skills in writing scientific proposals, mainly linked to project management courses. This study has offered a new understanding of the strategies for assessing the manuscript. The lecturer can formulate rubric assessments for scientific manuscripts innovatively to provide frequent social occasions to interact, contend, or argue about their improvement, rather than putting out the procedures and aptitudes of a scientific manuscript.

References

Bayne, S., & Ross, J. (2014). *The pedagogy of the Massive Open Online Course (MOOC): The UK view*. The Higher Education Academy. https://www.heacademy.ac.uk/resources/detail/elt/the_pedagogy_of_the_MOOC_UK_view.

Bangert-Drowns, R., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29–58.

Bazerman C. (2006). Analyzing the multidimensionality of texts in education. In J. L. Green, G. Camilli, P. B. Elmore, A. Skukauskaiti, & E. Grace (Eds.), *Handbook of complementary methods in education research (PAGE(S))*. American Educational Research Association.

Blackstone, J. H., Jr., Cox, J. F., III., & Schleier, J. G., Jr. (2009). A tutorial on project management from a theory of constraints perspective. *International Journal of Production Research, 47*(24), 7029–7046.

Braga Blanco, G. M., Belver Domínguez, J. L., & Fano Méndez, S. (2020). Evaluation is not a second phase in the pedagogical design of a sMOOC. A case study = La evaluación no es un segundo paso en el diseño pedagógico de un sMOOC. Un estudio de caso. Profesorado.

Bowles, K. (2014). Walking and learning. Wordpress. https://musicfordeckchairs.wordpress.com/2014/03/05/walking-and-learning/.

Chen, Y.-C., Hand, B., & McDowell, L. (2013). The effects of writing-to-learn activities on elementary students' conceptual understanding: Learning about force and motion through writing to older peers. *Science & Education, 97*, 745–771.

Chinn, P., & Hilgers, T. (2000). From corrector to collaborator: The range of instructor roles in writing-based natural and applied science classes. *Journal of Research in Science Teaching, 37*(3), 3–25.

Corneli, J. (2012). Paragogical praxis. *E-Learning and Digital Media, 9*, 267–272.

Corneli, J., & Danoff, C. (2011). Paragogy. In *Proceedings of the 6th Open Knowledge Conference*. Website name. http://ceur-ws.org/Vol-739/paper_5.pdf.

Deng, Y., Kelly, G. J., & Xiao, L. (2019). The development of Chinese undergraduate students' competence of scientific writing in the context of an advanced organic chemistry experiment course. *Chemistry Education Research and Practice, 20*(1), 270–287.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Psychology Press.

Fernández, E., Rodríguez, C., & Fueyo, A. (2014). Concepciones Pedagógicas y Comunicativas de los MOOCs: Estado del Arte y Prospectiva en el Marco de un Proyecto Europeo. III Workshop internacional sobre Creación de MOOC con anotaciones multimedia, Málaga, 5-7 de marzo.

Halawa, S., Greene, D., & Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. *eLearning Papers, 37,* 1–10. http://www.openeducationeuropa.eu/en/article/Dropout-Prediction-in-MOOCs-using-Learner-Activity-features.

Hase, S., & Kenyon, C. (2001). Moving from andragogy to heutagogy: Implications for VET. In *Proceedings of Research to Reality: Putting VET Research to Work: Australian Vocational Education and Training Research Association (AVETRA)*. AVETRA.

Keys, C. W. (1999). Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science. *Science Education, 83*, 115–130.

Lee, Y. (2015). The fragile beauty of peer-to-peer activism: The public campaign for the rights of media consumers in South Korea. *New Media & Society*. Advance online publication. https://doi.org/10.1177/1461444815582019.

Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science, 40*(2), 257–275.

Luo, H., Robinson, A. C., & Park, J.-Y. (2014). Peer grading in a MOOC: Reliability, validity, and perceived effects. *Online Learning Journal, 18,* 1–14. http://olj.onlinelearningconsortium.org/index.php/olj/article/view/429.

Salmon, G. (2002). *E-Tivities: The key to active online learning* (2nd ed.). Routledge.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing, 14*(3), 153–173.

Yin, R. K. (2014). *Case study: Research design and methods* (5th ed.). Sage.

Dessy Seri Wahyuni a lecturer and researcher of Informatics and Education Department, Graduate Program, Ganesha University of Education. She has experience as research leader of—Vocational Teacher Competencies in Vocational High School. Her teaching and research focus is related with the curriculum development, strategic collaboration of vocational education. She is actively publishing some articles. One of the articles that being published is The Effect of Agility and Product Innovation Efficacy in Indonesia. She is actively participating in certain professional societies, such as ADGVI and APTIKOM.

Gede Ariadi is a lecturer at Management Department, Faculty of Economic and Business, Universitas Kristen Satya Wacana and also consultant on several firms that move in education, operation management and vocational technology. His recent research interests include vocational industry educational management and learning technology. He is actively participating in certain professional societies, such as ASPADIN (Association of drinking-bottle industry Indonesia) and P4I (Association of Lecturers and Practitioners for International Trade).

Formative Assessment of E-Service Learning Using Learning Diaries and Group Reflections

Katharina Resch

Abstract

Service learning is a methodology that is now applied more often in higher education teaching across disciplines. It connects theory and practice by allowing students to develop or participate in a service in the community while at the same time reflecting on their learnings in class. Due to the complexity of service learning as a didactic approach, assessment practices are also multifaceted and have to take into account the perspectives of teachers, students and community partners. Assessment practices seem to be even more complex in digital spaces. This article reports on a study of assessment practices and related student learning in an e-service learning course using weekly learning diaries, group discussions and online presentations at the largest university in Austria in 2020-2021. The findings show the usefulness of structured learning diaries for formative assessment. In contrast, needs-based online group discussions do not exhibit the same impact. The study emphasizes the valuable learning opportunities in e-service learning for students but also their critical opinions on assessment practices.

Keywords

Assessment • e-service learning • Higher education didactics • Learning diaries

K. Resch (✉)
Centre for Teacher Education, University of Vienna, Vienna, Austria
e-mail: Katharina.resch@univie.ac.at

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_13

1 Introduction

Lecturers' aspirations to integrate practice-oriented methodologies into their teaching repertoire have increased in the higher education sector in recent years. Nonetheless, traditional forms of teaching, which are based on teacher input, principles of frontal teaching and shaping students' behaviour towards a pre-determined learning objective, are still commonly practiced (Roessingh & Chambers, 2011). More learner-centred forms of teaching require active student learning and have not yet become standard practice across disciplines, although they have gained more prominence (Gelmon, 2017). The Prague Communiqué (2001, p. 2) calls for students to be "competent, active and constructive partners" in higher education. Accordingly, their role has undergone a considerable shift—from merely 'customers' of education (Naidoo & Jamieson, 2005) to active participants in the learning process.

Service learning is a methodology that has recently been applied in higher education teaching across disciplines which adheres to the principle of learner orientation. It connects theory and practice by allowing students to develop or participate in a service in the community in order to gain an enhanced sense of civic engagement and to reflect in parallel on the experience in class (Bringle et al., 2006; Felten & Clayton, 2011; Furco, 2009). Service learning places its focus on an organized service activity that students provide (*service*) and on parallel reflection in class (*learning*), thus combining academic and civic learning. This, of course, requires the active participation of the student in both the service and the learning process. We assume that students learn differently inside and outside the classroom and thus grow personally in a different way in a practice setting than in a purely academic environment. When engaging in service learning, students learn to apply knowledge to real-life problems, which reinforces citizenship and civic engagement.

Assessing service learning is a complex task (Gelmon, 2017) as lecturers have to assess not only the classroom learning but also the service element, which takes place in the community. How academic and civic learning take place is the centrepiece of assessment in such courses, and finding appropriate assessment methods is therefore an essential component of effective teaching. This is also especially challenging in an online environment.

The recent COVID-19 pandemic forced the higher education sector into a period of remote learning in 2020, which remained the dominant mode of teaching well into 2021, thereby establishing a new reality in which online teaching and learning are the norm rather than a temporary phenomenon. To adhere to the required closures, different forms of online teaching, learning and assessment were mobilized. This was also the case for applied coursework, such as service learning courses, which likewise had to be (partly) transferred into online spaces (e-*service learning*; Waldner et al., 2012).

This article reports on a study of assessment practices and related student learning in an e-service learning course using weekly learning diaries and group discussions at the largest university in Austria in 2020-2021.

2 Theoretical Approach

2.1 Specificities of e-Service Learning

Traditionally, service learning activities take place without an online component (Waldner et al., 2012) – both the learning in class and the service are conducted onsite in the community, e.g. in an NGO, a school, or a business. During the COVID-19 pandemic, universities closed temporarily, causing many service learning courses to have to adapt spatially: classroom learning was replaced by hybrid or online learning, and onsite services in the community were replaced by hybrid or online services. We differentiate between five types of e-service learning courses, as shown in Table 1.

As already mentioned, traditional service learning is performed onsite in the classroom and the community. In the Type I. e-service learning hybrid, the instruction component between the lecturer and the students takes place online, and the service part onsite, e.g. in a school. The Type II. e-service learning hybrid is the opposite of Type I., with the instruction part taking place onsite at the university, and the service part completed online. The blended e-service learning hybrid is designed as a blended-learning environment both in the classroom and the community, and, last but not least, full e-Service learning takes place entirely online.

E-service learning courses of the Type I., Type II., blended or full e-service formats may have several advantages when it comes to reaching the learning

Table 1 Types of service learning (adapted from Waldner et al., 2012, p. 134)

	Onsite service	**Online service**
Onsite instruction	Traditional service learning	E-service learning hybrid type II
	Blended E-service learning hybrid type	
Online instruction	E-service learning hybrid type I	Full E-service learning

objectives. Firstly, if the service component is provided online, e-service learning can be carried out not only with partners in the university's local community but also with geographically dispersed organisations in rural areas or even global community partners without the need to travel (Waldner et al., 2012). Secondly, by integrating online components into the instructional part of a service learning course, geographically dispersed students, or students who are unable to attend lectures onsite on campus, can participate. This contributes to participatory and inclusive teaching in higher education, as it allows non-traditional students or students with disabilities to also access service learning (Waldner et al., 2012). Furthermore, the blended and full e-service learning formats also allow distance learning universities to implement the service learning approach in their curricula. Thirdly, e-service learning activities with internationally based partners enable students to pursue global citizenship and transcultural skills (Harris, 2017; Garcia-Gutierrez et al., 2017).

However, e-service learning can also present challenges for lecturers in terms of availability of technology, communication with students and community partners, course design and assessment of learning outcomes. In any event, technology should never take precedence over didactics. Due to the COVID-19 pandemic, lecturers were forced into a period of remote learning, which meant that course designs had to be adapted and re-adapted regularly as the situation progressed. This may have been viewed as burdensome for lecturers, since a service learning course usually requires more preparation by lecturers such as an established partnership with a community partner or active forms of feedback during the service phase (Resch & Dima, 2021).

Since aspirations to implement service learning have remained stable even during the pandemic and the resulting move to online learning, this raises the question of how to assess the effects of this teaching approach on students (Slepcevic-Zach & Gerholz, 2015). Attempts have already been made to assess service learning. Some studies show that students in service learning courses have a higher perception of their self-efficacy than those on traditional courses (Yorio & Ye, 2012). Others indicate that students develop civic responsibility and empathy with social issues (Govekar & Rishi, 2007). Some researchers are, however, less optimistic as real evidence on the impact of service learning is largely missing (Slepcevic-Zach & Fernandez, 2021).

2.2 Assessing Service Learning Courses

"Assessment serves a useful purpose as a mechanism to tell the story of what one has learned from one's work – articulating that learning for oneself as well as for others" (Gelmon, 2017, p. 33). The assessment of service learning courses to date represents a scarcely elaborated topic in research as it poses fundamental problems to higher education didactics and applied research. The complexity of service learning needs to be measured, namely the uptake of a service in a community – which may not be comparable to another service – and the students' learning processes in terms of academic and civic learning both in the community as well as in the classroom. Every service learning course is a unique combination of teacher, students, community partner, service performed and subject matter. Service learning is a complex didactic approach and, as such, requires approaches to assessment, evaluation and reporting that are suitable for capturing its complexity (Fromm, 2019; Gelmon, 2017).

The assessment process for a service learning course has to be designed in a multidimensional format, focus on feedback, be rooted in the respective context and be both informal, formal and continuous (Farber, 2011). Assessment of service learning should not take place only when the course is completed (*summative*) but also during the learning process in the community (*formative*).

Learning outcomes are statements about what is to be achieved and assessed at the end of a course and usually describe what is expected from a student in terms of understanding, knowledge and skills (Harden, 2002). In service learning courses, academic and civic learning produce outcomes related to cognition (e.g. content knowledge), abilities and skills (e.g. transferable skills, practical skills; Davies, 2002). It entails a collaborative initiative between students and the community that involves explicit learning objectives, a response to genuine community needs and systematic reflection on the part of the students (Scott & Graham, 2015). It comprises action and reflection similar to action research (Chambers & Lavery, 2017), whereby students learn through both the action/service and the reflection on it.

The role of the lecturer is to provide input and expertise regarding the service learning topics as well as guidance for students when they report difficulties. A reliable partner in the school community is also relevant for student learning. This ensures that students receive regular feedback from either the lecturer and/or the main contact person in the community. Reflection on service learning as an essential part of the assessment process encourages students to integrate experience, observation and knowledge as well as to examine theory in practice (Berman,

Table 2 Forms of reflection (adapted from Jakoby 2015)

Written reflection	Interactive reflection	Media-based reflection	Oral reflection
• Learning diaries • Service plans • Surveys	• Text-based discussions • Focus group discussions	• Drawings and maps • Videos • Art	• Presentations

2006; Jacoby, 2015). Hereby, different types of reflection activities can be useful (see Table 2), such as written, interactive, media-based and oral reflection.

The most valuable contribution that higher education makes to civic engagement is to equip graduates with the knowledge, skills and values that empower communities and society as a whole (Chalkley, 2006). A clear assessment plan can contribute to an understanding of civic engagement in service learning (Gelmon, 2017). According to Eyler (2002), there are multiple assessment formats which can be purposefully integrated and used in such a plan (before, during and after rendering a service in the community) (see Table 3).

Teaching service learning courses is sometimes a question of access to relevant pedagogical resources and tools. Oftentimes, lecturers only work on a subcontract basis, have a part-time contract or have to deal with crowded classrooms and poor university infrastructure (Resch & Dima, 2021). They must be willing to actively involve students in their coursework (Furco, 2003) and collaborate with community partners. In addition, identifying a relevant problem that is both socially and academically significant for students and community partners is time consuming (Reinders, 2016). Some faculty remain open-minded about these new endeavours, while others are sceptical, "demanding strong evidence for the value of this work" (Gelmon, 2017, p. 32). As indicated in the tables above, lecturers need to think of creative ways to assess learning outcomes when they design service learning courses. However, assessing service learning in its complexity could ultimately produce evidence on how service learning works.

2.3 The e-Service Learning Course

Service learning in the concrete case of teacher education students can mean rendering a direct or indirect service in schools for a primary (e.g. pupils) or secondary (e.g. school administration, school community, parents; Kaye, 2010) target group. While service learning has been frequently applied in teacher education worldwide (Hildenbrand & Schultz, 2015; Karayan & Gathercoal, 2005), it is

Table 3 Multiple forms of assessing service learning (adapted from Eyler, 2002)

	Before the service	During the service	After the service
Reflecting alone	• Writing a letter to myself • Setting personal goals	• Learning diaries	• Writing texts after the service (alone) • Producing videos or portfolios
Reflecting together at university	• Exploring expectations in the group • Querying experts in service learning	• Discussing hot topics and problems in class	• Group presentations in class
Reflecting together in the community	• Setting up a cooperation agreement • Conducting a needs analysis	• Lessons learned—reflecting in the community	• Group presentations in the community

still a rather new approach in higher education in Austria (Resch & Schrittesser, 2021).

The service learning course described in this paper was embedded in the 'University Cooperation Schools' Programme' (Resch & Schrittesser, 2021; Resch et al., 2022) at the University of Vienna, Austria. In this programme, schools regularly cooperate with the university for various purposes. The course was worth two credits, was embedded in the Bachelor's curriculum for teacher education, and involved two urban schools partnering with the largest teacher training institution in Austria. The author of this paper – as a lecturer in teacher education…– taught the e-service learning course in the 2020-2021 academic year together with a tutor. A total of 13 students enrolled in the course (see Table 4.), and four service learning projects were implemented. These projects were initiated and articulated by the needs of the associated school teachers, with whom the lecturer and tutor had established partnerships prior to the start of the course.

3 Methodology

This paper documents and critically discusses the assessment formats of the described e-service learning course – in particular the assessment data associated with the course collected for the 13 students who rendered digital service learning projects in schools. The different forms of assessment occurred using (1) weekly learning diaries, (2) focus group discussions and (3) online presentations, thus applying a mixture of oral and written assessment, individual and group assessment, and encouraging self-expression (Prinsloo et al., 2011). The assessments had to be arranged and adapted to fit the online learning spaces.

The main aim of the service learning assessment was to document and track student learning (rather than community (school) learning). The assessment was conducted formatively over the course of 8-10 weeks in the 14-week semester. Service learning was accompanied by the teacher and the tutor, who were both responsible for tracking progress through formative assessment. The assessment plan in the course contained three elements:

Learning diaries: The lecturer posed 16 questions (2 per week) over an 8-week period, which students were asked to answer individually on an online learning platform. The questions were posted every Tuesday in order to ensure continuous answering and reflection. The responses provided by a student were not visible to other students, only to the tutor and the lecturer. All questions were open-ended to allow students to write as much as they wanted and express themselves freely, thus granting learner autonomy (Han, 2011; Prinsloo et al., 2011).

Table 4 Overview of service learning projects

Title of the project	My fairy tale playground	Constructing simple robots	Working creatively with MS Teams	The art of motivating oneself
School community	Primary school, 12th district	Primary school, 12th district	Lower secondary school, 22nd district	Higher secondary school, 22nd district
E-service learning type	Blended E-service learning hybrid Onsite instruction in classroom Onsite and online service	E-service learning hybrid type II Onsite instruction in classroom Online service	E-service learning hybrid type II Onsite instruction in classroom Online service	E-service learning hybrid type II Onsite instruction in classroom Online service
Service objective	Promoting children's skills in analogue and digital cutting of photos of a fairy tale playground	Promoting children's skills in working with simple technologies, in particular constructing a fruit robot	Promoting children's skills in actively using a digital tool, its functions and interface for learning in class	Promoting teenagers' skills in motivating themselves to participate in digital learning during home schooling
Forms of assessment	Learning diaries Group reflections Online presentation	Learning diaries Group reflections Online presentation	Learning diaries Group reflections Online presentation	Learning diaries Group reflections Online presentation
Student characteristics *$N = 13$*	Teacher education students with specialisation in art $N = 3$ Female and male	Teacher education students $N = 3$ All female	Teacher education students $N = 4$ Female and male	Teacher education students $N = 3$ Female and male

Students received a brief explanation of the activity before the start of rendering the service in schools, describing the purpose of self-reflection in service learning and the possibility of answering openly (using text or creative methods like drawing or mind mapping). Examples of the questions asked include “What was your first impression of your onsite experience? Please describe.” (Question 3) or “Do you feel you are having the opportunity to make a real contribution? Why/Why not? Which benefits do you see in doing community schoolwork?” (Question 11).

Focus group discussions: Four focus groups with all 13 students were held online via Zoom by the course tutor and lecturer in the middle of the semester. The discussions lasted for 33 min on average, producing 133 min of material in total. Examples for the rather open and conversational questions are: How would you describe the progress you have recently made? What do you think constitutes a high-quality partnership in service learning in the associated school and your small group? All focus groups were audio-recorded on an external device and fully transcribed.

Online presentations: All four project groups held online presentations of their work in front of the course lecturer and tutor as well as representatives of the associated school.

While the learning diaries element in the assessment plan aimed at documenting individual student learning on civic issues, the focus groups and online presentations sought to identify factors of relevance to others conducting similar or parallel work. All data was initially audio-collected and transcribed in German, with key quotes subsequently translated into English by the author of this paper.

4 Main Findings

4.1 Weekly Learning Diaries

Phase I. Getting acquainted with service learning

The 13 students reflected on the methodology of service learning as such in the first few weeks of the course. They wondered what to expect and what they would experience.

> “Service learning sounds good: partly research, partly practice, partly civic engagement.” (CI, female, 25.10.2020)

They also expected to learn how service learning differs from other methodologies they were familiar with. The feelings mentioned by students at the beginning of the experience can be divided into positive (e.g. curiosity, optimism, enthusiasm, excitement, joy and motivation) and negative (e.g. anxiety, nervousness, feeling of being overwhelmed, fear and insecurity) sentiments.

> "I like surprises and that's why I am curious about the reality of the experience." (DH, female, 25.10.2020)

Some of the uncertainties expressed by students at this early stage were associated with a lack of experience in project development, concerns about the amount of time required to implement the service in the school, technical solutions and whether the implementation would match the planning. They also mentioned specific expectations related to social interactions in the schools as well as academic content.

> "What I am interested in in this course is working with children with special educational needs. There are four SEN children in the class (…), and three of them will participate in the service. I have never worked with children with disabilities before and don't really know what to expect" (DI, female, 25.10.2020).

> "I expect direct contact to pupils working in a service in a school, practical input and input about project management" (HD, male, 2.11.2020).

Phase II. Community partner phase

Students were very eager to get in contact with their respective schools to start the actual planning of the service activity, clarify open questions and find out how to interact with pupils in the online or hybrid environment. They reflected on this in different ways, some having had a positive experience with the school, and others having had to cancel appointments or experiencing technical disruptions.

> "My first impression after the Zoom meeting with the teacher (…) was very encouraging and positive. We were able to ask our questions and received friendly and helpful answers. I really liked the teacher's open-mindedness and the fact that she presented the pupils in the best light. She pointed out the children's advantages and talents and not their deficits. (…) I felt very comfortable. In addition, she radiates a pleasant calmness and confidence that makes me go to the meetings in a positive mood. She also supports us where she can, e.g. with group assignments, etc., but still gives us a free hand." (GC, female, 30.10.2020)

> "For technical reasons, we were not able possible to meet online. My fellow students, the teacher and I were online, but the Microsoft Teams link the teacher provided did not work. I recommended meeting on another digital platform (…), but the teacher did not consent. He wanted to try again the next day on Teams. The situation irritated me." (GME, female, 1.11.2020)

These examples show how the online environment shaped these first interactions with the schools. Most students first established contact with the headteachers, who they described as open-minded and helpful, then discussed the next steps for reaching out to pupils in line with their concepts in the subsequent planning phase.

Phase III. Planning phase

Students reflected on their ambivalent experiences in the planning phase, which lasted for around three to four weeks depending on the school. They quickly learned that their planning would likely need to be adapted when confronted with real-life needs and organisational structures. This meant being flexible in adapting the service as a whole or in parts, organizing elements in a different chronological order or changing the methodology.

> “We had to adapt our whole concept and look for new literature because the main activities in the school were completely different than we expected.” (GC, female, 9.11.2020)

> “We will have to shorten our concept, and some planned activities will have to be dropped. I learned that in this project we need to be very flexible and always have a Plan B, or even Plan C and D.” (GQ, female, 5.11.2020)

Students used the learning diaries to reflect on the process of planning the service, dealing with challenges of managing time and resources and coordinating the schedules of those involved.

> “We had planned two two-hour lessons but this week we found out that due to the class schedule we would only have one double lesson. So we will have to re-schedule the contents of our workshop." (PE, female, 10.11.2020)

> “During the week I was a bit frustrated because the project didn’t go as it should have. My colleagues and I had an online appointment with the class. In my opinion, everything went as expected during the lesson. Nevertheless, we agreed with the pupils to make an appointment outside the timetable. They had two days to make an appointment with me. I gave them various options and a deadline by which to decide. Unfortunately, that didn’t happen. One of the groups responded about 10 minutes before one of the possible dates, and the others did not respond at all. The situation frustrated me because I could not keep to our plan. It was a kind of conflict for me, but more with myself than with the pupils.” (GME, female, 26.11.2020)

Phase III. Implementing service learning

All four service activity projects were implemented in schools within the planned period, three as hybrid Type II activities, and one as a blended learning service. Students reported on their activities in detail and described their approach as taking one step after another, sticking to the plan, while at the same time remaining flexible.

> "The pupils have already started to draw [their favourite playgrounds], which we will take up in the next lesson. I have also already found an app which they can use to digitally cut their drawings. For this purpose, I researched various image editing apps and tried them out." (DH, male, 30.11.2020)

Technical infrastructure was a topic of reflection, both with regard to the students themselves and to the pupils.

> "All the children had their cameras and microphones on and paper and pens in front of them from the outset. During the video call (…), they were all active and answered questions willingly." (DH, female, 4.1.2021)

Some students reflected on the pace of the activities, which was quicker than in their previous experiences of internships, but also more structured and organised. Another topic of reflection was the small group experience at university: working together, balancing skills, coordinating schedules and creativity in small groups. Most students were glad that the small groups had worked well and reported positive experiences, e.g. being able to rely on others or learning from others in the group.

> "This week we started video production in class. I was positively surprised by my team members because of the creative and high-quality videos they produced." (HM, male, 3.12.2020)

> "Our online drawing session fascinated me, in particular the outputs. The teacher was so well organized. All pupils were present, well-informed and ready to work. The PowerPoint presentation reached its main aim. The children were very interested and motivated, and some ideas were already transferred into concrete images. It was so much fun to provide guidance for the children and watch them at work." (DM, female, 29.12.2020).

Students also reflected on bridging the theory-practice gap in teacher education. They realized the complexity of developing and implementing a plan in a real-life setting.

Phase IV. Reflecting on the overall experience of service learning

Students reported that they were more critical towards society after their service learning experience and that they acquired skills during the process which they previously did not have. They also grew academically due to the openness of the learning component.

> "The additional practical experience was a huge advantage for my academic growth." (HD, male, 16.1.2021)

> "I learned that theory and practice can be very different. Since I worked as a teacher for several years, I was able to transfer my experience to lesson planning and lesson design. It is important to pay attention to communication and be open to changes during the project, otherwise difficulties may arise." (GME, female, 16.12.2020)

The learning diaries also contained elements of self-reflection on their own identity as engaged citizens and their future role as engaged professional teachers. First, their self-evaluation changed – they started to think about themselves as more social or civic people than before. Second, they reported wanting to engage in civic matters more after the course had ended or wanting to integrate this kind of civic engagement in their professional role as a teacher in the future.

> "This [experience] changed me insofar as I personally want to be much more critical of and at the same time more engaged in society." (QG, female, 22.11.2020)

> "Through service learning I learned new ways to actively engage children and teenagers in social, ecological, political and cultural issues. I think service learning can give young people valuable experiences like self-confidence, self-efficacy, critical reflection, sense of responsibility, participation, cooperation and solidarity. Such experiences are elements of basic democratic education. My own civic engagement in this school led me to rethink my position as a future teacher and how little we do to actively engage students in civic matters. This just came to my mind now." (DH, female, 29.12.2020)

Overall, the experiences with service learning in schools were seen as positive and rewarding experiences. However, the students also criticized the level of recognition received for the service learning activities (only two credits) and noted that they would have liked to have had the "gold experience" (traditional face-to-face service learning). Ultimately, they viewed e-service learning as an alternative, but not the standard.

> "It is frustrating to want to do something but be so restricted by the pandemic." (GC, female, 26.11.2020)

The learning diaries were a good and suitable tool for assessing progress in the project groups – despite the fact that the students had to reflect individually on their weekly experience in them. Responses varied in length from a few sentences to several paragraphs per question. The learning diaries were more detailed and longer at the beginning of the service learning course, also describing expectations and feelings, and became shorter during the semester, when students were preoccupied with rendering the service. The learning diaries were honest, also reporting difficulties or not being able to answer questions.

> "I didn't learn anything from service learning this week." (MMT, female, 7.12.2020)
>
> "I can't answer question 14." (GC, female, 15.12.2020)

In assessment terms, these responses can be taken into account if they are written down and explained. Difficulties reported in the learning diaries were taken up as potential topics in the online focus group discussions.

4.2 Online Focus Group Discussions

It was important for the students to understand the community need right from the start of the course in order to also understand the concept of service learning. This enabled them to begin sketching, ordering, creating and initiating a service plan for the respective school. The students noted that the communication basis and an orientation on the needs of the school were necessary traits in the beginning of service learning.

> "You try to fulfil a need in some way." (Discussion 3, 283–284)

In the course of the service implementation in schools, they reflected on the reciprocity of the experience and the fact that the learning went in both directions: the students learned from the teachers in the schools and vice versa. The teachers themselves were relieved that students already had relevant skills for the service delivery, e.g. knowing how to cut photos digitally or being able to construct a photo collage. The students involved in the "My fairy tale playground" project used the skills they had acquired in prior teacher training to activate the motor functions of children's hands and fingers by drawing and creating photo collages and then cutting them both out in analogue and digital form.

> "What helped me personally was teaching methodology in art, which I was trained in." (Discussion 4, 72–73).

In another example, general skills in the use of digital tools for teaching were already available but had to be substantially expanded during the service experience.

> "I had worked with the tool before but this function [quiz] was new to me." (Discussion 3, 439–440)

Learning and experiencing digital tools has become a relevant part of teacher training, and students are motivated to acquire digital skills when they see a professional context in which they can be applied.

When reflecting on their roles in the service implementation in schools, the students were surprised that they were not treated as 'interns' but as 'experts' because they were rendering a service for which they needed specific skills (e.g. constructing a fruit robot, cutting photos in digital form, using functions in online communication tools, etc.). This can be subsumed as part of academic learning, understanding and taking on a role in a professional context in a school community. They felt that they were contributing to supporting the school in solving a practical problem.

> "The fact that she said we were the experts who were coming to the school. I had never experienced that before." (Discussion 1, 225–226)

Communication and cooperation with the headteachers were considered to have been on equal terms.

> "She didn't say, well you are students, and you still have to learn, and so on. She viewed us as experts, although we are not actually experts yet." (Discussion 1, 315–317)

When working with pupils in schools, students also experienced reciprocal learning. In the "Working creatively with Microsoft Teams" project, for example, the students asked pupils to produce a video to show what they had already learned about the online communication tool. They then gave them an exercise to learn a new function. Both the pupils and the students assumed active roles in the learning process and produced digital knowledge for each other.

> "I like the idea that we can turn pupils into experts. This is something you can actually use in a lot of pedagogical situations in school. Through this, they are much more motivated to participate in my opinion. […] I think that can be useful for different subjects." (Discussion 3, 309–317)

Here, the transfer process between the current service learning experience and its relevance for their future profession became visible. The students learned that they could apply digital tools in other subjects and topics as well. They felt in general that they had tried to learn everything they could about practice in school, about teaching lessons, teacher attitudes towards projects or the reception of people external to the school.

The focus groups were not as rich a source of data as the learning diaries. Since they were more conversational and allowing room to react to pressing concerns during service implementation, the focus groups were not particularly useful for course assessment purposes. Given the data they produced, it is clear that focus group discussions served a counselling, guidance and support function more than a formative assessment one.

4.3 Online Presentations

Each online presentation had a formal and a creative part. The formal presentation showed the duration of the service, the objectives and work packages associated with it and the answers to reflection questions. The creative part showed visual or audio results of the services (e.g. videos by students, drawings by pupils).

The core concepts of academic learning, civic learning and personal growth were reflected on in the final online presentations. Students saw the greatest benefit to lie in their personal growth rather than in academic or civic learning. When asked about their personal growth, they explained that they were challenged in terms of creativity, time management, designing realistic objectives, flexibility, (digital) communication and managing people in digital spaces. From an academic learning perspective, they stressed the positive effects of service learning on their specialized knowledge as teachers (e.g. didactics, psychology, motivating others), working with multi-level classes and getting to know distance education. In terms of civic learning, they stated having gained an enhanced sense of empathy and problem-solving skills.

5 Conclusions and Implications

All in all, the findings of this study should be taken with caution since service learning courses deviate in terms of their embedding in the curriculum, number of credits and, thus, hours of teaching, teaching methodology, community service and assessment plan. A context-sensitive comparison is therefore needed (Slepcevic-Zach & Fernandez, 2021). However, there are different learning fields to be taken from the assessment of such coursework. In general, students reported a positive increase in their academic, civic and personal skills after e-service learning, but most of all in their personal growth.

Assessing digital learning is difficult because communicating online means losing emotions, gestures and non-verbal signals. It may therefore be more difficult to build trust among students in e-service learning. In the course described in this paper, however, instruction did take place onsite, and trust between peer students was therefore established quickly and became a success factor, as students reported being able to rely on each other. When implementing e-service learning, clear lines of communication would appear useful and necessary, preceded by a clear hierarchy of project members. Waiting for answers in asynchronous communication (email) is more common in e-service learning than in traditional service learning, where questions are discussed among participants onsite, and may lead to misunderstandings. It is also recommended to organise online appointments in good time in order to create commitment (Hunter, 2007). Creating space for reflection online is also a necessity, although the online focus group discussions in the case described in this paper were in fact not particularly useful in assessment terms because they were only weakly linked to the assessment plan. In hybrid Type II. and blended service learning, instruction occurs in class – as in traditional service learning – so only a few differences were observed in terms of course design. Establishing a solid partnership with the community requires the same amount of effort regardless of which type of e-service learning is chosen.

Assessments are too often designed without a clear understanding of their resource implications. In our case, the students complained about the high workload, as is common in service learning (Gelmon, 2017). The combination of the weekly learning diaries and the need to adapt their service plans to a changing (digital) reality generated a high workload. Students with a consumer identity (Naidoo & Jamieson, 2005) might not find service learning suitable for their needs. This suggests a need to critically review the consequences of consumerism on higher education, since consumerist mechanisms may “reform academic values and pedagogic relationships to comply with market frameworks” (Naidoo & Jamieson, 2005, p. 271).

In terms of assessment data, the written data obtained in this case was lengthy because a lot of data was produced (learning diaries and online presentations). In the learning diaries, students revealed information that was otherwise not available to the lecturer or the tutor. While most students found diary writing tedious and time consuming, the diaries were relatively easy to design and handle online. Most students were unfamiliar with the concept of diary writing as a pedagogic tool. Learning diaries can be understood as a place for self-assessment, recording one's progress and reflecting on experiences and skills (Han, 2011). The assessment plan gave structure to the writing task by providing weekly questions, reminders and deadlines yet at the same time valuing learner autonomy (Benson, 2008).

In contrast, the focus group discussions did not provide as much rich data and were more conversational in nature. The group dynamics were not as visible as they could have been in a face-to-face setting as nonverbal communication was not observable in the digital space. However, the oral interaction generated new ideas in the group, which was not the case for the individual learning diaries. The focus groups revealed that working in small groups of three to four students created an environment of interdependence, collaboration and solidarity (Han, 2011). Given that the focus group discussions did not produce rich data, it remains open whether they could assume a more effective role in students' learning in the specific context of e-service learning courses.

The overall experiences with these forms of assessments show that the use of learning diaries and group reflections needs to be flexible to fit the specific work progress of students, while online presentations are well suited for assessing digital learning outcomes. Students on the course were provided opportunities to interact with their teacher educator and peers to discuss, generate and share knowledge. "Assessment provides a valuable mechanism for communicating the value of one's work. (…) The very nature of assessment necessitates a long-term perspective, as the assessment effort is never complete." (Gelmon, 2017, p. 43) To build a future evidence base on the effects of service learning on students, we hope to conduct further studies that capture the long-term perspective and trace how students develop and maintain civic skills after a service learning experience.

References

Benson, P. (2008). Teachers' and learners' perspectives on autonomy. In T. Lamb & H. Reinders (Eds.), *Learner and teacher autonomy: Concepts, realities, and responses* (pp. 15–32). Benjamins.

Berman, S. (2006). *Service learning: A guide to planning, implementing, and assessing student projects.* Corwin Press.
Bringle, R., Hatcher, J., & McIntosh, R. (2006). Analyzing Morton's typology of service paradigms and integrity. *Michigan Journal of Community Service Learning, 13*(1), 5–15.
Chalkley, B. (2006). Education for sustainable development: Continuation. *Journal of Geography in Higher Education, 30*(2), 235–236.
Chambers, D., & Lavery, S. (2017). Introduction to service-learning and inclusive education. *International Perspectives on Inclusive Education, 12,* 3–19.
Davies, A. (2002, Nov. 1). *Writing learning outcomes and assessment criteria in art and design.* http://arts.brighton.ac.uk/__data/assets/pdf_file/0003/67278/Writing-Learning-Outcomes-and-Assessment-Criteria-in-Art-and-Design.pdf.
Eyler, J. (2002). Reflection: Linking service and learning—Linking students and communities. *Journal of Social Issues, 58*(3), 517–534.
Farber, K. (2011). *Change the world with service learning. How to create, lead, and assess service learning projects.* Rowman & Littlefield Education.
Felten, P., & Clayton, P. H. (2011). Service-learning. *New Directions for Teaching and Learning, 128,* 75–84.
Furco, A. (2003). Issues of definition and program diversity in the study of service-learning. In S. Billig & A. S. Waterman (Eds.), *Studying service learning: Innovations in education research methodology* (pp. 13–33). Routledge.
Furco, A. (2009). Die Rolle von Service Learning im Aufbau einer gesellschaftlich engagierten Universität. In K. Altenschmidt, J. Miller, & W. Stark (Eds.), *Raus aus dem Elfenbeinturm? Entwicklungen in Service Learning und bürgerschaftlichem Engagement an deutschen Hochschulen* (pp. 47–59). Beltz.
Fromm, S. (2019). Evaluation im Kontext von Service Learning. In D. Rosenkranz, S. Roderus, & N. Oberbeck Eeds.), *Service Learning an Hochschulen. Konzeptionelle Überlegungen und innovative Beispiele* (pp. 110–115). Beltz Juventa.
García-Gutierrez, J., Ruiz-Corbella, M., & Del Pozo Armentia, A. (2017). Developing civic engagement in distance higher education: A case study of virtual service-learning (vSL) programme in Spain. *Open Praxis, 9*(2), 235–244. https://doi.org/10.5944/openpraxis.9.2.578.
Gelmon, S. B. (2017). *Assessing service-learning and civic engagement. Principles and techniques.* Campus Compact.
Govekar, M. A., & Rishi, M. (2007). Service-learning: Bringing real-world education into the b-school classroom. *Journal of Education for Business, 83,* 3–10.
Han, A. (2011). Gaining insights into student reflection from online learner diaries. *Language Learning in Higher Education, 1*(1), 195–210.
Harden, R. M. (2002). Learning outcomes and instructional objectives: Is there a difference? *Medical Teacher, 24*(2), 151–155.
Harris, U. S. (2017). Virtual partnerships: Engaging students in E-service learning using computer-mediated communication. *Asia Pacific Media Educator, 27*(1), 103–117. https://doi.org/10.1177/1326365X17701792.
Hildenbrand, S. M., & Schultz, S. M. (2015). Implementing service learning in pre-service teacher coursework. *Journal of Experiental Education, 38*(3), 262–279.
Hunter, D. (2007). The virtual student/client experience. *Journal of American Academy of Business, 12*(1), 88–92.

Jacoby, B. (2015). *Service-learning essentials: Questions, answers, and lessons learned*. Jossey-Bass.

Karayan, S., & Gathercoal, P. (2005). Assessing service-learning in teacher education. *Teacher Education Quarterly, 32*(3), 79–92.

Kaye, C. B. (2010). *The complete guide to service-learning: Proven, practical ways to engage students in civic responsibility, academic curriculum, and social action.* Free Spirit Publishing.

Naidoo, R., & Jamieson, I. M. (2005). Empowering participants or corroding learning?: Towards a research agenda on the impact of student consumerism in higher education'. *Journal of Education Policy, 20*(3), 267–281.

Prague Communiqué. (2001, May 19). Towards the European Higher Education Area. Communiqué of the meeting of European Ministers in charge of Higher Education in Prague on May 19th 2001. http://www.ehea.info/media.ehea.info/file/2001_Prague/44/2/2001_Prague_Communique_English_553442.pdf.

Prinsloo, P., Slade, S., & Galpin, F. (2011). A phenomenographic analysis of student reflections in online learning diaries. *Open Learning, 26*(1), 27–38.

Reinders, H. (2016). *Service Learning – Theoretische Überlegungen und empirische Studien zu Lernen durch Engagement.* Beltz Juventa.

Resch, K., & Dima, G. (2021). Higher education teachers' perspectives on inputs, processes, and outputs of teaching service learning courses. In L. Tauginienė & R. Pučėtaitė (Eds.), *Managing social responsibility in universities. organisational responses to sustainability* (pp. 117–136). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-70013-3.

Resch, K., & Schrittesser, I. (2021). Using the service-learning approach to bridge the gap between theory and practice in teacher education in Austria. *International Journal of Inclusive Education, 27*(10), pp. 1118–1132. https://doi.org/10.1080/13603116.2021.1882053.

Resch, K., Schrittesser, I, & Knapp, M. (2022). Overcoming the theory-practice divide in teacher education with the 'Partner School Programme'. A conceptual mapping. *European Journal of Teacher Education.* https://doi.org/10.1080/02619768.2022.2058928.

Roessingh, H., & Chambers, W. (2011). Project-based learning and pedagogy in teacher preparation: Staking out the theoretical mid-ground. *International Journal of Teaching and Learning in Higher Education, 23*(1), 60–71.

Scott, K. E., & Graham, J. A. (2015). Service-learning: Implications for empathy and community engagement in elementary school children. *Journal of Experiential Education, 38*(4), 354–372.

Slepcevic-Zach, P., & Gerholz, K.-H. (2015). Service-Learning – Entstehung, Wirksamkeit und konkrete Umsetzung. In E. Augustin, M. Hohenwarter, G. Salmhofer, & L. Scheer (Eds.), *Theorie, die ankommt. Grazer Beiträge zur Hochschullehre* (pp. 61–76). Leykam.

Slepcevich-Zach, P., & Fernandez, K. (2021). Effectiveness of service-learning. *Zeitschrift für Hochschulentwicklung, 16*(1), 77–96.

Waldner, L. S., McGorry, S. Y., & Widener, M. C. (2012). E-service-learning: The evolution of service-learning to engage a growing online student population. *Journal of Higher Education Outreach and Engagement, 16*(2), 123–150.

Yorio, P. L., & Ye, F. (2012). A meta-analysis on the effects of service-learning on the social, personal, and cognitive outcomes of learning. *Academy of Management Learning & Education, 11*(1), 9–27.

Katharina Resch is an educational sociologist and post-doctoral researcher at the University of Vienna, Austria. Her primary work focuses on inequalities in education, higher education research and didactics, diversity, and teacher education. She has initiated several European projects on Service-Learning aiming at strengthening the Service-Learning community in Austria and beyond. She is one of the founders of the Research Network "Engaged Universities" in Austria.

Validity and Fairness of an Admission Examination Using a Study-Related Learning Test

Ingrid Wahl, Christa Walenta, and Günther Wenzel

Abstract

Good admission procedures reflect the requirements of the aspired program and, thus, help select applicants who will finish their studies successfully. Accordingly, admission procedures should predict students' grades and dropout. A distance learning study program provided applicants with relevant learning materials in advance and used a study-related learning test on this material in an online setting to select students. In total 236 applicants participated in the four waves of the admission examination of which the best-ranked 82 were admitted. Results showed that (a) applicants in the first wave scored higher than applicants in the fourth wave, (b) the test can be regarded as fair concerning gender and age but not in terms of form of entrance and citizenship, and (c) the test is valid in predicting students' grade average in the first semester and shows a tendency in predicting dropout in the first semester but is not valid in predicting grade average and dropout in the second semester. These results are discussed regarding time of application, school experiences, language skills, and environmental factors. As the test informs applicants about the study program's content, proficiency level, study format, and the

I. Wahl (✉)
Department of Communication, University of Vienna, Vienna, Austria
e-mail: Ingrid.wahl@univie.ac.at

C. Walenta · G. Wenzel
Ferdinand Porsche FernFH, Wiener Neustadt, Austria
e-mail: Christa.walenta@fernfh.ac.at

G. Wenzel
e-mail: Guenther.wenzel@fernfh.ac.at

S. Hummel and M. Donner (eds.), *Student Assessment in Digital and Hybrid Learning Environments*, Doing Higher Education,
https://doi.org/10.1007/978-3-658-42253-0_14

necessity of self-regulated learning it is a good measure to clarify applicants' expectations.

Keywords

Admission procedure • Academic success • Distance learning • Entrance exam • Online admission test

1 Introduction

Universities of applied sciences in Austria have to conduct admission procedures if the number of applicants exceeds the available places. The aim is to select suitable applicants, who will finish their studies successfully (i.e., good grades, not dropping out). Good admission procedures help selecting applicants who will finish their studies successfully which in turn legitimates the effort for constructing and administering them (Messerer & Humpl, 2003; Schuler & Hell, 2008). Further, admission tests should be reliable, objective, economically applicable, fair across different groups of applicants, and should be accepted by applicants (Trost, 2003, as cited in Trost & Haase, 2005). Good admission procedures also reflect the required proficiency level of the aspired program, so that applicants can screen in advance whether they consider choosing the program (Arnhold & Hachmeister, 2004).

Professional experts and experts for distance learning established a profile of requirements for the distance learning study program "Business Administration & Psychology (Bachelor)". These requirements include students' cognitive skills, language skills in German as well as English, achievement motivation, but also being informed about the program's content and the readiness to learn relatively self-organised and being able to learn self-regulated. On the basis of these requirements and the fact that the number of applications for the study program exceeded the number of available places, an admission procedure including several tests and a structured interview was developed. For this procedure, applicants were provided with learning materials (i.e., text book, assisting online course) on basic content matters of the study program. They had to take an on-site admission test about the provided learning material. This, so-called study-related learning test was previously found to be valid and fair (Wahl & Walenta, 2017). However, after sending out the learning materials in spring 2020 the test had to be held online due to Austrian COVID-19 regulations. To avoid dissemination of the questions of the on-site test used in the previous years a new test for an online setting was constructed.

The current paper aims at evaluating the newly constructed online admission test's fairness and validity. Fairness is examined using socio demographic data of applicants and validity is assessed with students' study success. First, the requirement to learn self-regulated will be discussed since this is an important success factor in distance learning programs. Second, different admission procedures will be introduced and relevant quality criteria (i.e., validity, fairness, objectiveness, etc.) will be discussed. Third, it is reviewed which admission procedures inform about the required proficiency level and whether they capture implicitly or explicitly the ability to learn self-regulated. Fourth, results on the fairness and validity of the newly constructed study-related learning test are presented and discussed.

1.1 Self-regulated Learning: A Core Factor for Learning Success

Admission procedures should help selecting the most qualified applicants. Thus, they should not only assist in choosing the most qualified candidates in terms of cognitive or technical skills, but also the most motivated ones who have the potential to finish their studies in time. They should also provide information on the fitting of the applicants with the aspired study format (e.g., acquiring contents autonomously in a distance learning format or when studying while also working; Messerer & Humpl, 2003; Wahl & Walenta, 2018).

Simonson and Berg (2016, p. 1) define distance learning as a "form of education in which the main elements include physical separation of teachers and students during instruction and the use of various technologies to facilitate student–teacher and student–student communication". Distance learners – contrary to students in class – are irregularly present at their universities. Consequently, they are lacking face-to-face communication; however, they have the possibility to organize their learning flexibly by determining for example where, when, at which speed, with which intensity, and how they prefer to learn (Wenzel et al., 2019). This demands a high degree of self-organization when enrolling at a distance learning university (Kerres & Jechele, 2001). Accordingly, distance learners show a higher learning motivation and assess their learning activities to be more self-regulated compared to learners in traditional learning settings (Konrad, 2000).

Self-regulated learning is a process in which learning is actively planed, organized, edited, and evaluated (Knowles, 1975). Self-regulated learners have the skills to set their own learning goals as well as to choose and use appropriate techniques and strategies to achieve those goals. They also maintain their motivation, evaluate their goal achievement during and after the learning process, and rectify goals if necessary (Artelt et al., 2001). Motivational aspects (e.g., willingness to learn, interest) and environmental factors (e.g., learning environment, family, occupation) play a central role for the success of self-regulated learning (Deimann et al., 2008; Nenninger et al., 1996; Seeber et al., 2006; Straka, 2008). Distance learning students have to take more responsibility for the layout of their learning process than students on-site; thus, their ability for learning in a self-regulated way is particularly called for (Keller et al., 2004; McGivney, 2004). Accordingly, a questionnaire study found that successful distance learning students assessed their motivation to be higher compared to less successful distance learning students (Creß & Friedrich, 2000).

Suitable learning strategies also foster productive self-regulated learning (Schiefele et al., 2003; Seeber et al., 2006). Meta-cognitive learning strategies refer to planning, monitoring, and the endeavour to understand the learning material (Schober et al., 2016; Seeber et al., 2006). Successful distance learning students show more often the meta-cognitive strategy of self-monitoring than students on-site and less successful distance learning students (Konrad, 2000).

To predict an applicant's study success and to reduce dropouts in a distance learning program it is necessary to consider the special situation and requirements when selecting students. Thus, reliable admission procedures for distance learning study programs should evaluate implicitly and explicitly applicants' ability for self-regulated learning (Wahl & Walenta, 2017).

1.2 Admission Procedures at Universities

To select students for study programs admission procedures such as average grades, school achievement tests, intelligence tests, and admission interviews are used (Hell, 2006; Rindermann & Oubaid, 1999; Salvatori, 2001; Spiel et al., 2007; Trost, 2003, as cited in Trost & Haase, 2005). A less often used procedure are study-related learning tests, which assess applicants' knowledge about defined learning material relevant for the aspired study program. The following section describes the named admission procedures in more detail and presents studies concerning their validity and give information about their reliability, objectiveness, economical applicability, fairness, and acceptance. Further, it is reviewed whether they are suitable to convey the required proficiency level of the study program and whether they implicitly or explicitly capture the ability for self-regulated learning.

Average grades (i.e., HSGPA = high school grade point average in the US or average grade of school leaving examination in German-speaking areas) are widely used. They are easily accessible and predict students' success reliably for different study programs (Atkinson & Geiser, 2009; Geiser & Santelices, 2007; Gold & Souvignier, 2005; Kreiter & Kreiter, 2007; Kuncel et al., 2001; Salvatori, 2001; Troche et al., 2014). As grades are to some extent dependent on the teaching staff and on the school visited, they are regarded to lack reliability, objectiveness, and fairness; however, they are perceived to be economically applicable (Spiel et al., 2007; Trost & Haase, 2005) and are widely accepted by students (Hell & Schuler, 2005).

Approximately 12-13% of students at Austrian universities of applied sciences do not hold a conventional school leaving examination, but hold a relevant professional qualification and have to pass additional tests in study relevant subjects (e.g., vocational school leaving certificate, certificate of general educational development, qualification examinations in relevant subjects; Nitsch, 2006; Statistik Austria, 2021). Thus, these students cannot be enrolled based on their school leaving examination's average grade. Furthermore, distance learning students often do not start to study directly after school and are often already working and are therefore older than on-site students (Dibiase, 2000; Peters, 1975). This raises the question whether it is meaningful to use school achievements dating back several years. Additionally, basing student selection on the applicants' average high school grades is not intended by the University of Applied Sciences Studies Act (Fachhochschulgesetz, 1993 § 11 as amended; Messerer & Humpl, 2003). When using high school grades applicants are not informed about the proficiency level of the aspired study program or other institutional conditions and universities are not provided with any information about the applicants' study habits, motivation, or their ability to learn in a self-regulated manner.

Spiel et al. (2007) introduced *study-related learning tests* in which applicants receive in advance learning material relevant for the study program's field and take an exam in the form customary for the aspired study program. This test introduces the basic conditions of the study program and examines technical knowledge together with needed cognitive abilities. Furthermore, applicants with less current learning experiences or non-native speaking applicants have the opportunity to actively learn for the test, which might compensate for disadvantages these groups face. A first evaluation showed that study-related learning tests predict students' success (i.e., average grades, dropout) and can therefore be assumed to be valid. It was also shown that these tests are fair concerning different sociodemographic groups (Wahl & Walenta, 2017). It could be assumed that study-related learning tests represent an applicants' first academic performance. Thus, also studies on predicting future study achievements based on current test performances could be informative concerning the validity of study-related learning tests. Brandstätter and Farthofer (2003) showed that test achievements in the first year of studies predicted the study success of the following semesters. Also, grades in a bachelor study program predict students' study success in a similar master program (Hell, 2006; Troche et al., 2014). Students in study programs reflecting their interests are more successful than students whose interests are not reflected in their studies (Brandstätter et al., 2001; Hasenberg & Schmidt-Atzert, 2014). Thus, the insight applicants gain through the preparation for the study-related learning test might have a positive effect. Further, the preparation for the exam and the results of the exam can be regarded as a work sample (Kasper & Furtmüller, 2005). Work samples are often used for personnel selection and show a high prognostic validity (Hunter & Hunter, 1984; Robertson & Downs, 1989; Roth et al., 2005; Schmidt & Hunter, 1998). Using work samples in admission procedures is accepted to be reliable, objective, economically applicable, and fair (Spiel et al., 2007). Also, they are widely accepted by students (Hell & Schuler, 2005). As study-related learning tests are simulating the studying situation and the examination situation applicants are informed about the required proficiency level of the aspired study program and universities can also implicitly draw conclusions whether applicants are capable of self-regulated learning.

School achievement tests are standardized tests on field-specific school knowledge relevant for a study program (Rindermann & Oubaid, 1999; Trost, 2003, as cited in Trost & Haase, 2005). Examples are standardized tests on foreign languages like the TOEFL (Test of English as a foreign language) which are presented during admission procedures (Trost & Haase, 2005). Contrary to study-related learning tests school achievement tests only examine knowledge acquired at school and neglect knowledge and competences specific for a study program (Spiel et al., 2007). International school achievement tests are the ACT (American College Test), the SAT and the SAT II subject tests (Scholastic Assessment Test), the AP exams (Advanced Placement), the DAT (Dental Admission Test), the GRE subject tests (Graduation Record Examination), the SweSAT (Swedish Scholastic Assessment Test), and the Chilean achievement test for mathematics, verbal language skills, social sciences, and science (Atkinson & Geiser, 2009; Geiser & Santelices, 2007; Koljatic et al., 2013; Kuncel et al., 2001; Lyrén, 2008; Richardson et al., 2012). Especially the SAT II subject tests and the AP exam are valid in predicting study success (Atkinson & Geiser, 2009; Geiser & Santelices, 2007), whereas, the Chilean school achievement tests are similarly valid as other assessment procedures (Koljatic et al., 2013). Overall, school achievement tests are rated to be valid, reliable, objective, and economically applicable (Spiel et al., 2007; Trost, 2003, as cited in Trost & Haase, 2005); however, fairness issues arise as minorities face more disadvantages in school achievement tests than in other admission procedures (Geiser & Santelices, 2007). In German-speaking countries standardized school achievement tests – other than tests on foreign languages – are not available for student admission testing. School achievement tests can be used to convey a study program's proficiency level; however, those tests cannot assess the ability for self-regulated learning either implicitly or explicitly as they only examine knowledge acquired at school.

For successfully finishing one's studies cognitive abilities such as deductive reasoning, verbal competences, learning aptitude, and retentiveness are needed. Cognitive abilities can be measured with *intelligence tests* (Spiel et al., 2007). It was shown that standardized tests assessing cognitive abilities predict students' study success and are therefore valid (Allareddy et al., 2012; Hell et al., 2007; Koljatic et al., 2013; Kreiter & Kreiter, 2007; Poropat, 2009; Richardson et al., 2012; Rohde & Thompson, 2007; Salvatori, 2001; Strenze, 2007). Intelligence tests are regarded to be reliable, objective, economically applicable, and fair (Spiel et al., 2007); however, students meet them with little acceptance (Hell & Schuler, 2005). These tests are not able to communicate the proficiency level of a study program and they do not assess the ability for self-regulated learning.

Admission interviews are often used for selecting students. Their prognostic validity is rather low; however, structured interviews show better results than unstructured interviews (Hell, 2006). Different studies which show significant and non-significant correlations between admission interviews and study success indicate an insufficient validity (for an overview see Salvatori, 2001). Admission interviews are also regarded as scarcely reliable, objective, economically applicable, and fair (Salvatori, 2001; Spiel et al., 2007); however, students accept them widely (Hell & Schuler, 2005). Further, admission interviews provide the possibility for a personal meeting, which allows students as well as universities to gather further information, which might be relevant for each side to draw the final decision. During admission interviews information about the required proficiency level can be communicated. Although the ability for self-regulated learning and expectations regarding learning preferences can be asked explicitly in admission interviews, applicants could try to present themselves as more positive.

As reported above, most of the mentioned admission procedures are frequently used in different settings and therefore, are already widely tested regarding their fairness and validity. However, study-related learning tests lack these verifications, as so far they were only tested once (Wahl & Walenta, 2017). Thus, the current study aims to examine whether the newly constructed study-related learning test for a study program can also be regarded as fair and valid.

2 Method

2.1 Participants

During spring 2020 a total of 236 applicants participated in the admission procedure and finished the newly constructed study-related learning test. Of these applicants 170 were female and 66 male; with age ranging between 19 and 55 years ($M = 29.88$, $SD = 7.65$). As proof for university entrance 217 held a school leaving examination and 80 were allowed to study in Austria based on another qualification. An Austrian citizenship is held by 217 applicants whereas 19 hold a citizenship of another country.

The best-ranked applicants were invited to a follow-up personal online interview and were offered a place in the study program "Business Administration & Psychology (Bachelor)"; however, five declined their allocated place leaving 82 applicants who enrolled. Of these 60 were female and 22 male aged between 20 and 53 years ($M = 30.39$, $SD = 6.96$). As proof for university entrance 66 held a school leaving examination and 16 were allowed to study in Austria due to another qualification. An Austrian citizenship is held by 77 of the enrolled applicants whereas 5 hold a citizenship of another country.

2.2 Procedure and Material

In preparation for the study-related learning test, all applicants received the learning material via e-mail at least three weeks before the test. The text book consisted of an introduction to the basics of business administration and business psychology to the extent of 105 pages and contained excerpts from study materials used in the first semester, learning control questions, and solutions. In addition, applicants were invited to an online course that includes general information about the program, a learning quiz, and a discussion forum. The online course corresponded to the content of the text book.

In this way, applicants were able to experience the study conditions of the study program and acquired the basic knowledge necessary for starting their studies of Business Administration and Psychology. Further, applicants also gained an initial insight into the relevant contents of the study program and could decide in advance whether they are interested. Due to the independent preparation for the study-related learning test, which is similar to the learning situation in the later distance learning phases, it can be assumed that the ability of self-regulated learning is also indirectly ascertained by the learning test.

The admission examination was held in four waves according to incoming applications. In order to reduce the possibility that the questions of the newly constructed study-related learning test were disseminated competency-based questions were constructed, in three different versions using comparable topics and answers. This type of questions allows to measure higher-order thinking (e.g., application, analysis) rather than only recalling facts (Gyll & Ragland, 2018; Scully, 2017).

As applicants took the test online at home, using competence-based questions also reduced the probability for simply looking up answers on the internet. In the first three waves one version of each question was provided and in the fourth wave one of the three versions was selected in turns. The new test consisted of 49 questions regarding the provided learning material, with 13 items showing a statement which could be answered as true or false and 36 multiple-choice questions with four answering alternatives of which one to three alternatives were correct.

A statement which should be answered as true or false was: "To operate successfully means for commercial companies to be able to continuously meet the necessary payments in terms of time and amount." An example of a multiple-choice question regarding business administration is: "Which of the following statements do not represent a conflict between microeconomic and macroeconomic goals? (a) To reduce costs, vacancies will no longer be filled; (b) As a result of the requirement to offer products on the market as cheaply as possible, production takes place in low-wage countries; (c) Since consumers are willing to pay higher prices for organic products, production is being switched to organic products; (d) Impairment of the environment is accepted for rapid economic growth." An example of a multiple-choice question in business psychology was: "A market researcher wants to test experimentally whether a new advertisement has an influence on buying behaviour. In this context is advertising (a) the independent variable; (b) a confounding variable; (c) the dependent variable; (d) an active variable."

The maximum test score achievable was 170 points (i.e., 2 points for correctly answering a true or false question and a maximum of 4 points for correctly answering a multiple-choice question). Sociodemographic data (i.e., applicants' gender, age, highest education, citizenship) and study success (i.e., average grades in the first and second semester, enrolment status) were collated from the university's database.

3 Results

3.1 Comparison of the Four Waves

To analyse whether there are differences between the four waves a univariate analysis of variance with the four waves as independent variable and the achieved test score of the study-related learning test as dependent variable was conducted. The analysis revealed a significant effect between subjects ($F(3232) = 3.58$; $p = 0.015$; $\eta^2 = 0.04$). The Scheffé test showed that scores of the first wave were significantly higher than scores of the fourth wave; however, no differences between either of these waves and the second or third wave was found. Means and standard deviations are provided in Table 1.

Table 1 Means and standard deviations of test scores on the study-related learning test by waves, participants' gender, participants' form of entrance, and participants citizenship

	N	*M*	*SD*
Waves			
First wave	59	112.70^{a}	18.76
Second wave	61	105.28^{ab}	19.30
Third wave	64	104.26^{ab}	18.61
Fourth wave	52	101.13^{b}	21.74
Applicants' gender			
Female	170	106.10^{a}	19.32
Male	66	105.54^{a}	21.37
Applicants' form of entrance			
School leaving examination	156	109.75^{a}	17.36
Other form of entrance	80	98.52^{b}	22.33
Applicants' citizenship			
Austrian	217	107.09^{a}	19.18
Other country	19	92.84^{b}	23.29

Note. Different suprascripts for an independent variable indicate differences on a level of $p < 0.05$.

(a) the first three waves do not differ (i.e., all have an "a"), (b) the first and the fourth wave differ (i.e., they have no suprascript in common), and (c) that the last three waves do not differ (i.e., they all have a "b")

3.2 Fairness

As admission tests should be fair concerning different groups of applicants, different sociodemographic variables were used to compare test results. Three separate t-tests with participants' gender, whether they held a school leaving examination or are allowed to study in Austria due to another qualification, and whether they are Austrian citizens or citizens of another country as independent variables and applicants' test scores on the study-related learning test as dependent variable were performed. No differences between the test scores of women and men ($t(234) = 0.20$; $p = 0.845$) were found. Applicants with a school leaving examination had significantly higher test scores than applicants with another qualification ($t(234) = 4.25$; $p < 0.001$). Also, applicants from Austria performed better on the study-related learning test than applicants from other countries ($t(234) = 3.05$; $p = 0.003$). Means and standard deviations are depicted in Table 1.

To check whether applicants' age influences their test score a Spearman's correlation analysis was conducted. No significant correlation effect was found between applicants' age and their test score on the study-related learning test ($r_{sp} = -0.07$; $p = 0.275$).

3.3 Validity

Spearman's correlations were used to analyse the study-related learning test's validity on bases of the relation between enrolled students' performance and their grades in the first and second semester. For grades in the first semester a significant correlation was found ($r_{sp} = -0.26$; $p = 0.021$) but not for grades in the second semester ($r_{sp} = -0.17$; $p = 0.176$). Thus, in the first semester students with higher test scores achieved better school grades, were grades range from 1 (very good) to 4 (sufficient). Correlation coefficients are shown in Table 2.

To examine whether the study-related learning test can predict dropouts two separate Mann–Whitney-U-Tests[1] with students' enrolment status in the two semesters as independent variables and scores on the test as dependent variable were used. No significant differences were found between students still enrolled and students who dropped out after the first semester ($U = 153.50$; $p = 0.070$);

[1] In fact, results of the study-related learning test should be the independent variable (i.e., predictors) and enrolment status should be the dependent variables (i.e., criterion). Following this logic, a hierarchical binary logistic regression should be conducted. Exchanging the independent and dependent variables does not affect results but increases comprehensibility.

Table 2 Spearman's correlation coefficients of applicants' age and students' average grades in the first and second semester related to test scores on the study-related learning test

	Test scores on the study-related learning test
Applicants' age	−0.07
Students' average grades in the first semester	−0.26*
Students' average grades in the second semester	−0.17

Note. * indicates significant correlations on a level of $p < 0.05$; grade averages relate to positively passed exams and range from 1 (very good) to 4 (sufficient)

Table 3 Medians, 25% quartile, and 75% quartile of test scores on the study-related learning test by enrolment status of the first and second semester

	N	*Md*	25% quartile	75% quartile
First semester's enrolment status of students				
Enrolled	75	124.00[a]	117.33	132.00
Dropped out	7	116.00[a]	111.33	130.67
Second semester's enrolment status of students				
Enrolled	70	124.00[a]	117.33	132.00
Dropped out	12	122.00[a]	113.00	130.67

Note. Different suprascripts for an independent variable indicate differences on a level of $p < 0.05$

however, a tendency is discernible. For the second semester no significant differences were found for the compared groups ($U = 362.00$; $p = 0.446$). Medians, 25% quartiles, and 75% quartiles are depicted in Table 3.

4 Conclusion and Implications

Study-related learning tests are used for student selection through assessing applicants' knowledge about previously handed out material relevant for the study program. Due to Austrian COVID-19 regulations, a study-related learning test of a distance learning study program had to be newly constructed and was held online in four test waves. The present paper examines this test's fairness concerning different groups of applicants and this test's validity in selecting future successful students. First, applicants in the first wave had better results than applicants in the fourth wave. Second, the test is fair concerning gender and age; however, it discriminates applicants who do not hold a school leaving examination and who are not Austrian citizens. Third, the study-related learning test is valid in predicting students' average grades in the first semester; however, it cannot predict students' grades in the second semester or students' dropout.

Applicants were allocated to the four waves in sequence of submitting their applications, with very early submissions being tested in the first wave and very late submissions being tested in the fourth wave. A study considering assignments' submission time showed that early submissions predict good results, whereas, submissions closer to the deadline are less likely to score well (Młynarska et al., 2016). Thus, early applications might reflect a considerate decision for a particular study program accompanied with high commitment to the aspired goal of studying, whereas late applications might indicate a less thoughtful decision with less commitment. Accordingly, early applicants might be better prepared for the study-related learning test than later applicants. This could explain the found differences between the first and the fourth wave of the admission test.

People not holding a school leaving examination have fewer school experiences than people holding a school leaving examination do. The absence of experience with taking tests – especially with a multiple-choice answering format – might be responsible for the disadvantages applicants without a conventional school leaving examination face. The University of Applied Sciences Studies Act reflects this and orders to consider applicants' form of entrance in bachelor's degrees accordingly (Fachhochschulgesetz, 1993 § 11 as amended) to ensure equal educational opportunities and inclusion. This accommodation helps applicants with less school experiences to start studying, although, they might otherwise not be selected because of their performance in admission procedures. Thus, the study-related learning test at hand might still be fair as it is likely that applicants not holding a school leaving examination perform worse in admission procedures than applicants holding a school leaving examination. Still, this group should be especially supported (e.g., supplementary courses for levelling, additional test questions in the assisting online course).

Citizens of other countries than Austria whose native language might not be German perform worse in the study-related learning test than Austrian citizens. In this regard the study-related learning test is not fair. Multiple-choice questions must be easy to understand and must follow a simple structure, otherwise especially – but not exclusively – non-native speakers often struggle with answering them. Abedi (2006) showed performance differences between native English speakers and non-native English speakers in assessments and argues that these differences are not due to lacking knowledge of the contents but may be due to lacking language skills. Questions should be scrutinised to omit difficult phrasing so that knowledge of the materials' content is assessed instead of language skills.

The previously used study-related learning test was found to be valid (Wahl & Walenta, 2017); however, the newly constructed test is only partially valid. It predicted the grade average of the first semester and showed a tendency to predict dropouts in the first semester. The result on dropouts might be affected by the group size of students as only seven students dropped out in the first semester. For the second semester neither grade average nor drop out could be predicted by test scores achieved in the study-related learning test. An explanation for the lack of predictive power in the second semester is that students get accustomed to the study program's proficiency level during the first semester. As academic challenges are mastered, environmental factors concerning students' families or occupations might interfere with students' study success (Deimann et al., 2008; Nenninger et al., 1996; Seeber et al., 2006; Straka, 2008). These obstacles can arise for every student and are unrelated to their academic skills.

Work samples have a good prognostic validity (Hunter & Hunter, 1984; Robertson & Downs, 1989; Roth et al., 2005; Schmidt & Hunter, 1998). Taking an exam under similar conditions as the aspired study program could be regarded as a work sample (Kasper & Furtmüller, 2005). Further, study-related learning tests provide applicants with information about future study conditions and inform about basic knowledge of the study program before starting the studies. Therefore, they can make an informed decision, which might reduce dropout. The low dropout rate found in the current study might be a first indication of this assumption.

Trost (2003, as cited in Trost & Haase, 2005) postulates that additionally to validity admission procedures should show the quality criteria of reliability, objectivity, economic applicability, fairness, and acceptance. The current study did not evaluate the study-related learning tests' reliability, but due to the similar construction used in school achievement tests it could be assumed as given. As true or false questions and multiple-choice questions can be used the study-related learning test is objective and economic in administering. Due to the overlapping content with the study program and the resulting high level of face validity, it can be assumed that applicants regard it as transparent and therefore meet it with a high level of acceptance (Spiel et al., 2007).

Overall, the main implication is that other study programs could also use study-related learning tests in addition to their existing admission procedures. For that purpose, they need to provide learning materials conveying their basic contents in the form customary for the study program and construct a corresponding test. However, the prognostic validity of each newly constructed study-related learning test needs to be tested to ensure that successful future students are selected. A further implication of using study-related learning tests is that applicants can decide whether they are comfortable with the study conditions and whether they are still interested in the intended subject. These considerations of applicants could considerably reduce dropouts especially at the start of the study program and therefore decrease universities' costs for study places not used. Additionally, as study-related learning tests can also be seen as a form of work sample, applicants might be more willing to accept them compared to other admission procedures.

In addition to average grades and enrolment status, other criteria such as study satisfaction, successful graduation, and professional success provide information about study success (Schuler & Hell, 2008). Future research could use these criteria when evaluating admission procedures; however, high demands are placed on the collection of the needed personal data.

References

Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record, 108*(11), 2282–2303.

Allareddy, V., Howell, T. H., & Karimbux, N. Y. (2012). Association between students' dental admission test scores and performance on comprehensive clinical exams. *Journal of Dental Education, 76*(2), 168–173. https://doi.org/10.1002/j.0022-0337.2012.76.2.tb05243.x.

Arnhold, N., & Hachmeister, C.-D. (2004). *Leitfaden für die Gestaltung von Auswahlverfahren an Hochschulen.* Centrum für Hochschulentwicklung. http://www.che.de/downloads/Gestaltung_Auswahlverfahren_AP52.pdf.

Artelt, C., Demmrich, A., & Baumert, J. (2001). Selbstreguliertes Lernen. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiß (Eds.), *Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 271–298). Leske + Budrich.

Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher, 38*(9), 665–676. https://doi.org/10.3102/0013189X09351981.

Brandstätter, H., & Farthofer, A. (2003). Einfluss von Erwerbstätigkeit auf den Studienerfolg. *Zeitschrift für Arbeits- und Organisationspsychologie, 47*(3), 134–145. https://doi.org/10.1026//0932-4089.47.3.134.

Brandstätter, H., Farthofer, A., & Grillich, L. (2001). Die Stabilität der Studienwahl als Funktion von Interessenkongruenz, Selbstkontrolle und intellektueller Leistungsfähigkeit. *Psychologie in Erziehung und Unterricht, 48*(3), 200–218.

Creß, U., & Friedrich, H. F. (2000). Selbst gesteuertes Lernen Erwachsener: Eine Lernertypologie auf der Basis von Lernstrategien, Lernmotivation und Selbstkonzept. *Zeitschrift*

für Pädagogische Psychologie, 14(4), 194–205. https://doi.org/10.1024//1010-0652.14.4.194.
Deimann, M., Weber, B., & Bastiaens, T. (2008). *Volitionale Transferunterstützung (VTU) – Ein innovatives Konzept (nicht nur) für das Fernstudium* (IfBM.Impuls - Schriftenreihe Des Instituts Für Bildungswissenschaft Und Medienforschung). http://ifbmim-puls.fernuni-hagen.de/2008-01-Volitionale-Transferunterstuetzung.pdf.
Dibiase, D. (2000). Is distance education a Faustian bargain? *Journal of Geography in Higher Education, 24*(1), 130–136. https://doi.org/10.1080/03098260085216.
Fachhochschulgesetz. (1993). https://www.ris.bka.gv.at/Doku-mente/BgblPdf/1993_340_0/1993_340_0.pdf.
Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes.* http://www.escholarship.org/uc/item/7306z0zf.
Gold, A., & Souvignier, E. (2005). Prognose der Studierfähigkeit. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 37*(4), 214–222. https://doi.org/10.1026/0049-8637.37.4.214.
Gyll, S., & Ragland, S. (2018). Improving the validity of objective assessment in higher education: Steps for building a best-in-class competency-based assessment program. *The Journal of Competency-Based Education, 3,* e01058. https://doi.org/10.1002/cbe2.1058.
Hasenberg, S., & Schmidt-Atzert, L. (2014). Internetbasierte Selbsttests zur Studienorientierung. *Beiträge zur Hochschulforschung, 36*(1), 8–28.
Hell, B. (2006). Auswahl- und Zulassungsverfahren als Beitrag zur Qualitätssicherung im tertiären Bildungsbereich. In Hochschulrektorenkonferenz (Ed.), *Hochschulrektorenkonferenz* (Vol. 10, pp. 25–35). Hochschulrektorenkonferenz. https://www.uni-hohen-heim.de/studieneignung/publikationen/Vortrag%20FH%20Ber-lin%202006%20Hell.pdf.
Hell, B., & Schuler, H. (2005). Verfahren der Studierendenauswahl aus Sicht der Bewerber. *Empirische Pädagogik, 19*(4), 361–376.
Hell, B., Trapmann, S., & Schuler, H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. *Empirische Pädagogik, 27*(3), 251–270.
Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72–98. https://doi.org/10.1037/0033-2909.96.1.72.
Kasper, H., & Furtmüller, G. (2005). Die Studieneingangsphase als Auswahlinstrument für die Zulassung von Studierenden. *Zeitschrift für Hochschuldidaktik, 6*(Dezember), 19–33. http://www.zfhe.at/index.php/zfhe/article/view/164.
Keller, H., Beinborn, P., Boerner, S., & Seeber, G. (2004). Selbstgesteuertes Lernen im Fernstudium. Ergebnisse einer Studie an den AKAD Privathochschulen. *Schriften der Wissenschaftlichen Hochschule Lahr,* 1–61.
Kerres, M., & Jechele, T. (2001). Didaktische Konzeption des Tele-Lernens. In L. J. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia* (2nd ed., pp. 267–282). Beltz. http://mediendidaktik.uni-due.de/sites/default/files/kerres-jechle4issing_0.pdf.
Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers.* Follett.
Koljatic, M., Silva, M., & Cofré, R. (2013). Achievement versus aptitude in college admissions: A cautionary note based on evidence from Chile. *International Journal of Educational Development, 33*(1), 106–115. https://doi.org/10.1016/j.ijedudev.2012.03.001.

Konrad, K. (2000). Selbstgesteuertes Lernen: Differentielle Effekte unterschiedlicher Handlungsfelder und demographischer Variablen. *Unterrichtswissenschaft, 28*(1), 75–91.
Kreiter, C. D., & Kreiter, Y. (2007). A validity generalization perspective on the ability of undergraduate GPA and the medical college admission test to predict important outcomes. *Teaching and Learning in Medicine, 19*(2), 95–100. https://doi.org/10.1080/10401330701332094.
Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*(1), 162–181. https://doi.org/10.1037/0033-2909.127.1.162.
Lyrén, P.-E. (2008). Prediction of academic performance by means of the Swedish Scholastic Assessment Test. *Scandinavian Journal of Educational Research, 52*(6), 565–581. https://doi.org/10.1080/00313830802497158.
McGivney, V. (2004). Understanding persistence in adult learning. *Open Learning: The Journal of Open, Distance and e-Learning, 19*(1), 33–46. https://doi.org/10.1080/0268051042000177836.
Messerer, K., & Humpl, S. (2003). *Bewerbung—Auswahl—Aufnahme: Das Aufnahmeverfahren an österreichischen Fachhochschul-Studiengängen* (Vol. 7). WUV.
Młynarska, E., Greene, D., & Cunningham, P. (2016). *Indicators of good student performance in moodle activity data.* ArXiv:1601.02975 [Cs]. http://arxiv.org/abs/1601.02975.
Nenninger, P., Straka, G. A., Spevacek, G., & Wosnitza, M. (1996). Die Bedeutung motivationaler Einflußfaktoren für selbstgesteuertes Lernen. *Unterrichtswissenschaft, 24*(3), 250–266. https://doi.org/10.1007/978-3-663-14403-8_8.
Nitsch, F. (2006). *Hochschulstatistik.* Statistik Austria.
Peters, O. (1975). *Hinweise zu einem allgemeinen didaktischen Konzept für die Entwicklung von Studienmaterialien im Fernstudium.* Fernuni Hagen. http://deposit.fernuni-ha-gen.de/1695/1/ZP_001.pdf.
Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322–338. https://doi.org/10.1037/a0014996.
Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. https://doi.org/10.1037/a0026838.
Rindermann, H., & Oubaid, V. (1999). Auswahl von Studienanfängern durch Universitäten—Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs. *Zeitschrift für Differentielle und Diagnostische Psychologie, 20*(3), 172–191. https://doi.org/10.1024//0170-1789.20.3.172.
Robertson, I. T., & Downs, S. (1989). Work-sample tests of trainability: A meta-analysis. *Journal of Applied Psychology, 74*(3), 402–410. https://doi.org/10.1037/0021-9010.74.3.402.
Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*(1), 83–92. https://doi.org/10.1016/j.intell.2006.05.004.
Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology, 58*(4), 1009–1037. https://doi.org/10.1111/j.1744-6570.2005.00714.x.

Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Sciences Education, 6*(2), 169–175. https://doi.org/10.1023/a:1011489618208.
Schiefele, U., Streblow, L., Ermgassen, U., & Moschner, B. (2003). Lernmotivation und Lernstrategien als Bedingungen der Studienleistung. *Zeitschrift für Pädagogische Psychologie, 17*(3), 185–198. https://doi.org/10.1024//1010-0652.17.34.185.
Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262.
Schober, B., Jöstl, G., Klug, J., Wimmer, B., Spiel, C., Steuer, G., Schmitz, B., Ziegler, A., & Dresel, M. (2016). Kompetenzen zum Selbstregulierten Lernen an Hochschulen: Das Projekt PRO-SRL. In U. Zahn-Elliott (Ed.), *Bildungsforschung 2020: Zwischen wirtschaftlicher Exzellenz und gesellschaftlicher Verantwortung* (pp. 184–192). BMBF. https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/40770.
Schuler, H., & Hell, B. (2008). Studierendenauswahl und Studienentscheidung aus eignungsdiagnostischer Sicht. In H. Schuler & B. Hell (Eds.), *Studierendenauswahl und Studienentscheidung* (pp. 11–17). Hogrefe.
Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research, and Evaluation, 22,* Article 4. https://doi.org/10.7275/swgt-rj52.
Seeber, G., Boerner, S., Keller, H., & Beinborn, P. (2006). Strategien selbstorganisierten Lernens bei berufstätigen Studierenden. Ausgewählte Ergebnisse einer empirischen Untersuchung. *Journal für Sozialwissenschaften und ihre Didaktik, 2,* 1–19. https://doi.org/10.4119/jsse-368.
Simonson, M., & Berg, G. (2016). *Distance learning*. Encyclopedia Britannica. https://www.britannica.com/topic/distance-learning.
Spiel, C., Litzenberger, M., & Haiden, D. (2007). Bildungswissenschaftliche und psychologische Aspekte von Auswahlverfahren. In C. Badelt, W. Wegscheider, & H. Wulz (Eds.), *Hochschulzugang in Österreich* (pp. 479–552). Grazer Universitätsverlag.
Statistik Austria. (2021). Ordentliche Studierende an Fachhochschulen im Wintersemester 2020/21 nach Vorbildung. https://pic.statistik.at/wcm/idc/idcplg?IdcSer-vice=GET_PDF_FILE&RevisionSelectionMethod=LatestRelea-sed&dDocName=121270.
Straka, G. A. (2008). Selbstgesteuertes Lernen und Qualität der Lehre. *Neue Didaktik, 2008*(1), 1–12. https://doi.org/10.25656/01:7325.
Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*(5), 401–426. https://doi.org/10.1016/j.in-tell.2006.09.004.
Troche, S. J., Mosimann, M., & Rammsayer, T. H. (2014). Die Vorhersage des Studienerfolges im Masterstudiengang Psychologie durch Schul- und Bachelorstudienleistungen. *Beiträge zur Hochschulforschung, 36*(1), 30–45.
Trost, G. (2003). *Deutsche und internationale Studierfähigkeitstest. Arten, Brauchbarkeit, Handhabung* (Vol. 51). ITB Consulting; Deutscher Akademischer Austauschdienst.
Trost, G., & Haase, K. (2005). *Hochschulzulassung: Auswahlmodelle für die Zukunft. Eine Entscheidungshilfe für Hochschulen.* Stifterverband für die deutsche Wissenschaft.

Wahl, I., & Walenta, C. (2017). Evaluierung des Aufnahmeverfahrens eines Fernstudiengangs anhand des Studienerfolges: Vorhersage des Studienerfolges durch studienbezogene Lerntests, Englischtests, objektive Leistungsmotivations-tests. Intelligenztests und Eignungsinterviews. *Zeitschrift für Evaluation, 16*(1), 73–96.

Wahl, I., & Walenta, C. (2018). Mehr Zeit und bessere Vereinbarkeit durch Blended-Learning? Befragungsergebnisse von Studierenden eines berufsbegleitenden Studiengangs. In I. Buß, M. Erbsland, P. Rahn, & P. Pohlenz (Eds.), *Öffnung von Hochschulen: Impulse zur Weiterentwicklung von Studienangeboten* (pp. 233–254). Springer VS.

Wenzel, G., Walenta, C., & Wahl, I. (2019). Flexibilität und Struktur am Beispiel einer Lehrveranstaltung im Blended-Learning-Design. *Zeitschrift für Hochschulentwicklung, 14*(3), 91–101. https://doi.org/10.3217/zfhe-14-03/06.

Ingrid Wahl is a Senior Researcher at the Department of Communication at the University of Vienna. Her teaching experience ranges from (online courses) on statistics to business psychology to scientific work to supervising bachelor and master theses of (online) students on various topics. Among her research interests are tax psychology, work psychology, diversity, distance learning, teleworking, and communication.

Christa Walenta is the head of BA in Business Administration and Psychology at FernFH Distance-Learning University of Applied Sciences. Her main fields of teaching and research are business psychology, leadership psychology and management topics, diversity in organizations, personnel and organizational development as well as distance learning and working in virtual space.

Günther Wenzel is a lecturer at the Institute for Business Administration and Psychology at FERNFH Distance-Learning University of Applied Sciences. His research focuses on marketing and sales, market and consumer psychology, finance, and its teaching in distance learning.